UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M. FANNO"

CORSO DI LAUREA MAGISTRALE IN
Management for Sustainable Firms

TESI DI LAUREA

"AI FOR SUSTAINABILITY: A PATENT-BASED ANALYSIS AND REGIONAL INNOVATION PATTERNS"

RELATORE:

CH.MO PROF. Andrea Ganzaroli

LAUREANDA: Elena Vavilova

MATRICOLA N. 2106165

ANNO ACCADEMICO 2024 – 2025

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

Elena Vavilova

Table of Contents

List of Figures

List of Tables

1. Introduction

Nowadays, there are multiple crises in economics, the environment, and society that change our world in many aspects. Economic growth is expected to slow from 2.9% in 2024 to 2.3% in 2025, Economic Policy Uncertainty Index in January 2025 had the highest value, and in many regions across the world the economy is entering a recession phase *("Global economy under pressure could slow to 2.3%, signals UN Trade and Development," 2025*). Natural resources are becoming increasingly scarce and costly due to the significant rise in demand for food, energy, water, land, and minerals (*Foresight, n.d.*). The problem of CO2 emissions is becoming increasingly urgent, as it has risen over the last 50 years (*Ritchie et al., 2023*). All the aspects show that there is a need for a big change in order to improve the situation. Twin transition and development of smart cities, which combine both digitalization and sustainability, are trying to solve the problem and create a more sustainable future.

Sustainable Development Goals represent the key problems and goals that are vital for both developing and developed countries. According to *"Key findings: Exploring the SDGs through patents," 2024*, the number of patents related to Sustainable Development Goals (SDGs) is increasing. Trend is highly connected with the critical situation in the environment and society.

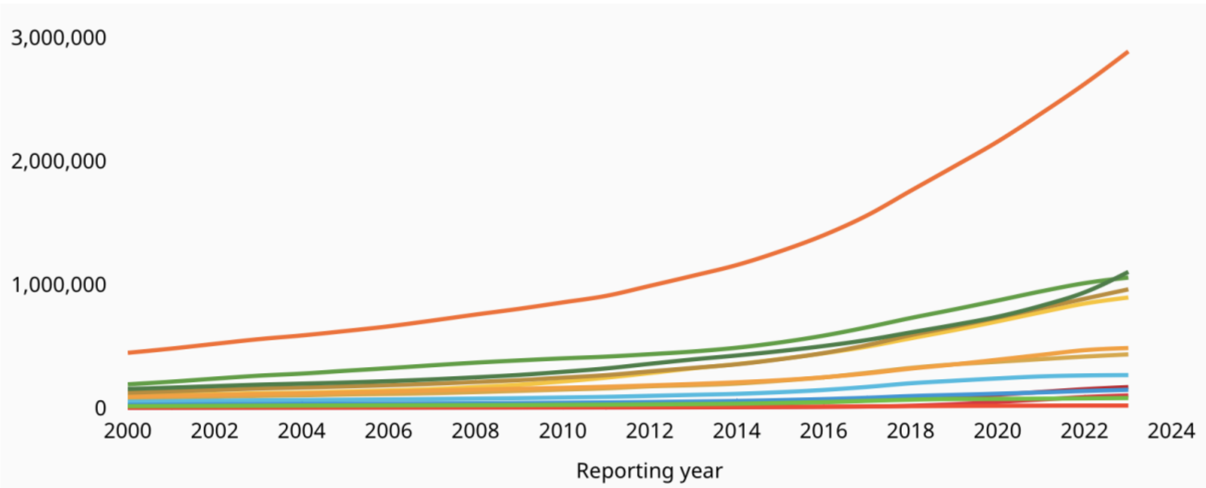

Figure 1. Active Patent Families Related to SDGs from 2020 to 2023

According to the provided statistics, the trend is rising for most of the SDGs, however, the most drastic change can be seen for SDG 9 (Industry, Innovation, and Infrastructure), SDG 13 (Climate Action), SDG 3 (Good Health and Well-Being), SDG 12 (Responsible Consumption and Production), and SDG 7 (Affordable and Clean Energy).

