A methodology for linking the Energy-related Policies of the European Green Deal to the 17 SDGs using Machine Learning

Koundouri, Phoebe and Theodossiou, Nicolaos and Stavridis, Charalampos and Devves, Stathis and Plataniotis, Angelos

January 2022

A methodology for linking the Energy-related Policies of the European Green Deal to the 17 SDGs using Machine Learning

Prof. Dr. Phoebe Koundouri, Professor at the School of Economics of the Athens University of Economics and Business, Director of the Research Laboratory in Socio-Economic and Environmental Sustainability (ReSEES), Director of the Sustainable Development Unit and the EIT Climate-KIC Hub of Greece, of the ATHENA Information Technology Research Center, elected member of the World Academy of Arts and Sciences (WAAS), elected president of the European Association of Environmental and Resource Economists (EAERE) and Co-chair of the Sustainable Development Solutions Network Europe (UN SDSN Europe) and Greece (UN SDSN Greece). She is also a member of the National Committee on Climate Change. Personal website: www.phoebekoundouri.org. pkoundouri@aueb.gr

Prof. Nicolaos Theodossiou, Department of Civil Engineering, AUTh, Chair of UN SDSN Black Sea, niktheod@civil.auth.gr

Charalampos Stavridis, PhD Candidate, Aristotle University of Thessaloniki, cstavrid@civil.auth.gr

Stathis Devves, PhD Candidate, Athens University of Economics and Business, devves@aueb.gr

Angelos Plataniotis, PhD candidate, National and Kapodistrian University of Athens, aplataniotis@econ.uoa.gr

Abstract

The European Green Deal (EGD) was published in December 2019 with the ambition of being Europe's new growth strategy, making it climate neutral by 2050, and ensuring its citizens a sustainable, prosperous, and inclusive future. The energy sector is central to this ambition, as the European Commission's objectives are, among others, to increase the efficiency of energy production by establishing a fully integrated, interconnected, and digitalized EU energy market. The EGD was the starting point for the publication of a large number of Policy and Strategy documents for achieving Sustainability in Europe. One of the first attempts to systematically correlate the policy areas of the European Green Deal with the 17 Sustainable Development Goals (SDGs) was made in the first report of the UN SDSN's Senior Working Group for the Joint Implementation of the SDGs and the EGD, which was published in February 2021, where the EGD framework was linked to each of the 17 SDGs using textual analysis. Building on this methodology, in this chapter we extend the manual linkage of policy texts to SDGs, by using Natural Language Processing and Machine Learning techniques to automate it, focusing on Energy-related documents derived by the EGD.

Keywords: European Green Deal, Policies, Sustainable Development Goals, Deep Learning, Natural Language Processing, Semantics.

Introduction

The European Green Deal (EGD)

Since the launching of the **European Green Deal (EGD) (European Commission, December 2019)** by the European Commission at the end of 2019, a significant number of Policy and Strategy documents have been introduced to support the effective implementation of it through the mobilization of funds and the engagement of the various actors from both public and private sector.

The EGD serves as Europe's long-term growth plan and as a roadmap for the establishment of new Policies and Strategies at both central European and national levels. It is a comprehensive strategy that aspires to make the European continent carbon-neutral by 2050, as well as resource-efficient, cutting-edge in terms of technology, and socially just.

The Energy sector is one of the fundamental Pillars to achieve the EGD objective of climate-neutrality, considering that the production and consumption of energy are responsible for more than 75% of greenhouse emissions globally and Buildings are responsible for about 40% of energy consumption (**World Resources Institute, 2020**).

Below we provide a list of the most important Policies and Strategies ***related to Energy, that were*** published in the past two years and we accompany each one with a brief description, highlighting their main points. Then we provide a method to link this set of Policies to the 17 Sustainable Development Goals of the United Nations Agenda 2030 (**United Nations, 2015**). Next, we provide a Machine Learning Method to make this linkage in an automated way. Finally, a discussion of the results is made and directions for future research are given.

Energy-related Policies derived from the EGD

A New Industrial Strategy for Europe

Europe's new industrial strategy was published in March 2020 (**European Commission, March 2020**). Its key priorities are:

1. **Commonly created transition pathways across relevant industrial ecosystems**. Challenges faced by 14 industrial ecosystems are presented in the Annual Single Market Report 2021. as they have been identified so far, but also of the transformative initiatives already undertaken to achieve the Green and Digital (Twin) Transition has to be implemented by specific and focused policies, such as clean and competitive European steel with specific measures supporting the relevant industry.
2. **Supporting the business case for the green and digital transitions**. Real business cases, and a resilient regulatory framework, infrastructure, finance for innovation financing, raw materials innovation, clean energy, demand-side measures for climate neutral, production circularity, up-skilling in labour potential, synergies between the sustainable and digital transitions. Digital solutions, such as digital twins in advanced manufacturing, can help optimization of processes in all ecosystems.
3. **Investment in skills** is a highly appreciated component of investing in labour. The European Skills Agenda supports the green and digital transitions. Mobilization of the

private sector is a critical parameter. Large-scale skills partnerships per ecosystem, skilling commitments in the critical industry fields (automotive, microelectronics, aerospace, defence industry).

EU Hydrogen Strategy

Europe's Hydrogen Strategy was published in July 2020 (**European Commission, July 2020**). The establishment of a resilient and Sustainable Eco-system for Hydrogen in Europe consists of:

- **Electricity-based hydrogen** through water electrolysis procedure. It is up to renewable electricity production to characterize as fully Green the Hydrogen Production.
- Renewable hydrogen may also be produced through the reforming of biogas (instead of natural gas) or biochemical conversion of biomass, if in compliance with sustainability requirements, therefore, **'Clean hydrogen'** refers to renewable hydrogen.
- '**fossil-based hydrogen with carbon capture**' means that greenhouse gases emitted can be captured. The variable effectiveness of greenhouse gas capture (maximum 90%) needs to be taken into account.
- **'Hydrogen-derived synthetic fuels'** refer to a variety of gaseous and liquid fuels based on hydrogen and carbon. For synthetic fuels to be considered renewable, the hydrogen part of the syngas should be renewable. Synthetic fuels include for instance synthetic kerosene in aviation, synthetic diesel for cars, and various molecules used in the production of chemicals and fertilizers.

*The **main Key actions** are Strategic investments agenda; Boosting demand for and scaling up production (incentives and supporting regulations, such as a* **Carbon Contracts for Difference scheme;** Design enabling market rules to the deployment of hydrogen; *Promoting research and innovation in hydrogen technologies;* Strengthen EU leadership in international fora for technical standards, regulations and definitions on hydrogen.

The Annual Sustainable Growth Strategy of 2021 (7 technology flagship Areas)

The Annual Sustainable Growth Strategy of 2021 for Europe was published in September 2020 (**European Commission, September 2020**).

The green and digital transformations required unprecedented expenditures in re- and upskilling. Reforms to direct public and private investments toward climate and environmental initiatives are required to complete the green transition. The EU-ID and other important digital public services, including the legal and medical systems, need to be revitalized and be accessible by all and the public administration and services, in general, will be improved by Digitalisation.

A Recovery and Resilience Facility will help rebuild and prepare the next generation, but the transition process must be equitable for all Europeans to avoid further inequalities, ensure societal support, and promote social, economic, and territorial cohesion.

Reskilling and upskilling are key to advancing the green and digital transitions, boosting innovation and growth potential, and assuring quality employment and social inclusion. The issues addressed by the Annual Sustainable Growth Strategy (ASGS) 2021 include employment, skills, health, and education. Building renovation projects will help the economy by creating jobs, lowering energy costs, improving living conditions, and reducing energy poverty. Also, by setting the European Green Deal as its cornerstone, aspires to maintain global health, prosperity, and happiness.

In the face of Energy, the EU should speed up the deployment of renewable energy and hydrogen, as well as intensify efforts to improve building energy efficiency. The promotion of future-proof clean technologies to accelerate the use of sustainable, accessible, and smart transportation will make European cities and regions cleaner. The EU's Recovery and Resilience Facility will help the Member States achieve climate neutrality and help to meet the Paris climate goals.

Cross-border and multi-country projects are necessary to promote specific types of investments, such as energy interconnectors, transportation networks, and forward-thinking digital and green projects.

Chemicals strategy for Sustainability

Chemicals strategy for Sustainability was published in October 2020 (**European Commission, October 2020**).

Human Health and a **toxic-free environment** are the keywords for reflecting the Chemical Strategy. These 2 keywords should be protected within an ambitious approach to pollution tackling from all kinds of sources in our common life and activities.

During the full lifecycle of new chemicals and materials, their safety and sustainable utilization and disposal must be ensured. The complexity and global context of manufacturing and supply chains for critical chemicals (i.e.pharmaceuticals) should be treated with **resilient value chains** and diversified sourcing towards climate-neutral and circular economy pathways. Toxic-free material cycles and clean recycling from the EU should lead to a world-widely spread benchmark. Any **substances of concern in products and recycled materials** should be minimized.

Financial instruments with regards to Innovation Calls of EU Commission aspire to support the deployment of the necessary infrastructure to develop the **use, transport and storage of electricity** from renewable/carbon-neutral energy sources, for the production of chemicals. Another dimension is the priority for KPIs of measurement of the industrial transition. The support of investments in sustainable innovations will anticipate the contamination of waste. In that mode, increasing recycling will become safe. The exports of waste (plastics and textiles) will be reduced. cleaner production process and Technologies and Process approach will be supported to be cleaner, with regards to **innovative Business Models**, and financial access with priority to **Start-ups.**

EU Strategy to reduce methane emissions

EU Strategy to reduce methane emissions was published in October 2020 (**European Commission, October 2020**) to address the challenge of Agricultural Transition. Dietary changes will be promoted through the actions relevant to Farm to Fork (F2F) Strategy. Biogas production will be supported for agricultural waste via National Strategic Plans under the Common Agricultural Policy. The gas market regulatory framework will be under review, to promote the facilitation for distributed and locally connected production of biogas. Best practices and technologies, feed and breeding changes, will be promoted for the reduction of agricultural emissions. Improvement of leak detection and repair (LDAR), regarding leaks on fossil gas infrastructure, production, transport and use will be an obligation.

Future legislation will include venting, flaring and standards for the overall supply chain, and support to the World Bank 'Zero Flaring' initiative. An expert group will develop life-cycle methane emissions methods, especially for the agricultural sector. Partner countries shall be engaged in the global coordination of energy-sector methane emissions, with regards to diplomatic action of the EU. Methane emissions from international partners will be reduced, under standards, targets (of emission reduction) and incentives for any kind of (fossil) energy carrier imported to the EU.

A Renovation Wave for Europe

Europe's Renovation Plan was published in October 2020 (**European Commission, October 2020**). The main Priorities in Renovation Wave and initiations on Energy Poverty are: Waste and renewable heat and cool into energy systems; Extension of the use of emission trading to emissions from buildings; Minimum proportions of renewable energy in buildings; Green public procurement criteria related to life cycle and climate resilience; Eco-design and energy labelling; Creating green jobs, upskilling workers and attracting new talent within the Frame of European Skills Agenda; Deep renovation standard; Horizon Europe and the R&I co-creation space; Building Information Modelling; 2050 whole life-cycle performance roadmap; Public buildings and social infrastructure showing the way; Tackling energy poverty and worst-performing buildings; Launching the Affordable Housing Initiative piloting 100 renovation districts; Reviewing the General Block Exemption Regulation and Energy and Environmental Aid Guidelines; European Bauhaus platform (combination of sustainability with art and design); digitalization in the construction sector.

EU Commission Recommendation on Energy Poverty

Europe's Recommendation on Energy Poverty was published in October 2020 (**European Commission, October 2020**). Energy Poverty has become a real challenge for the European Union, since almost 34 million citizens cannot withstand their lack of access to affordable energy for warming their households, feeding their appliances, isolating their buildings from extreme weather/temperature conditions and finally living in a comfortable home environment.

European Commission Recommendation on Energy Poverty outlines specific indicators for the proper assessment of energy-poor households, such as a) indicators for the comparison of energy expenditures versus income, b) a self-assessment on which is the current access of each household in energy services (ability on heating, cooling, warming), c) explicit monitoring by real measurements of physical parameters/figures, such as room temperatures, as a feedback of the current situation and potential / or implemented interventions, d) monitoring of the ability for households to pay for their energy bills, the quality of their home, the number of energy disconnection, as a measurement of people weakness to finance their energy consumption/ needs, e) last but not least, the electricity/fuels prices charging the end-users.

Long-Term Renovation Strategies in every Member State should take into account to erase barriers to investments relevant to energy efficiency, according to the residential profiles that are seeking an energy upgrade.

EU funding programs that are enhancing cohesion policies should be included to develop their distributional effects with regards to energy transition models and the support of vulnerable society profiles.

Finally, Energy Saving Companies (ESCO) and the relevant contracts should be mentioned as key elements to anticipate the upfront costs that are raising barriers to the vulnerable groups and to provide reliable financing solutions for the necessary energy performance interventions.

EU Strategy to harness the potential of offshore renewable energy for a climate-neutral future

The offshore renewable energy Strategy of Europe was published in November 2020 (**European Commission, November 2020**). The EU has the great advantage of having the biggest variety in sea basins, so the EU is in a position to develop offshore renewable energy. Spatial planning in maritime space is necessary and there are successful pilot projects specialized in the identification of environmental benefits of offshore wind and aquaculture. Molluscs, algae and multi-performed offshore platforms are these examples running in Belgium, Germany, France, Netherlands, Portugal (MERMAID, TROPOS, Edulis, Wier en Wind).

On the other hand, coordination between Transmission System Operators and Regulatory Authorities is necessary to support offshore renewables interconnection between local and central grids. Integration of large scale or distributed offshore energy systems into the grids should follow the regulatory rules. EU Sustainable Taxonomy and Grids development, innovation applications and technical maturity can support private capital investment.

Research and innovation can also support the choice of the proper materials (i.e corrosion resistance etc.) and ensure high capacity rates and long-term efficiency and "sufficiency". Re-skill and Up-skill is an important challenge with the potential of high-quality employment opportunities to skilled workers affected by the transition. The Circular Economy approach is important to focus on recycling and re-use of renewable machinery equipment components (i.e. wind turbine blades), affecting the modern design and provoking niche and non-generic applications.

New technologies and methods should be developed and deployed for this, increasing the value retention of products and services within the manufacturing sector. Last but not least, through Green Deal diplomacy, the EU is engaged with its international partners to develop an ecosystem of offshore renewables in cooperation with low-income countries and emerging markets.

Smart Mobility Strategy

The Strategy of Europe for Smart Mobility was published in December 2020 (**European Commission, December 2020**). Mass Transportation and mobility is a crucial parameter in the social life of citizens. But it has to cut off the costs derived by air-water, greenhouse emission pollution. Environmental impacts affect human health, wellbeing and biodiversity. To protect these public goods, especially the non-market goods, the mobility sector has to become sustainable, green digitalised and serve the energy transition towards clean, low carbon, green fuels or other energy carriers that are facilitating modern transportation solutions.

Charging and fuel infrastructure for zero-emission vehicles, the supply chain for renewable and low-carbon fuels, can support developed green mobility in cities with more healthy and environmentally friendly conditions.

On the other hand, digitalisation and Artificial Intelligence services can enhance circularity and efficiency in mass production and travel for vehicles, aircraft and shipping, considering mobility as a service (MAaS). Therefore, safe, comfortable, reliable, interconnected (for both urban and rural areas) and affordable for all mobility, especially for the vulnerable groups of people, should be maximized via digitally powered competitive advantages and logistic chains with increased resilience. All these provide good chances for all in social life, the development of new skills and, as a result, new and better opportunities for attractive jobs.

Commission has settled down certain milestones for European Smart and Sustainable Mobility:

By 2030, at least 30 million zero-emission cars and 80 000 zero-emission heavy-duty vehicles should be in operation. By 2050, only these categories should be in traffic. 2) Zero-emission ocean-going vessels and large zero-emission aircraft will be introduced in the market by 2030 and 2035, respectively. 3) Traffic on High-speed rail utilization should be twice by 2030 and 3 times by 2050. 4) Transport by inland waterways and short sea transport systems should be increased by 25% by 2030 and by 50% by 2050.

Updating the 2020 New Industrial Strategy: Building a stronger Single Market for Europe's recovery

The European Commission published an update for the Industrial Strategy in May 2021 (**European Commission, May 2021**). The Upgrade of the New Industrial Strategy relies on a) the monitoring of industry trends and the **Key Performance Indicators (KPI's)** of competitiveness b) making the Single Market more resilient to anticipate crises and support the recovery. c) the toolbox development for the release of strategic dependencies. d) the boost of Digital & Energy Transition.