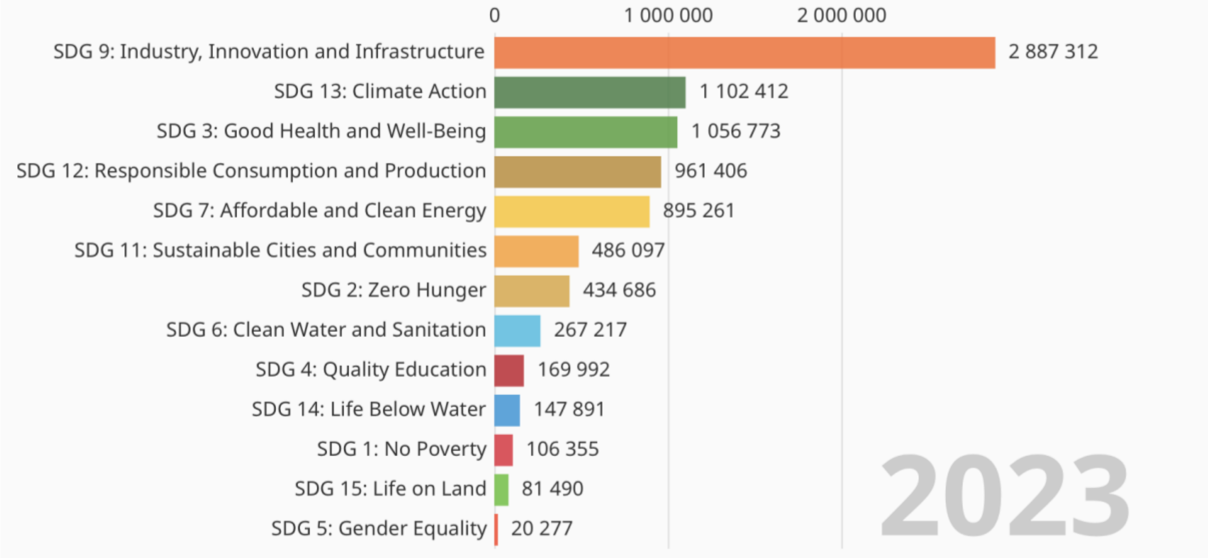

Figure 2. Patents Related to SDGs in 2023

Another big change that can be noticed is the use of AI. Currently, it is already applied in various fields, and the trend is only rising. The AI market size is growing every year and, according to *Statista, 2025*, it is expected to grow even more rapidly.

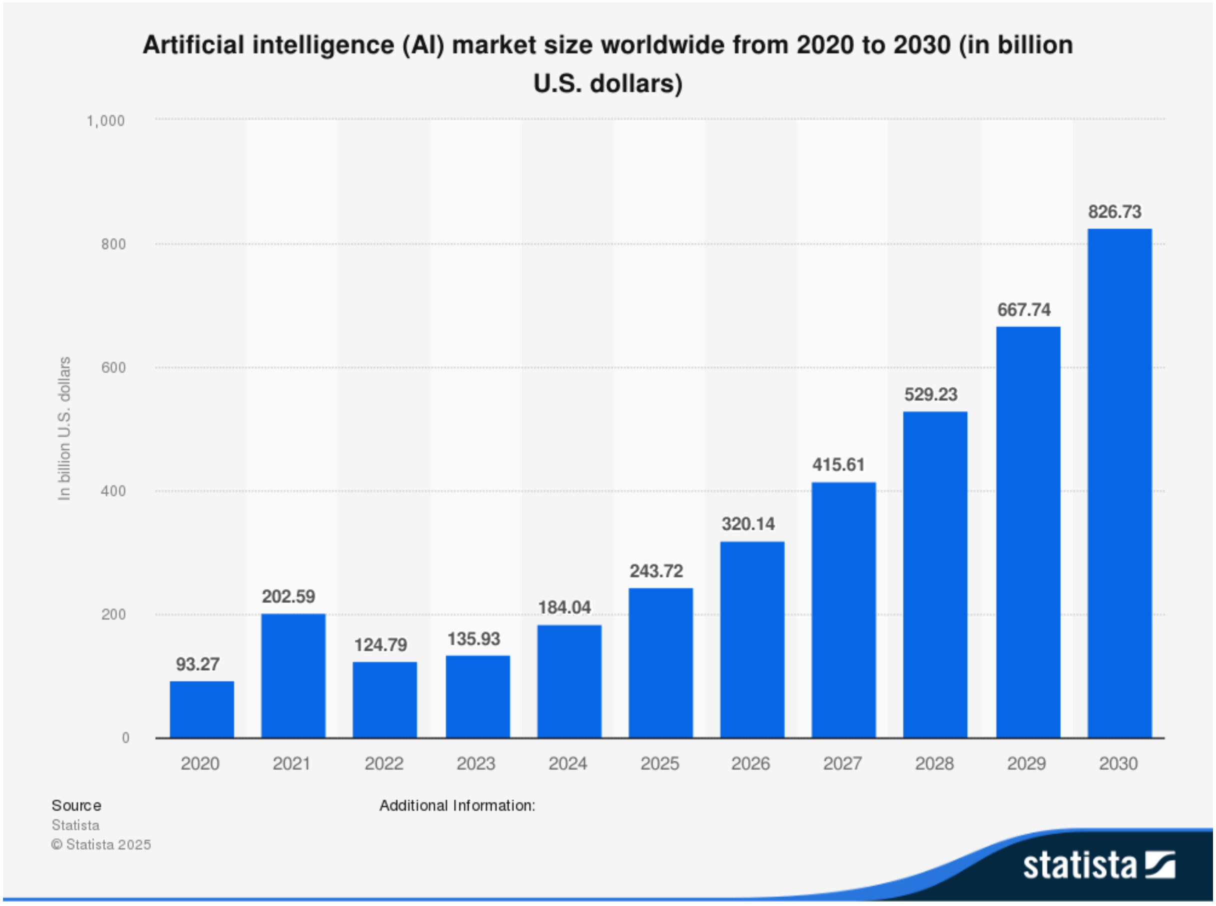

Figure 3. AI Market Size Worldwide from 2020 to 2030

As can be seen from the graph, by 2030 the market size is expected to be almost 10 times bigger than in 2020. Both sustainability and AI are becoming a big part not only for people and businesses, but also for research and innovations. More and more innovations are related to sustainability, and in many cases, the new technologies, such as artificial intelligence and machine learning, are used.