Commission in combination with the current tools (i.e. Single Market Scoreboard) will monitor main indicators of the competitiveness of the European economy such as:

1. Single Market integration, based on indicators on intra-EU trade across the Member States, to support the innovation ecosystems.
2. Productivity growth.
3. International competitiveness, for open and fair access to export markets.
4. Public and private investment, as a percentage of GDP, performs with regards to the capabilities of both digital and energy potential transformation of the economy.
5. R&D expenditure as a percentage of GDP, to indicate the innovations support.

For the Single Market Resilience, the Key Actions are (1) The Single Market Emergency Instrument, (2) The Annual Single Market Report, (3) A potential consolidated template for the declaration of workers posting, (4) The Surveillance of European and imported products, (5) The Capital or Equity financing of the SME's (under InvestEU).

The Commission is preparing the Alliance on processors and semiconductor technologies and the Alliance for Industrial Data, Edge and Cloud.

The Commission is also considering the preparation of an Alliance on Space Launchers for the autonomous European access to space, as well as an Alliance on Zero Emission Aviation to provide the future fuels and energy carriers for aircraft (e.g. hydrogen, electric), with regards to the European 2050 climate neutrality objective, as a complementary Group to the Renewable and Low-Carbon Fuels Alliance.

Important Projects of Common European Interest (IPCEIs) will be still supported by the Commission, especially in cases the market alone cannot deliver breakthrough innovation cannot be implemented throughout the market itself, such as in the case of batteries and microelectronics.

The Commission will strengthen the synergies between the Green and Digital transitions. Digital twins for the development of the digital copies, historic data, predictive analytics, and all Asset Management Performance of Critical Industrial Equipment will reinforce the reliability of all involved ecosystems.

European Green Digital Coalition will implement the monitoring of green/digital combination solutions and the derived impact. The carbon neutrality target of ICT technologies, according to the commitments of the Digital Compass for the reduction of the environmental footprint with regards to the digital technologies.

The twin transition will be supported by the Investment in skills. On the other hand, the twin transition will enhance the Just recovery. The Pact for Skills is a tool of the European Skills Agenda. The European Labor force will be improved in upskilling and reskilling throughout the mobilization of Private Enterprises, Institutes, Academia and all the relevant Stakeholders.

The European Green Deal and the 17 SDGs

The **Agenda 2030** for Sustainable Development, agreed by all UN Member States in 2015, contains **17 Sustainable Development Goals (SDGs)** and 169 objectives (Table 1). The Agenda is a pledge to eradicate poverty and achieve sustainable development on a global scale by 2030, considering three pillars of sustainable development – *economic*, *social*, and *environmental*. The SDGs are global in scope and universal in application, considering the different national specificities, capacities, and stages of development, as well as specific difficulties. Thus, all countries share responsibility for achieving the SDGs and each has a critical role to play locally, nationally, and globally, under the United Nations dogma of "leaving no one behind" (**United Nations, 2015**).

Goal	The ambition of the Goals
Goal 1 - No Poverty	End poverty in all its forms everywhere
Goal 2 - Zero Hunger	End hunger, achieve food security and improve nutrition and promote sustainable agriculture
Goal 3 - Good Health & Well Being	Ensure healthy lives and promote well-being for all at all ages
Goal 4 - Quality Education	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
Goal 5 - Gender Equality	Achieve gender equality and empower all women and girls
Goal 6 - Clean Water & Sanitation	Ensure availability and sustainable management of water and sanitation for all
Goal 7 - Affordable & Clean Energy	Ensure access to affordable, reliable, sustainable, and modern energy for all
Goal 8 - Decent Work & Economic Growth	Promote sustained, inclusive, and sustainable economic growth, full and productive employment, and decent work for all
Goal 9 - Industry, Innovation & Infrastructure	Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation
Goal 10 - Reduced Inequalities	Reduce inequality within and among countries

Goal 11 - Sustainable Cities & Communities	Make cities and human settlements inclusive, safe, resilient, and sustainable
Goal 12 - Response Consumption & Production	Ensure sustainable consumption and production patterns
Goal 13 - Climate Action	Take urgent action to combat climate change and its impacts
Goal 14 - Life Below Water	Conserve and sustainably use the oceans, seas, and marine resources for sustainable development
Goal 15 - Life On Land	Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, halt and reverse land degradation and halt biodiversity loss
Goal 16 - Peace Justice & Strong Institutions	Promote peaceful and inclusive societies for sustainable development, provide access to justice for all, and build effective, accountable, and inclusive institutions at all levels
Goal 17 - Partnerships for the Goals	Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development

Table 1 The 17 Sustainable Development Goals of UN Agenda 2030.

The EGD and the 17-SDGs context share common objectives, meaning that the implementation of EGD policies would at the same time support actions that will contribute to achieving various SDGs. In (**Sachs, J., Koundouri, P., et al., 2021**) a methodology is presented to link the objectives of the EGD Policy Areas with those of the 17 SDGs, which is based on a text-mining exercise to match specific parts of the EGD document to all relevant SDGs. Their results are very vibrant on the relationship between the two frameworks. Figure 1 demonstrates the relation between SDGs and EGD policy areas. Dark green represents a straight connection between EGD and SDG, according to the number of the EGD text extracts that are conceptually similar to the SDG ambition. Light green coloured cells illustrate indirectly associations between EGD and SDGs, and white show a weak or no obvious linkage.

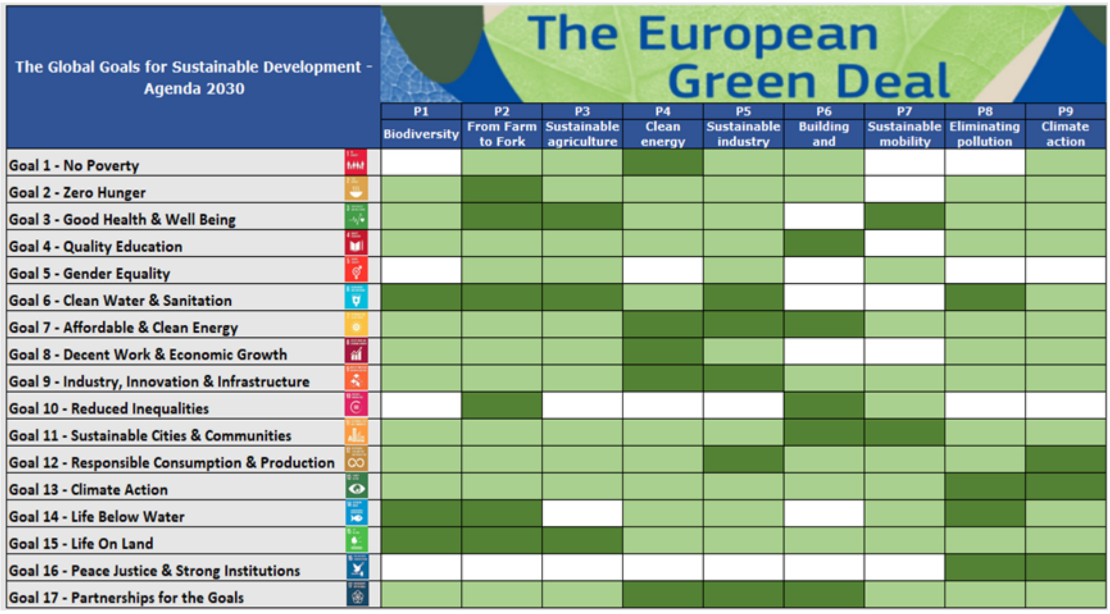

Figure 1 The correlation between the 17 SDGs and the European Green Deal Policy areas. Source: ***Sachs, J., Koundouri, P., et al., 2021***

The Sustainable Development Goals (SDGs), like the Paris Climate Agreement, call for profound changes in every country, requiring coordinated efforts by governments, civil society, research, and business. However, stakeholders lack a broad consensus on how to operationalize the 17 SDGs, but (**Sachs et al. 2019**), suggest 6 major categories of Transformations that integrate SDGs in public policy interventions, namely: (1) education, gender and inequality; (2) health, well-being and demography; (3) energy decarbonization and sustainable industry; (4) sustainable food, land, water and oceans; (5) sustainable cities and communities; and (6) digital revolution for sustainable development. Each transformation is accompanied by recommended key investments and regulatory concerns and as a result, government structures can be used to operationalize transformations while still respecting the 17 SDGs' interdependencies (Figure 2).

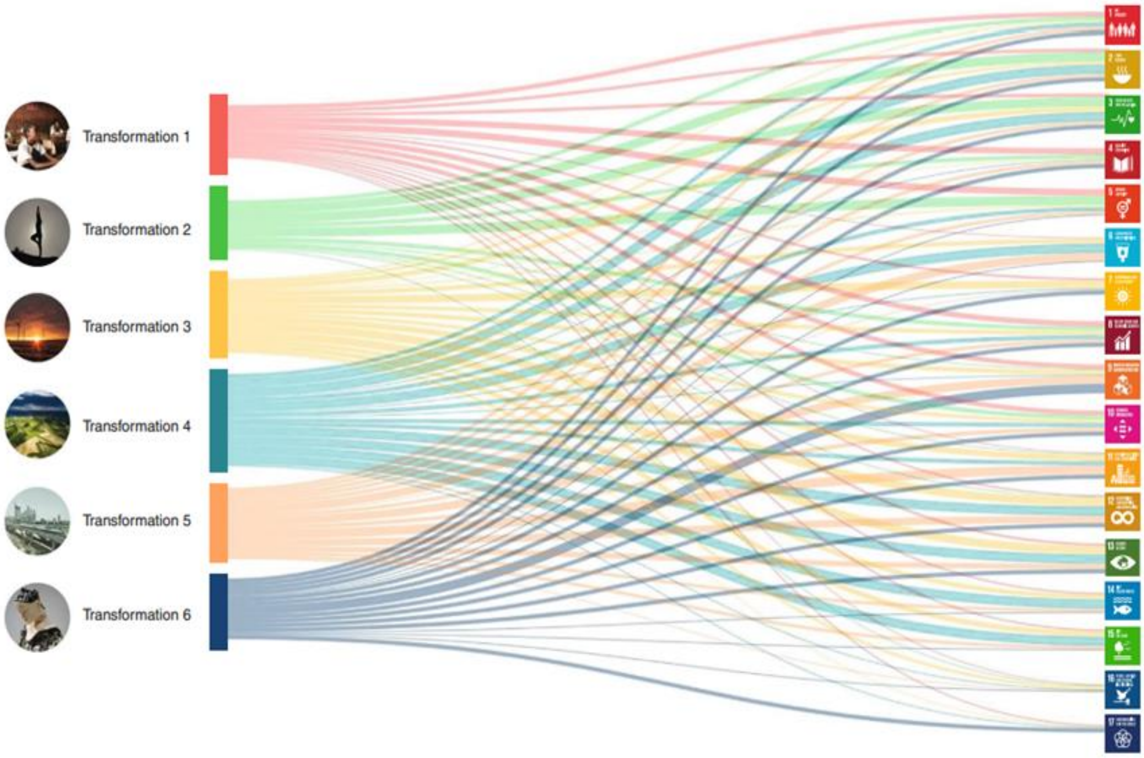

Figure 2 The translation of the 17 SDG into 6 Transformations. Source: Sachs et al. 2019

Alignment between Energy-related Policies and the 17 SDGs

One of the political guidelines of Ursula von der Leyen (**von der Leyen U., 2019**), the President of the European Commission, is that the process of **European Semester**[1] must be reoriented and become an instrument that integrates the **United Nations Sustainable Development Goals (SDGs) derived from Agenda 2030 (United Nations, 2015)**, which constitute the most widely accepted framework for sustainable development, globally.

It is believed that the integration of SDGs into the European Policy Framework, will help ensure that Europe's pathway in achieving climate neutrality, will happen within a comprehensive economic framework that gives equal opportunities to everyone, and properly.

Below, we present a method used to identify how each Policy mentioned above, is related to the 17 Sustainable Development Goals and, thus, contributes to their achievement in various levels.

The linkage is made by identifying phrases or sentences in each Policy text that are conceptually related to specific SDGs. Then, assuming that the greater the number of relevant phrases or sentences, the greater the influence of the Policy on the SDGs, we assign a score to show the level of impact, using a 4-point scale, as follows:

- 3, the Policy document directly affects the SDG outcomes;
- 2, the Policy document reinforces the SDG outcomes;
- 1, the Policy document enable the SDG outcomes;
- 0, the Policy document do not interact with the SDG

[1] The European Semester serves as a framework for the integrated monitoring and coordination of economic and employment policies across the European Union. Since its inception in 2011, it has become a well-established forum for discussing the fiscal, economic, and employment policy challenges confronting EU countries on a yearly basis

	SDG 1-No poverty	SDG 2-Zero hunger	SDG 3-Good health and well-being	SDG 4-Quality education	SDG 5-Gender equality	SDG 6-Clean water and sanitation	SDG 7-Affordable and clean energy	SDG 8-Decent work and economic growth	SDG 9-Industry, innovation and infrastructure	SDG 10-Reduced inequalities	SDG 11-Sustainable cities and communities	SDG 12-Responsible consumption and production	SDG 13-Climate action	SDG 14-Life below water	SDG 15-Life on land	SDG 16-Peace, justice and strong institutions	SDG 17-Partnerships for the goals
A New Industrial Strategy for Europe (COM/2020/102)	1	2	1	2	0	0	3	2	3	0	1	2	2	1	2	2	2
EU Hydrogen Strategy	1	0	0	2	0	0	3	2	3	1	2	2	3	0	0	2	1
7 technology flagship Areas, ASGS for 2021	0	0	2	1	1	0	2	3	3	3	3	2	2	0	1	2	1
Chemicals strategy for Sustainability	0	1	3	0	0	0	1	0	3	0	1	2	3	3	3	1	0
EU Strategy to reduce methane emissions	1	3	1	1	0	0	2	1	2	0	1	2	1	1	1	1	1
A Renovation Wave for Europe	1	0	0	1	0	0	3	1	2	0	3	2	3	1	1	1	1
EU Commission Recommendation on Energy Poverty	3	0	0	0	0	0	2	2	0	3	1	1	2	0	0	0	0
EU Strategy to harness the potential of offshore renewable energy for a climate neutral future	0	0	0	1	0	0	3	2	3	0	2	1	3	2	0	2	2
Smart Mobility Strategy	0	1	2	0	0	0	3	0	3	2	2	2	3	2	0	0	1
Updating the 2020 New Industrial Strategy: Building a stronger Single Market for Europe's recovery	1	2	1	2	0	0	3	2	3	0	1	2	2	1	2	2	2

Table 2 Connection between Policies and the 17 SDGs

A Machine Learning method to evaluate the connection between Policy documents and the 17 SDGs

Within the scope of work undertaken by the authors, was the utilization of Machine Learning for three main reasons: a) to validate the linkages found by the classical approach, b) to create a tool that will serve as a fast classifier for present and future work and c) to discover any new possible connections between the SDGs and the policy documents scanned that were not identified during the classical approach in the first place

As (**Wang et al., 2009**) suggested, Machine Learning is the "fundamental way to make the computer intelligent" and it offers a fast accumulation of knowledge. Thus, it is widely used in Natural Language Processing problems to aid in text analysis.

Natural Language Processing (or NLP for short) is a set of methods for making the human language accessible to computers (**Eisenstein, 2019**) or else a set of methods that allow the analysis, the understanding and translating of the meaning of human language from computers. The core idea is that for a machine to "understand" what is written or spoken, the message is necessary to break down into a numerical format.

While NLP is widely used in our everyday lives, such as spam e-mail detection, machine translation (**Wu et al, 2016**), chatbots and virtual agents development, text classification and summarization, there are major challenges in its implementation. The Natural Language is quite ambiguous. For example, a given text may consist of contextual words and phrases, homonyms and synonyms, errors and slang and may express irony or sarcasm. The above cannot be easily understood and deciphered by a machine

For our study, we have developed and deployed two different models. The first one is a simple Information Retrieval Model using Bag-of-Words, while the second one is a more developed and complicated one using Deep Learning Techniques (BERT).

Information Retrieval

Information Retrieval (IR), as the term reveals, is the field involved with the search and retrieval of information. It is concerned with all the activities related to the organization, processing and access of information of all forms and formats (**Chowdhury, 2010**). The purpose of IR is the rapid retrieval of documents, texts and information in general, based on a user's query. An example of IR, and a rather advanced one, is the Google Search Engine.

Bag-of-Words (BoW), a term first used linguistically by (**Zellig, 1954**), is among the prevalent techniques used for textual Information Retrieval (**Passalis & Tefas, 2018**). It is a simplifying representation of a document, in which words of a text are presented as a multiset (Bag) disregarding grammar and syntax while keeping simplicity.