Most studies show that there is a connection between artificial intelligence and sustainability: AI and new technologies can be applied for sustainable development. At the same time, there are several risks that appear because of the high usage of AI. The most common issues are related to environmental and ethical aspects. It can be said that artificial intelligence has two sides: while there are clear benefits to its applications, there are also potential hazards. Another important factor that is discussed in the thesis is to what extent regional collaboration and the presence of science and technology parks can foster innovations.

Accordingly, based on the theoretical knowledge and previous studies, the main research question of this thesis combines AI and sustainability, is defined as follows: "How is artificial intelligence connected with sustainability in innovations, and how can patent data be used to identify trends and regional specialization, and collaborations?"

The aim of the research is to define the contribution of countries and regions to sustainable development with the help of artificial intelligence, based on patent information. The leaders among European regions and separate applicants will be defined. Based on the provided data, it is expected to find already existing regional collaborations and clusters, the most common technological classes, and technological clusters. Results of the analysis can become a baseline to define possible competitors or partners for the regions based on the level of development of separate technological sections.

Current research uses a mixed approach: qualitative research design is supported by qualitative analysis for a deeper understanding. The research is based on the secondary data collected from two databases that contain information about patents, Orbit and REGPAT. Usage of different data sources is beneficial for getting more complete information. Patents are selected by filtering keywords and IPC or CPC codes. Besides databases, statistical information was used for additional information about relevant regions. In the qualitative research, information from reports, publications, and official websites was collected. Various data analysis techniques were used in the research: descriptive statistics, visualization, regional co-occurrence, RTA, and collaboration analysis. There are also several limitations that have appeared during the research.

The structure of the thesis was designed to support the steps that were necessary. The parts outline the progression of the work, which started from the theoretical part, then the methodology for the practical part was selected. Based on these two parts, an analysis of the data was conducted. The final part includes the main findings from the previous steps. To support the material, the main algorithms are included in the report. A brief overview of the main chapters is presented below:

- **Literature Reviews:** Existing literature about sustainability and artificial intelligence will be analyzed. Afterwards, the topics of AI as an enabler for sustainability, risks of AI and possible solutions, collaboration, and spillover effect will be covered. Since the practical part is based on the patent analysis, the different ways in which patent data can be used for analyzing innovations will be provided.
- **Methodology:** In this chapter, all the necessary information about classifications and tools will be given. The practical part includes an analysis of quantitative information, and several formulas that are used will be explained in detail. Additionally, methodology contains information about limitations that influenced the research.
- **Presentation of the Results and Data Analysis:** This is the first practical part, which will concentrate on descriptive statistics and calculations to define the most active regions, applicants, find collaborations, and the strongest technological classes. Basic statistics will be provided for countries worldwide, and then a deeper analysis of the European regions will be conducted. Besides that, the part will cover the analysis of the most outstanding regions to find the reasons for the great performance of the selected regions in Europe.
- **Conclusion:** The conclusion provides a brief overview of the conducted research, including findings from the literature review and data analysis. Possible applications of the results will also be presented in this chapter.
- **Appendix:** This part will contain the main parts of the code that were used for the practical part. All the uploads, imports of modules, or simple commands will be omitted.

2. Literature review

In this chapter, the main definitions related to sustainability and artificial intelligence will be covered, and the possible benefits and threats of AI in sustainability will be examined. Since the research is based on the patent analysis, the literature review will analyze how patent data can be used for measuring the impact. Practical application of the findings will be presented in the analytical part.

2.1 Artificial Intelligence and Its Applications

According to the publication of *High-Level Expert Group on Artificial Intelligence, 2019*, AI systems can be defined as the software or hardware that can act both in physical and digital dimensions by taking the steps such as: data collection and interpretation, reasoning of knowledge, processing the information obtained from the data, and choosing the action that is the most suitable for the defined goal. AI is considered to behave in an intelligent (rational) way. Some of the more advanced AI systems (learning rational systems) have good adaptability: not only do AI act based on the initial analysis, but also adjust their behavior based on the reaction of the environment to the previously taken actions.

There are six subfields or components of AI, provided by *Khan, 2024*:

- **Cognitive computing** enables the simulation of human thought processes, as well as assisting people in the decision-making process and problem-solving. This subfield has several goals: improving human-computer interaction in terms of reasonability and naturality; acquiring of new knowledge for the continuous learning; solving advanced problem where multiple factors have to be considered; improving adaptability of the system to keep it efficient and effective; supporting decision-making processes for more informed and optimized decision; ensuring that the system keeps ethical and responsible manner; fostering collaboration of different disciplines for better human understanding. Cognitive computing can be combined with machine learning to improve the abilities and adapt to situations faster.
- **Computer vision** enables the understanding of visual information, such as videos and photos, in order to make decisions based on the extracted information. Computer vision can help to extract the particular features, for example shapes and colors, recognize patterns that were learned previously, detect objects, and classify the visual information. A combination of computer vision and deep learning creates the opportunity to learn from multiple examples.
- **Machine learning** helps to develop systems that can learn on a regular basis and improve themselves based on the learning process. Machine learning is commonly used

for data analysis, predictions, classification, and decision-making. There are common steps that are necessary for the successful learning process: data collection, data preparation, training, learning, testing, and improvement.

- **Neural networks** can be considered as a fundamental concept. It is based on the functioning of the human brain and is frequently used in neurology, since it can mimic the human brain: various algorithms allow to categorize data and capture the relationship between the objects. There are multiple objectives and applications of neural networks: pattern recognition, data transformation, making predictions, data classification, making decisions, generalization for broader utility, and adaptation.
- **Deep learning** is meant to make use of the neural network's capabilities for working with complex data. Deep learning can allow to work with a lot of data from different sources in order to imitate the human decision-making process. Due to its advancement, deep learning is commonly used in complex tasks, such as voice commands on smartphones and self-driving cars.
- **Natural language processing (NLP)** is created to work with human language, specifically with recognition, analysis, interpretation, and general understanding. The most common applications of NLP are sentiment analysis, translation, chatbots, and summarizing texts.

All the subfields of AI can be widely applied in multiple industries, and currently, the number of sectors is getting higher. AI is used not only for business purposes, but also for science, which can be seen in *"Researchers and innovators invited to shape Europe's AI Strategy in Science," 2025*. Currently, there is a new initiative in the EU called "The European Strategy for Artificial intelligence (AI) in Science". It is created to foster scientific research by helping scientists to adopt AI for these purposes. This initiative is highly applied in the key areas, such as climate change, health, clean technologies, astronomy, medicine, etc. The European Strategy for AI is expected to drive innovation in the EU and attract talented researchers.

2.2 Sustainability and SDG Goals

"Cambridge English Dictionary: Meanings & Definitions," 2025 defines sustainability as the quality of causing little or no damage to the environment and therefore being able to continue for a long time. Based on this, it can be said that sustainable development has to meet the needs of the present generation but without depriving future generations of the opportunity to meet their needs.

Three pillars are most commonly used when discussing sustainability: environmental, economic, and social. These aspects do not exist separately – they interact with each other, and the intersection of all three pillars can be considered as "sustainability":

- **Environmental sustainability** refers to the protection of the natural environment and ecological balance on the planet. Several factors are affecting the environmental situation, such as pollution, extensive natural resource consumption, and losses of biodiversity.
- **Economic sustainability** concentrates on providing well-being through economic growth and innovation activities while avoiding violating the abilities of the next generations. Economic sustainability can be influenced by the level of financial stability, corporate responsibility, and collaboration between the public and private sectors.
- **Social sustainability** includes the support of the well-being of people and communities through access to basic needs, human rights, development, healthcare, and education. Several obstacles can prevent achieving social sustainability, such as poverty, conflicts, inefficiency of governmental institutions and corruption, and discrimination.

Despite the broad usage of the "three pillars" concept, currently, there are more modern and complex ways to describe sustainability. According to *Purvis et al., 2018*, one of the contemporary concepts is the UN's sustainable development goals (SDGs). In 2015, The 2030 Agenda for Sustainable Development was adopted by all the UN members. The central part of the Agenda is seventeen goals – SDGs, that are considered to be crucial in society *("THE 17 GOALS | Sustainable Development," n.d.*). SDGs emphasize the necessity for both developed and developing countries to work together. All the sustainable development goals are presented in the picture below:

Figure 4. Sustainable Development Goals

The approach of having not three aspects, but more, helps to show a more integrated approach and emphasize the critical issues that currently exist. A complex framework allows to show the issues common to both developed and developing countries. Such representation of the global problems makes the situation clearer and more defined. There are many approaches to combat the global goals at the state, organizational, and personal levels. Moreover, many R&D activities are currently addressing global issues. As it was mentioned earlier, more and more patents are related to SDGs.

2.3 AI as an Enabler of Sustainability

Artificial intelligence has traits of GPT (*Martinelli et al., 2021*), and one of the areas where it has started to be applied is sustainable development. In the research of *Biggi et al., 2025*, the role of AI as an enabler for sustainable development is discussed in a detailed way. First of all, the application of AI is relatively new, which can become a reason for a lower positive response in the short term due to not enough evidence of the positive impact. Nevertheless, the research shows that the influence and usage of AI are getting broader and can be used for the stimulation of green innovations. Green innovations include various aspects, which show that artificial intelligence has the possibility of being applied in different ways in order to tackle the environmental and economic issues.