The BoW technique was used for the study, to quickly find similarities between energy policy documents and the 17 Sustainable Development Goals (SDGs). Specifically, 17 different vocabularies were built, each consisting of words and phrases describing the essence of each SDG. For instance, SDG 4("Quality Education") was described with keywords such as: “primary education", "secondary education", "reading", "mathematics", “organized learning", "information technology", "equal access", "training", "literacy", "numeracy", "fixed level of proficiency", "learner", "curricula", "scientific programs", "teachers", "Learning opportunities", "School enrolment", "Secondary education", "Teacher training", "Universal education", ”Vocational training", while SDG 6 (“Clean Water and Sanitation”) was described with keywords, such as "drinking water", "safe water", "water service", "water sanitation", "hygiene", "defecation", "hand-washing facilities", "shower", "wastewater treatment", "water-use", "freshwater”, "water management" etc.

Next, policy documents were read in a .html form and were preprocessed by lowercasing the documents, removing numbers, punctuation, stop words, contractions and whitespaces. Both the vocabularies and the pre-processed policy documents were transformed into vectors by using CountVectorizer provided by the sci-kit-learn.

To calculate the similarity between each policy token and each vocabulary token was calculated using the cosine similarity, to ignore the magnitude of the vectors compared.

Results 1

In Table 3 the correlation between each Energy Policy with the SDGs is presented. Higher scores present higher similarities.

Short Name	**SDG1**	**SDG2**	**SDG3**	**SDG4**	**SDG5**	**SDG6**	**SDG7**	**SDG8**	**SDG9**	**SDG10**	**SDG11**	**SDG12**	**SDG13**	**SDG14**	**SDG15**	**SDG16**	**SDG17**
A New Industrial Strategy	0,451	0,333	0,361	0,417	0,278	0,352	0,552	0,512	0,533	0,471	0,423	0,520	0,398	0,427	0,307	0,422	0,556
Hydrogen Strategy	0,468	0,410	0,375	0,385	0,259	0,410	0,659	0,485	0,556	0,494	0,473	0,512	0,545	0,418	0,346	0,394	0,541
7 technology flagship Areas, ASGS for 2021	0,538	0,367	0,402	0,468	0,344	0,401	0,544	0,525	0,590	0,485	0,522	0,525	0,497	0,388	0,346	0,422	0,504
Chemicals strategy for Sustainability	0,500	0,422	0,489	0,447	0,359	0,491	0,528	0,499	0,596	0,480	0,494	0,568	0,484	0,436	0,453	0,441	0,551
Strategy to reduce methane emissions	0,468	0,455	0,402	0,385	0,287	0,468	0,582	0,431	0,508	0,456	0,502	0,545	0,585	0,388	0,453	0,408	0,470
Renovation Wave	0,523	0,410	0,415	0,454	0,337	0,452	0,666	0,534	0,596	0,508	0,605	0,584	0,568	0,436	0,403	0,422	0,556
Recommendation on Energy Poverty	0,459	0,294	0,280	0,278	0,238	0,296	0,475	0,392	0,368	0,375	0,393	0,343	0,390	0,309	0,226	0,316	0,379
Strategy for offshore renewable energy	0,508	0,404	0,368	0,454	0,313	0,401	0,639	0,508	0,579	0,485	0,486	0,557	0,497	0,527	0,392	0,387	0,546
Smart Mobility	0,552	0,427	0,451	0,425	0,352	0,444	0,604	0,555	0,606	0,542	0,577	0,565	0,528	0,480	0,392	0,477	0,566
Updating the 2020 Industrial Strategy	0,451	0,360	0,368	0,393	0,313	0,392	0,544	0,530	0,527	0,485	0,469	0,525	0,398	0,418	0,307	0,435	0,561

Table 3 Correlation between each Energy Policy with the SDGs

The highest similarity score of each policy does not exceed 0.70, meaning that the algorithm did not find a high correlation with a certain SDG. This can be due to the preprocessing of the documents and the vocabularies’ construction. Although this might seem disappointing at a first sight, one can draw several conclusions after carefully examining the results.

Figure 3 Correlation between each Energy Policy with the SDGs in Bar Chart

To further understand the correlation and to easily compare the results with the "human" approach described in the previous chapter, we have assigned the values of 0, 1, 2 and 3 according to a range of similarity scores, as follows:

- Similarity scores below 0.3 are transformed to 0;
- Similarity scores between 0.3 and 0.4 are assigned to 1;
- Similarity scores between 0.4 and 0.5 are assigned to 2;
- Similarity scores greater than 0.5 are assigned to 3;

By implementing the aforementioned allocation Table 3 is transformed as seen below:

Short Name	SDG1	SDG2	SDG3	SDG4	SDG5	SDG6	SDG7	SDG8	SDG9	SDG10	SDG11	SDG12	SDG13	SDG14	SDG15	SDG16	SDG17
A New Industrial Strategy	2	1	1	2	0	1	3	3	3	2	2	3	1	2	1	2	3
Hydrogen Strategy	2	2	1	1	0	2	3	2	3	2	2	3	3	2	1	1	3
7 technology flagship Areas, ASGS for 2021	3	1	2	2	1	2	3	3	3	2	3	3	2	1	1	2	3
Chemicals strategy for Sustainability	2	2	2	2	1	2	3	2	3	2	2	3	2	2	2	2	3
Strategy to reduce methane emissions	2	2	2	1	0	2	3	2	3	2	3	3	3	1	2	2	2
Renovation Wave	3	2	2	2	1	2	3	3	3	3	3	3	3	2	2	2	3
Recommendation on Energy Poverty	2	0	0	0	0	0	2	1	1	1	1	1	1	1	0	1	1
Strategy for offshore renewable energy	3	2	1	2	1	2	3	3	3	2	2	3	2	3	1	1	3
Smart Mobility	3	2	2	2	1	2	3	3	3	3	3	3	3	2	1	2	3
Updating the 2020 Industrial Strategy	2	1	1	1	1	1	3	3	3	2	2	3	1	2	1	2	3

Table 4 Transformed similarity scores on a 4-point scale

As one can easily notice, all energy strategies seem to be highly connected to SDG 7 ("Affordable and Clean Energy"), following SDG 9 ("Industry and Infrastructure"), SDG 12 ("Responsible Consumption and Production"), and SDG 8 ("Decent Work and Economic Growth"), goals that are expected to be connected to energy production and management. On the other, energy policies seem to be less related to SDG 5 ("Gender Equality"), SDG 15 ("Life on Land") and SDG 3 ("Good Health and Well-Being"), which seems again sensible.

Comparing and contrasting the table with the one produced from the classical approach of reading and rating each policy's connection to the SDGs, it can be observed, that the algorithm, in general, did well. Of course, due to the use of similar keywords in some vocabularies and due to the semantic similarity of some of the SDGs, the similarity scores produced by the algorithm are, in most cases, higher.

A great difference is observed in SDG 6 ("Clean Water and Sanitation"), in which there was no correlation found following the classical approach, but some connection with most policies was produced by the algorithm. This can mean either that different and maybe, more targeted keywords should be used as a vocabulary for describing SDG 6, or that there might be a connection, found by the algorithm, that was not detected by the classical approach of reading through the policies and scoring their similarity.

Another odd and unexpected result is the high similarity scores produced in SDG 1 ("No poverty"), as the Classical approach produced none or little correlation between SDG 1 and the scanned energy policies. Again, this might have happened due to the words and expressions chosen as the Goal's describing vocabulary.

Of course, one must not forget that a machine is not human as it lacks semantic capabilities, which are crucial for understanding what a text communicates. Considering that the SDGs are closely related and interconnected, a machine cannot easily capture the essence of each SDG. For instance, "No poverty" and "Zero Hunger" are two very interrelated Goals that, in some cases, relevant texts or phrases cannot be easily labelled or even classified based on their correspondence with the first or the second one.

After we compared and contrasted the results with the classical approach, we decided to take a step further and try to build an even better and more developed algorithm that would perform better and have semantic capabilities.

Deep Learning

As said in the previous chapter, the similarity tool developed showed possible correlations between the SDGs and the scanned policy documents. To make a step further, we decided to deploy a model that has the semantic capability, since our goal was to find the semantic similarity[2] of the EU energy policies and the SDGs. To do so, we used deep learning.

Deep learning refers to very large neural networks with many layers (deep), that "allow computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction" (**LeCun et al, 2015**). Simply put, machines learn from experience by representations that are expressed in terms of

[2] Semantic Similarity is the task of determining how similar two sentences are, in terms of what they mean

other, simpler representations (Goodfellow et al, 2016). In 2017, Google Research introduced The Transformer, a deep learning model, which is based on attention mechanisms, dispensing with recurrence and convolutions entirely (Vaswani et al, 2017). This innovation led to the development of a wide range of models based on transformers, allowing the processing of entire sequences without the need for labelled data in pre-training.