Another recent research provided by *Tao, 2024* explores if artificial intelligence has an effect on green productivity and how big the influence is. The data from fifty-nine countries in the period from 2008 to 2019 was chosen. The author highlights that AI has a solid base for sustainable development. As it was indicated before, AI can be applied in various ways, but the research highlights three factors that can be seen as the most important mediators: attraction of high-skilled labor, acceleration of renewable consumption, and weakening stock market. Besides the role of AI in sustainable development, the research highlights the necessity to have additional regulations at the governmental level to increase commitment to sustainability.

Examination of the relationship between AI adoption and the efficiency of green innovations was discussed by *Feng et al., 2024*. The researchers concentrate not only on the absolute volume of green innovations but also on the efficiency of such innovations in China. Efficiency is calculated by finding the ratio of output and input: output is presented as the logarithm of all green patent applications, while input is shown by the R&D expenditures. The results of the research highlight the strong positive impact of AI on the outcomes of sustainable innovations. The findings show that the usage of artificial intelligence combined with absorptive capacity, innovation capability, and adaptability of the company has a great contribution to sustainable innovations.

Research presented by *Yadav et al., 2024*, shows examples of how AI can be applied in order to achieve the Sustainable Development Goals. One of the most important advantages of AI is its ability to process big data and various types of data, from pictures to information obtained from sensors. AI can be used for forecasting because of its advanced capabilities. Moreover, advanced AI algorithms are used for better efficiency and resource optimization, which can, for instance, lead to waste reduction and optimization of transportation networks. The author defines several applications to tackle SDGs:

- **SDG 1 (No Poverty):** Application of AI can be used to create microfinance systems, analyze and visualize how poverty is distributed in particular regions for better targeting. A decision-making algorithm similar to a scoring system can be used for an automated check for social welfare programs eligibility.
- **SDG 2 (Zero Hunger):** The usage of artificial intelligence for agriculture and the food industry can be used to optimize agriculture practices, make better predictions for yield, and reduce poor nutrition and food scarcity.
- **SDG 3 (Good Health and Well-being):** AI can be used to diagnose and detect diseases, especially at early stages, for preventive medicine and early intervention. New algorithms can also be beneficial in the stage of development of new drugs. Centralized information about patients combined with AI can be the basis for personalized recommendations and health tracking accessible to people.

- **SDG 4 (Quality Education):** AI and new technologies can help to achieve better education by creating personalized learning experiences and making the resources available for a wider range of people, adapted for individuals from different backgrounds and with different abilities.
- **SDG 5 (Gender Equality):** Data bias, which poses a threat of discrimination, can be noticed while using AI. Artificial intelligence can also be used for creating policies and preventing violence based on gender discrimination.
- **SDG 6 (Clean Water and Sanitation):** Water monitoring systems, water conservation, and distribution can be powered by AI in order to avoid water leaks, poor quality, and overconsumption.
- **SDG 7 (Affordable and Clean Energy):** AI can be used for predictive maintenance of energy infrastructures, demand forecasting and optimization, smart distribution and management.
- **SDG 8 (Decent Work and Economic Growth):** Platforms supported by AI can be beneficial for a better match between job-seekers and employers. Predictive analytics can be a source of valuable information about potential skill development. AI-powered models in finance can be used for forecasting and planning.
- **SDG 9 (Industry, Innovation, Technology and Infrastructure):** Predictive maintenance is one of the important topics for many industries: it can help to predict and prevent critical issues. Moreover, AI can be used to improve supply chains in terms of efficiency and optimization.
- **SDG 10 (Reduced Inequalities):** Artificial intelligence can be used to analyze the data and use the results of the analysis for the creation of better and more inclusive policies. AI can help with defining possible inequality and discrimination in society.
- **SDG 11 (Sustainable Cities and Communities):** Many aspects of smart cities are based on artificial intelligence and digitalization of the systems. A more detailed overview will be presented later.
- **SDG 12 (Responsible Consumption and Production):** Usage of AI can increase transparency of supply chain, help with smart resource management to reduce waste and overconsumption.
- **SDG 13 (Climate Action):** AI is broadly used for optimization, including optimization of the resources that can foster the reduction of greenhouse gas emissions.
- **SDG 14 (Life Below Water):** Artificial intelligence can provide an opportunity to have autonomous underwater vehicles and devices to explore the ocean, monitor the marine ecosystem, and help with conservation.

- **SDG 15 (Life on Land):** AI-powered algorithms can be used to detect areas affected by deforestation using the analysis of satellite pictures, and predict possible ecosystem and biodiversity issues.
- **SDG 16 (Peace, Justice and Strong Institutions):** AI can assist the decision-making process, detect human rights violations, and prevent crimes.
- **SDG 17 (Partnerships for the Goals):** Artificial intelligence is beneficial for defining potential partnerships and improving collaborations and engagement via platforms for global initiatives.