As a result, and taking particularly into account the ambiguity of the Natural Language (Lexically, Syntactically, Semantically, Anaphorically, Pragmatically), we have fine-tuned a pre-trained transformer-based model to find the similarity score of each policy document with each SDG.

BERT, which stands for "Bidirectional Encoder Representations from Transformers", was firstly introduced by Google Research in 2018 (Devlin et al, 2018) and revolutionized the approach of modelling. Standard Language Models are unidirectional, limiting the architectures that can be used for pre-training, while BERT is a bidirectional transformer pre-trained by using masked language modelling objective and next sentence prediction. Data used for its pretraining are the Toronto Book Corpus (approx. 800 million words) and Wikipedia (approx. 2,500 million words). For the study, the "bert-base-uncased" (12-layer, 768-hidden, 12-heads, 110M parameters) model was used.

To fine-tune the pre-trained BERT model, we have used the latest available OSDG Community Dataset, along with the description of the targets and indicators of each Goal. As the OSDG Community Dataset does not include texts for SDG16 and SDG17 yet, we have used expressions and sentences linked to those specific goals from the humanely scanning related policies.

The OSDG Community Dataset is the result of the work of more than 1,000 volunteers from all over the world using the OSDG Community Platform, who label sentences and paragraphs according to their relationship with each SDG. Each labelling exercise is a binary decision problem, in which the volunteer has to answer positively or negatively if the presented excerpt is connected to the presented SDG or not.

The dataset is formed after text excerpts of paragraphs deriving from public documents, such as reports, policies and publication abstracts. Furthermore, some documents originate from UN-related sources (e.g. SDG-Pathfinder and SDG Library). The released dataset (OSDG, UNDP IICPSD SDG AI Lab, & PPMI. (2021) constitutes 32,115 labelled document excerpts and it contains the referred SDG, the number of volunteers that classified the connection to the SDG as negative, the number of volunteers that classified the connection to the SDG as positive and the agreement score based on the formula:

$$agreement = \frac{|labels_{positive} - labels_{negative}|}{labels_{positive} + labels_{negative}}$$

For the presented research, data used were pre-selected using the following criteria:

1. $labels_{positive} > labels_{negative}$, as we needed to only use data related to an SDG

2. agreement > 0.6, as we needed to be sure that the volunteers agreed to the labelling

This pre-selection process produced 14,280 excerpts, in which minor corrections occurred, such as joined words separation and wrong letter replacing (e.g. in some words the letter ‘r’ was ‘i’) etc. In the final set of excerpts, we have added the indicators and the descriptions for each of the 169 targets of the SDGs, as retrieved from the website sdg-tracker.org, leading to a total of 15,083 text excerpts for fine-tuning the model.

The number of text excerpts used for each SDG is presented below:

SDG	Number of excerpts used for fine-tuning
1	970
2	726
3	1639
4	1993
5	1948
6	1132
7	1391
8	759
9	605
10	385
11	1081
12	232
13	991
14	613
15	480
16	66
17	72

Table 5 Number of text excerpts used for each SDG

The model was developed in Python, using PyTorch and Scikit-Learn. 80% of the text excerpts were used as training data and 20% as testing data. Adam Optimizer was used as an optimizer, while the Cross-Entropy Loss was chosen as the optimization criterion. Next, the model was trained for 10 Epochs with a learning rate of 10-5, resulting in an accuracy score of 0.889.

Considering that the training data used for each SDG are not of the same size and the fact that a text excerpt is most probably linked to more than one SDG, the accuracy score is acceptable. In other words, the algorithm simply answers the question “how probable is the X energy policy document to be related to the Y SDG”

By deploying the trained model, the same ten energy policy documents were classified, in the form of .html files as published by the European Commission, giving the results presented below.

Results 2

Similarity scores calculated by the model are presented in Table 6. The higher the score the bigger the probability for a scanned policy to be linked to a certain SDG. It is noted, that to get better insight, one should view the presented results with a qualitative and not with a quantitative look.

POLICY DOCUMENT	SDG Correlation																
Column1	**SDG1**	**SDG2**	**SDG3**	**SDG4**	**SDG5**	**SDG6**	**SDG7**	**SDG8**	**SDG9**	**SDG10**	**SDG11**	**SDG12**	**SDG13**	**SDG14**	**SDG15**	**SDG16**	**SDG17**
A New Industrial Strategy	0,00%	0,04%	0,02%	0,04%	0,03%	0,02%	0,11%	0,21%	99,11%	0,02%	0,03%	0,19%	0,04%	0,03%	0,03%	0,02%	0,05%
Hydrogen Strategy	0,05%	0,09%	0,09%	0,03%	0,02%	0,11%	92,81%	0,07%	0,14%	0,05%	0,12%	1,19%	2,36%	0,11%	0,04%	0,62%	2,08%
7 technology flagship Areas, ASGS for 2021	0,05%	0,47%	1,58%	0,14%	0,08%	0,34%	1,64%	0,19%	92,25%	0,17%	0,21%	0,16%	0,94%	0,23%	0,10%	0,26%	1,17%
Chemicals strategy for Sustainability	0,04%	0,15%	1,81%	0,08%	0,12%	1,60%	0,97%	0,06%	0,21%	0,08%	0,27%	92,41%	0,25%	0,81%	0,31%	0,43%	0,39%
Strategy to reduce methane emissions	0,13%	0,07%	0,16%	0,03%	0,05%	0,41%	84,73%	0,13%	0,08%	0,13%	0,38%	4,44%	4,25%	0,12%	0,09%	1,39%	3,39%
Renovation Wave	0,01%	0,01%	0,02%	0,03%	0,02%	0,04%	0,08%	0,03%	0,07%	0,03%	98,96%	0,26%	0,24%	0,03%	0,03%	0,09%	0,03%
Recommendation on Energy Poverty	0,03%	0,03%	0,03%	0,02%	0,01%	0,15%	98,30%	0,05%	0,03%	0,02%	0,07%	0,41%	0,22%	0,06%	0,01%	0,20%	0,35%
Strategy for offshore renewable energy	0,01%	0,01%	0,01%	0,01%	0,01%	0,02%	99,20%	0,05%	0,03%	0,01%	0,04%	0,26%	0,13%	0,03%	0,01%	0,06%	0,11%
Smart Mobility	0,01%	0,01%	0,03%	0,02%	0,01%	0,02%	0,04%	0,01%	0,07%	0,01%	99,64%	0,04%	0,02%	0,02%	0,01%	0,03%	0,01%
Updating the 2020 Industrial Strategy	0,03%	0,10%	0,09%	0,13%	0,12%	0,11%	1,16%	3,69%	87,95%	0,11%	0,12%	5,61%	0,16%	0,12%	0,12%	0,10%	0,28%

Table 6 Total Similarity Scores

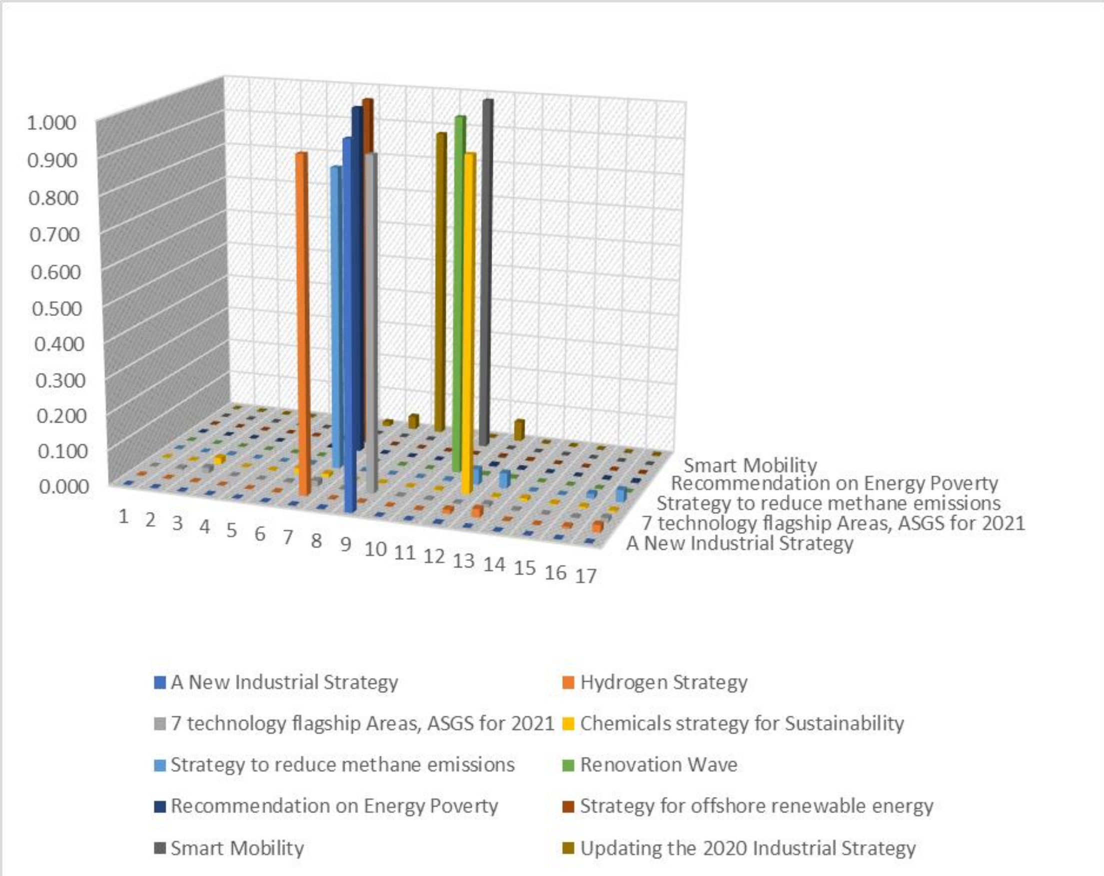

Figure 4 Total Similarity Scores in Bar Chart

Surprisingly, not all Energy Policies scanned are closely related to SDG 7 ("Affordable and Clean Energy"), as was expected. Only "Hydrogen Strategy", "Strategy to reduce methane emissions", "Recommendation on Energy Poverty" and "Strategy for offshore renewable energy" found to be closely related to SDG 7. "A New Industrial Strategy", "7 Technology flagship Areas" and "Updating the 2020 Industrial Strategy" were found to be probably more related to SDG 9 "Industry, Innovation and Infrastructure". On the other hand, "Renovation Wave" and "Smart Mobility" seem to be more related to SDG 11 "Sustainable Cities and Communities", while "Chemicals Strategy for Sustainability" seems to be closer to SDG 12 "Responsible Consumption and Production".

To understand the correlation between the policies and the SDGs, and to draw meaningful conclusions, the rest of the percentages should be examined as well. To do so, we excluded the highest score of each row (policy document) and by re-adjusting the percentages, we get Table 7:

Short Name	SDG1	SDG2	SDG3	SDG4	SDG5	SDG6	SDG7	SDG8	SDG9	SDG10	SDG11	SDG12	SDG13	SDG14	SDG15	SDG16	SDG17
A New Industrial Strategy	0,54%	4,04%	2,31%	4,86%	3,61%	2,50%	12,25%	23,41%		2,13%	3,89%	21,38%	4,21%	3,15%	3,83%	2,08%	5,80%
Hydrogen Strategy	0,76%	1,27%	1,19%	0,39%	0,34%	1,58%		0,97%	2,02%	0,66%	1,71%	16,52%	32,81%	1,58%	0,59%	8,65%	28,97%
7 technology flagship Areas, ASGS for 2021	0,70%	6,04%	20,37%	1,85%	0,97%	4,44%	21,13%	2,50%		2,19%	2,74%	2,09%	12,15%	3,02%	1,29%	3,38%	15,15%
Chemicals strategy for Sustainability	0,58%	2,04%	23,87%	1,08%	1,60%	21,09%	12,72%	0,79%	2,75%	1,08%	3,60%		3,24%	10,65%	4,13%	5,65%	5,13%
Strategy to reduce methane emissions	0,87%	0,48%	1,07%	0,19%	0,34%	2,72%		0,85%	0,53%	0,86%	2,52%	29,09%	27,81%	0,81%	0,58%	9,08%	22,20%
Renovation Wave	1,11%	1,04%	2,24%	2,95%	2,03%	4,31%	7,97%	3,24%	6,55%	2,52%		25,18%	22,84%	2,95%	3,32%	8,46%	3,29%
Recommendation on Energy Poverty	1,55%	2,01%	1,66%	0,90%	0,75%	8,99%		3,00%	2,05%	1,02%	4,15%	24,03%	13,17%	3,38%	0,81%	11,79%	20,74%
Strategy for offshore renewable energy	1,33%	0,92%	1,64%	1,16%	1,06%	2,95%		6,15%	3,98%	1,63%	4,39%	32,83%	16,39%	4,17%	0,91%	7,12%	13,35%
Smart Mobility	1,56%	1,67%	8,49%	5,26%	3,96%	6,58%	10,11%	2,96%	20,42%	3,22%		10,09%	6,09%	5,42%	3,52%	8,37%	2,28%
Updating the 2020 Industrial Strategy	0,25%	0,82%	0,74%	1,10%	0,97%	0,87%	9,62%	30,67%		0,95%	1,03%	46,54%	1,30%	0,99%	0,99%	0,86%	2,30%

Table 7 Correction Table

The differences between each policy can be now clearly understood. For instance, energy policies that are related more to industry, such as “A New Industrial Strategy” and “Updating the 2020 Industrial Strategy” appear to be more connected to SDG 8 “Decent Work and Economic Growth” and SDG 12 “Responsible Consumption and Production”, rather than SDG 7. On the other hand, policies that had their highest similarity score to SDG 7 or SDG 11 are, on a lower level, more connected to SDG 12 and SDG 13 “Climate Action”.

Another interesting insight is that “Chemicals Strategy for Sustainability” seems to be closely connected to SDG 3 “Good Health and Well-being” and SDG 6 “Clean Water and Sanitation”. SDG3 also has an adequate similarity score with the “7 Technology Flagship Areas”, while almost all energy policy documents seem to be closely connected to SDG 17, even though it had very few labelled text excerpts in comparison to the rest of the SDGs.

Discussion on the results

In this study, three different approaches were used to find a correlation between EU energy policy documents and the SDGs. The first one referred to a “classical approach”, which consists of reading the policy documents, derived from the EGD and relate to energy, and finding phrases and extracts related to each SDGs. The second one uses a simple Information Retrieval Bag-of-Words Technique, which seeks correlation between the energy policy documents and the SDGs using cosine similarity of vectors, formed by the aforementioned documents and sets of keywords describing each SDGs, referred to as vocabularies. The third one, involves the fine-tuning of a pre-trained BERT model (deep learning), using over 15,000 excerpts from documents, which were labelled by the OSDG Volunteer Community.

Assuming that the “classical approach” is a standard and concrete way of finding connections between two documents or a document and a concept, although time-consuming, its results were used as a benchmark for the effectiveness of the two other methods.

Overall, both machine approaches worked, providing a quick way of classifying a policy document regarding the 17 SDGs. The first algorithmic approach, performs well, as it captures the correlation of energy-related Policies with SDG 7, SDG 9 and SDG 12 as expected. However, some deviations were observed, namely unexpected connections with SDG 1, SDG 5 and SDG 6. Most probably this is due to the vocabularies used, or the fact that the algorithm does not have semantic capabilities to decipher the true meaning and the use of words and phrases. Nevertheless, by focusing on each policy document separately, it is obvious that although there are a few cases where the algorithm produced connections to the related SDGs, none was found through the classical method of scanning the policy text and scoring its relativity to the SDGs. But we must have in mind that, since the algorithm is incapable of semantic capabilities, the results are based purely on the frequency of terms appearing in each policy document. This explains why no policy was classified with a 0,00-similarity score in Table 6. Results that may seem quite peculiar are the strong connection between all policies with SDG 1 and SDG 6. These discrepancies may occur again due to the words used to form the vocabularies describing SDG 1 and SDG 6 and their frequency in the examined policy documents. In Conclusion, the 1st algorithmic approach performs well in capturing the overall connection between policies and the SDGs, but not so well in identifying which SDGs are mostly related to the documents under examination. As a result, the 1st algorithm could be used for a quick first glimpse of the overall relativeness between SDGs and the policy documents.