The examples mentioned above represent how powerful artificial intelligence can be in specific sectors. At the same time, the concept of a smart city can represent how AI can be used in different fields for urban development and sustainability. In *"Smart cities," n.d.*, the European Commission indicates that smart cities use digital solutions for the benefit of both businesses and inhabitants. The concept shows how advanced technologies are used to tackle such issues, as optimization of transportation systems, achievement of energy efficiency, improvement of public safety, and improvement of healthcare. The research of *McKinsey Global Institute, 2018* provides examples of how technologies used in smart cities can improve the quality of life from different perspectives:

- **Public Safety and Crimes:** Even though currently technologies are not advanced enough to fix crimes fast enough, they can be used as a preventive tool. Artificial intelligence can help with predictive policing and real-time crime mapping, which can influence the rate of burglaries, homicides, fires, and road traffic. Systems based on AI can optimize the work of the call centers, which will lead to faster responses. Real-time traffic detection can be beneficial not only for reducing the number of car accidents, but also for optimizing the routes of emergency vehicles to get to the place faster.
- **Transportation:** As it was mentioned before, AI can be used to detect traffic in real life, which can let people spend less time on commuting. Besides that, artificial intelligence is one of the tools for predictive maintenance and root optimization, which is important for the safety, reduction of delays, as well as for reducing emissions.
- **Health System:** Technologies can provide the necessary tools to improve the health systems in order to help inhabitants have long, productive, and healthy lives, and telemedicine is one of the examples. It allows for monitoring and consulting patients remotely. Another benefit of artificial intelligence is data analysis: data can be used to identify groups with higher risks, and then use the results for various preventive actions, such as reminders and monitoring.

- **Improvement of Environment:** Automation systems used both in commercial buildings and households can help to reduce electricity consumption and emissions. Air quality sensors, which can identify the sources of pollution, help to work on protective and preventive actions. Water sensors can not only detect leakages from the pipes, which cause enormous water losses, but also empower people to change their own behavior in case of overconsumption.

The report emphasizes that smart cities have the potential to reach up to seventy percent of the Sustainable Development Goals by using technologies, which highlights how AI can enable sustainability.

2.4 Risks of AI and Possible Solutions

AI is becoming an important part of households, companies, and innovation activities. On the one hand, as it was discussed in the previous part, AI has a positive impact on sustainable development, while on the other hand, there are a lot of concerns regarding the potential risks.

Research by Yadav et al., 2024 which was mentioned before, provides not only the applications of artificial intelligence for sustainability but also potential threats. The first problem related to AI is data bias: as it was explained before, models are learning and training based on the data, and if the data is initially biased, incomplete, or has poor quality, it can influence the results. Data bias, which is cited as one of the most common ethical concerns, can cause discrimination against some groups of people because the data provided for algorithms was initially not correct or complete. Another AI-related threat that is commonly mentioned is job displacement: artificial intelligence and new technologies are used for the automation of processes, which leads to the reduction of people needed for specific tasks. While AI can improve efficiency, it can cause higher inequality in the society, higher unemployment rates, and poverty. While automation can help to provide the productivity needed to achieve future economic growth, at the same time, it can cause job displacement problems. According to *McKinsey Global Institute, 2017*, even though less than five percent of all jobs are fully automated, for sixty percent of jobs automation level is around thirty percent.

Research by *Stahl, 2021* covers other ethical issues related to AI, including data protection and information privacy problems, covered in a detailed way. Deeper insights that can be found with the help of artificial intelligence require better models and more data for training. Finding patterns using artificial intelligence raises the issue of privacy, since information can be found even without direct access to personal data. Besides the problems mentioned above, there are cybersecurity concerns. With the development of the systems, new types of cyberattacks are also becoming more advanced, which means that systems are becoming increasingly vulnerable.

"United Nations Environment Programme (2024) Artificial Intelligence (AI) end-to-end: The Environmental Impact of the Full AI Lifecycle Needs to be Comprehensively Assessed - Issue Note.," n.d. provides information about possible environmental issues that can be caused by AI. First, the fast growth of AI and its computational power requires the ability to process more data and use more advanced models, which leads to the need for more resources. The increase in resource usage can cause more greenhouse gas emissions, pollution, overconsumption of minerals, energy, and water. Second, to work with more complex problems, there is a need for better hardware and larger data centers. Not only are more resources needed for the production, but also the problem of e-waste and maintenance. The research also distinguishes between direct and indirect effects of AI. In order to tackle all the mentioned threats, the note provides recommendations that can help to evaluate the impact and prevent the negative effects of AI. The possible actions include: standardization of impact measurement (with the focus on direct effects first), development of the frameworks for reporting about direct effect of AI for companies, empowerment of end users to contribute to the reduction of environmental impact, concentrating on the research about algorithms' optimization, usage of renewable energy sources, make a deeper research on potential effects.