On the other hand, the second algorithmic approach, which utilizes deep learning techniques to calculate the similarity between SDGs and energy policy documents, performs much better. Overall, its results seem to agree with the “classical approach”, but, unlike both the latter and the 1st algorithmic approach, they provide a better insight on scoring the identified relationship. This scoring can be of great use, for a quick but adequate evaluation of policy documents, thus providing a useful tool for highlighting the interlinkages between Policy and Strategy documents, like the ones examined here, and the SDGs. Moreover, what seems quite interesting, is the fact that some energy-related policy documents are somewhat relevant to SDG 6. This correlation can be also found in the results of the first algorithmic approach but not with the “classical approach”. This means that there might be a connection between some policies and “Clean Water and

Sanitation” that one must look more closely to find out. After all, one of the advantages of utilizing Machine Learning is the fact that they quickly identify insights and correlations, not easily observed by the human eye. For instance, SDG 6 (Clean Water and Sanitation) targets safe drinking for all, therefore it is necessary to review the progress in Water-Use efficiency and the reduction of water-stress levels. These 2 important indicators are dependent on the maintenance of the water distribution systems and the investments in innovative technologies within the industrial, manufacturing and agricultural systems. Moreover, water-use efficiency and savings are highly associated with energy savings as well, due to the water-use intensity in the energy systems. Similarly, relief of water stress is essential for the activities associated with Energy- Agriculture- Food Industry Nexus. To sum up, there are real technical initiatives that are proving hidden relationships and implicit correlations between Policies and SDG’s and ML utilizing can support this in such accuracy, as the correspondent technical libraries can be embedded within the ML approach and process.

On the other hand, similarity scores with SDG 1 were low, confirming the results of the classical method. Finally, one must highlight the fact that policies are somewhat connected to SDG 16 and 17, which were the two goals that had the fewest labelled text excerpts used for training, showing that without “Peace. Justice and strong institutions” and without building “Partnerships for the Goals” little can be achieved.

Conclusions - Ideas for further research

The presented work gives oversight on how the main Energy Policies of the European Union connect to the 17 SDGs and roughly answers, in short, the question "Where is the EU heading, considering the announcements and the promises made for a sustainable future". The work included three different methodologies targeting grading the relativity of each policy with each SDG.

The first methodology, referred to as the "classical approach", included the reading of each document and grading its SDG-relativity using a scale of 0-3, where 0 stands for "no relativity" and 3 for "great relativity". The second methodology included using BoW Techniques (Information Retrieval) to compare the similarity of any given document to a vocabulary of keywords, formatted by the research group. Each document was preprocessed and transformed into a vector. The document-vector and each vocabulary-vector were compared using the Cosine Similarity and their similarity scores were produced. The third methodology, included the fine-tuning of a deep learning BERT model, using more than 15,000 labelled texts in total.

By utilizing Natural Language Processing techniques one can find a quick and adequate correlation between the examined policy documents and the 17 SDGs, but further examination is needed to validate the results produced. A clearer view with better accuracy is given after utilizing Deep Learning Techniques with semantic capabilities, which produced insights that need further investigation to be confirmed.

Overall, this study utilized Natural Language Processing Techniques and Machine Learning to calculate the interconnectivity of policy documents and SDGs, setting a concrete work that can be used in future research.

For instance, one could further improve the results by including more labelled texts in the training data set, especially for Goals SDG 12, 16 and 17. Furthermore, considering the interconnectivity and relativity of the 17 SDGs with each other, one could use multi-labelling and voting ensembles, to get better results for each scanned document, or deploy a Neural Information Retrieval Model for better queries. This work could help in utilizing Machine Learning in the Six Transformational Pathways towards Sustainability, that base upon the interconnectivity of the SDGs.

Additionally, to improve the public dissemination and understanding of the connection between policies and the SDGs, a web platform could be set up, open to anyone interested in finding more information regarding a policy document and its connection to the SDGs. The platform could host the aforementioned developed tools, that would be regularly maintained and upgraded to provide more accurate results from time to time.

The classical approach will be still more popular since human working groups will prefer with no doubt to discuss and exchange ideas or even have arguments about the relevance between policy documents and ideas or commonly agreed targets (such as SDG's). But this concrete and the fair process can be supported in the future by ML or similar tools since these tools can be upgraded, to mention and remind explicit and implicit relationships between policies and targets. For instance, an upgraded ML tool, as it presents the result of relevance, can include also a short description, proving and giving feedback (by a short example) why they give a "point 3 or point 2" relevance between the x SDG and the A Policy, presenting a short text when computer censor is over this specific score assessment. This means that the algorithm should exploit all possible correlations between crucial document keywords and synonymous ideas that are leading to specific Targets and their semantic relations.

References

Chowdhury, G.G. (2010), Introduction to Modern Information Retrieval, London: Facet Publishing

Eisenstein J., “Introduction to Natural Language Processing”, 2019, MIT Press, Cambridge: United States

European Commission, 10 March 2020, A New Industrial Strategy for Europe (COM(2020) 102 final), Brussels https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0102&from=EN

European Commission, 11 December 2019, The European Green Deal (COM/2019/640 final), Brussels https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1576150542719&uri=COM%3A2019%3A640%3AFIN

European Commission, 14 October 2020, A Renovation Wave for Europe - greening our buildings, creating jobs, improving lives (COM(2020) 662 final), Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0662&from=EN

European Commission, 14 October 2020, Chemicals Strategy for Sustainability (COM(2020) 667 final), Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0667

European Commission, 14 October 2020, on an EU strategy to reduce methane emissions (COM(2020) 663 final), Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0663

European Commission, 17 September 2020, Annual Sustainable Growth Strategy 2021 (COM(2020) 575 final), Brussels https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0575&from=en

European Commission, 19 November 2020, An EU Strategy to harness the potential of offshore renewable energy for a climate neutral future (COM(2020) 741 final), Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0741&from=EN

European Commission, 5 May 2021, Updating the 2020 New Industrial Strategy: Building a stronger Single Market for Europe’s recovery (COM(2021) 350 final), Brussels

European Commission, 8 July 2020, A hydrogen strategy for a climate-neutral Europe (COM(2020) 301 final), Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0301&qid=1641991695109&from=EN

European Commission, 9 December 2020, Sustainable and Smart Mobility Strategy – putting European transport on track for the future (COM(2020) 789 final), Brussels, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0789&from=EN

European Commission, October 2020, Official Journal of the European Union (L 357/35), COMMISSION RECOMMENDATION (EU) 2020/1563 on energy poverty, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32020H1563&from=EN

Goodfellow I., Bengio Y., Courville A. (2016). Deep Learning. MIT Press. Available at: www.deeplearningbook.org

LeCun Y., Bengio Y and Hinton G. (2015), Deep Learning, Nature 521, 436-444, DOI: https://doi.org/10.1038/nature14539

OSDG, UNDP IICPSD SDG AI Lab, & PPMI. (2021). OSDG Community Dataset (OSDG-CD) (2021.09) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5550238

Passalis, N., & Tefas, A. (2018). Learning bag-of-embedded-words representations for textual information retrieval. Pattern Recognit., 81, 254-267

Sachs, J., Koundouri, P., et al., 2021. Transformations for the Joint Implementation of Agenda 2030 for Sustainable Development and the European Green Deal - A Green and Digital, Job-Based and Inclusive Recovery from the COVID-19 Pandemic. Report of the UN Sustainable Development Solutions Network, https://resources.unsdsn.org/transformations-for-the-joint-implementation-of-agenda-2030-the-sustainable-development-goals-and-the-european-green-deal-a-green-and-digital-job-based-and-inclusive-recovery-from-covid-19-pandemic

Sachs, J.D., Schmidt-Traub, G., Mazzucato, M. et al. Six Transformations to achieve the Sustainable Development Goals. Nat Sustain 2, 805–814 (2019). https://doi.org/10.1038/s41893-019-0352-9

United Nations, October 2015, Transforming our world: the 2030 Agenda for Sustainable Development (A/RES/70/1), https://sdgs.un.org/2030agenda

Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. ArXiv, abs/1706.03762.

von der Leyen U., 2019, Political guidelines for the next Commission (2019-2024) - "A Union that strives for more: My agenda for Europe", https://ec.europa.eu/info/files/political-guidelines-new-commission_en

Wang H., Ma C. and Zhou L., "A Brief Review of Machine Learning and Its Application," 2009 International Conference on Information Engineering and Computer Science, 2009, pp. 1-4, doi: 10.1109/ICIECS.2009.5362936

World Resources Institute (By Mengpin Ge, Johannes Friedrich and Leandro Vigna), February 6, 2020, 4 Charts Explain Greenhouse Gas Emissions by Countries and Sectors, https://www.wri.org/insights/4-charts-explain-greenhouse-gas-emissions-countries-and-sectors

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K. & others (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144, .

Zellig S. H. (1954), Distributional Structure, WORD, 10:2-3, 146-162, DOI: 10.1080/00437956.1954.11659520