Besides the proposed actions, which are not mandatory but recommended, there are official regulations regarding AI, such as the EU AI Act, which entered into force on 1 August 2024 (*"Regulation - EU - 2024/1689 - EN - EUR-LEX," n.d.*). The Act represents the first legal framework for AI on a worldwide level. The main goal of the regulation is to encourage innovation and AI development while building trust in AI by providing protection from possible risks. The Act defines different levels of AI risks, from minimal to unacceptable. The regulation is expected to be fully applicable with several exceptions, in two years from the indicated date, on 2 August 2026. Even though the EU AI Act makes a contribution to risk mitigation, it still highlights that some specific challenges are not addressed.

To sum up, currently, artificial intelligence is still relatively new and complex. The rapid increase of AI usage, fast development, and different applications cause concerns regarding safety, environmental, and ethical problems that can appear. All the mentioned resources define both benefits and threats caused by artificial intelligence, which leads to the need for having regulations and methods to mitigate the risks. The EU AI Act is the starting point of AI regulations, but it is still in the initial stage, and the effect of the implementation can be seen after some time.

2.5 Combination of AI and Sustainability: Twin Transition

Twin transition, or dual transition, is an example of how artificial intelligence and sustainability can be combined. Twin transition is also considered to be crucial for smart cities. The report of *Muench et al., 2022* emphasizes the importance of the twin transition in Europe and defines what are the elements of successful transformation. Even though the concept of dual transition includes simultaneity, it does not imply interconnectedness. As it was mentioned before, there are several problems that can be caused by artificial intelligence.

Digital technologies provide exceptional tools that can foster and support the green transition, such as better monitoring systems, analytical tools, visualization, simulation, and forecasting. The study mentions that the success depends not only on the technologies, but also on policies, consciousness, and awareness among people. Based on these issues, the key success factors that are vital for the transition were defined for different areas:

- **Social Requirements:** First, the development and innovations have to be beneficial for all people in the society to avoid any discrimination. There is a need to raise awareness and inclusivity to show the necessity of behavioral change: people have to understand the importance of their actions, and how the transformation can benefit the society. Since the personal data is needed for the research and analysis, it is important to provide necessary protection and anonymization of the information, and avoid the gathering of unnecessary data.
- **Technological Requirements:** Well-developed research ecosystems are one of the key success factors since they provide the necessary equipment and environment that can favor innovation. Technologies have to be appropriate and reliable for the development, and data governance is required for data security.
- **Environmental Requirements:** The entire lifecycle of green-digital technologies has to be tracked to evaluate the resource consumption and reduce the environmental footprint. Governance and market mechanisms have to be adapted to favor the diffusion of green and digital technologies.
- **Economic Requirements:** There is a need to have a diverse market: instead of having monopolies or oligopolies, companies of all sizes have to be present to ensure competition and innovation. Employees have to be provided with special training to be able to handle new green-digital technologies. To incentivize the transition towards green-digital solutions, markets have to take the environmental costs since especially in the beginning, the new solutions might not be profitable.

- **Political Requirements:** Policies and standards have to support the low entry barriers, be aligned with long-term goals and other regulations to have a stable and consistent framework, and create conditions for higher investments in green-digital technologies.

One of the factors that is not always taken into account while talking about twin transition is the regional heterogeneity. Different regional peculiarities can affect the trajectory and success of the twin transition (*Faggian et al., 2024*). The results of the research show that the advancement of the region highly depends on the scientific capabilities and foundations, especially in green and digital technologies. Success in these technologies individually creates success in the transformation that combines both aspects. Another factor that affects the twin transition is the type of region, because rural and urban areas have different prerequisites. Rural areas are more likely to succeed if they are closer to the natural resources, but for urban areas, local innovation ecosystems that include companies, institutions, people, and relevant activities are more important to foster innovation and twin transition. The study emphasizes the necessity to create policies that will be adapted to the specific region based on the gaps and strengths.

2.6 Collaborations and Knowledge Spillovers

One of the aspects that can affect the innovation intensity is collaboration across regions and firms. Several studies examine the effect of the collaborations and the possible spillover effect.

The research of *Fan et al., 2019* provides the analysis of Chinese cities. The paper gives an overview of the impact that different types of collaborations, intra-regional (within the same region) and inter-regional (across different regions), can have on overall innovation intensity. The research highlights that collaborations of strong innovators within the same region have overall a high impact not only on the local efficiency, but also have a spillover effect to the neighbouring regions. The level of intra-regional collaboration depends on the connections of firms, universities, and research institutions. As for the inter-regional joint efforts, it still has an effect, but weaker than in the case of intra-regional connection. The importance of collaborations across different regions is explained by the overall diffusion of innovations.

Another research study on how innovation efficiency in a region is influenced by external knowledge obtained through a collaboration network (*Hazir et al., 2014*). For the research, patenting activity in ICT field was examined. The findings highlighted that both for short and long-distance connections there is a positive effect of spillover, since regions have an opportunity to learn from each other. According to the researchers, knowledge slipover that happens through collaborations that are supported by the European Union’s Framework Programs for Research and Technological Development (FP) has a significant effect in the patenting activity of the regions. Additionally, if the region i has more interactions and stronger connection with region j, then it is more likely to get more benefits from the spillovers from the other region. Such a connection shows that there is a necessity to have more strategically organized innovation systems.

Collaborations appear not only across regions, but also in science and technology parks. The effect of knowledge spillover in STPs was examined by *Montoro-Sánchez et al., 2011*. The results of the empirical studies show that the presence of the company in a science and technology park gives it an advantage because of the network and habitat that fosters innovations. In general, STPs help to stimulate the knowledge spillover and technology diffusion among businesses, universities, and research institutions. Besides the spillover effect, the habitat is favoring the establishment of collaborations among the residents.

In conclusion, all of the provided studies highlight the positive effect of collaboration and the spillover effect, both intra-regional and inter-regional. Science and technology parks are favoring networking and fostering innovations.

2.7 Patent Data for Measuring Innovation

There are several methods to protect intellectual property rights, and one of them is patenting. According to *"Cambridge English Dictionary: Meanings & Definitions," 2025,* a patent is "a legal right to make or sell an invention for a particular number of years". The definition highlights monopoly rights for the invention and the limited period of the exclusive rights for the patent.

World Intellectual Property Organization, 2021, provides detailed information about patents, including specific features of patents and steps of the procedure for granting a patent. Patent information is beneficial for avoiding duplicates in R&D processes, estimating the value of patents, making key business decisions, and identifying key trends in different technologies. Patents have a specific structure, which includes information about applicants, inventors, description (explanation of the invention and its novelty), claims (legal definitions of the subject which is patented), priority filling (the first filing), priority date (the date of the first filing), filing date (date of the application submission in a particular office), designated states (list of countries where the rights can be used), legal status (used to identify if the patent was granted), citations and references (references to the related technologies), document kind status (identificatory of the specific type and status of the patent), country codes, classifications (codes that are used to define the classification of the technologies related to the patent).

The usage of patent data analysis for measuring innovations is becoming more and more common due to the database expansion, better computational methods, and more powerful tools, and a better understanding of the whole system (*Nagaoka et al., 2010*).

One of the parameters that can be examined is the number of patent applications, which can be tracked during a specific period of time. The number of patents can be found for each country, region, applicant, inventor, or technological area. Research by *Sobolieva et al., 2021* shows how the techniques can be applied. The findings of the analysis show that there is a connection between the number of patents for a specific technology and the trends in technological development worldwide. Some fields, like computer technologies and AI, can be seen as the basis for new trends and further development of technologies.

The findings show that patent data can be valuable for fostering and measuring innovations. The data provided in the patent can help to define the main trends in specific areas, get information about the most active applicants or inventors on different levels. The practical application of patent data will be presented in the analytical part of the thesis.

3. Methodology

In the following chapter, the detailed methodology of the current research will be presented. It will cover the process of selecting the necessary patents and creating the dataset for further analysis, all the software and modules that were used to process the data, as well as limitations. All the classifications that will be used in the analytical part are provided.

3.1 IPC and CPC Classifications

There are two of the most common methods to classify patents worldwide – IPC (The International Patent Classification) and CPC (Cooperative Patent Classification) classifications.

The IPC classification was established by the Strasbourg Agreement of 1971, and it provides the division of patents based on different technological areas. The classification was created to standardize the rules of division and organize the patents for easier access to technical and legal information. As it is stated in the Guide to the International Patent Classification (*WIPO, 2025)*, IPC classification enables the analysis of technological progress, assessment of novelty, and inventive step in each patent.

CPC classification was launched by EPO and USPTO and represents another system that is similar to IPC *("About CPC | Cooperative Patent Classification," n.d.)*. According to *Mailänder, 2016*, IPC and CPC classifications are highly similar to each other, but CPC can be considered as a finer subdivision of IPC classification. Because of this, CPC codes can be easily converted into IPC codes. There is an exception – class Y02 in CPC which cannot be found in IPC.

In the research, both CPC and IPC classifications will be used for the selection of AI and sustainability-related patents, but the co-occurrence of IPC codes only will be analyzed.

3.2 AI-related Patents

To get the necessary data, it was decided to use the WIPO classification methodology, provided by *WIPO, 2019* for patents related to artificial intelligence. This classification is based on the following components:

- Cooperative Patent Classification (CPC) codes related to AI;
- Keywords related to artificial intelligence: words related to core concepts;
- Non-specific classes that contain keywords related to artificial intelligence: words that are commonly related to computing or mathematics and that are often used in artificial intelligence. This approach allows to include not only classes that initially imply the usage of new technologies, but also other classes where artificial intelligence was used.

There are many fields that contain text information, but only some of them were selected for the research: English title, English abstract, English claims, and the Object of invention. It was decided to exclude the full text from the analysis so as not to have false positive results.

The data was downloaded from Orbit – software that allows to get all the data about patents. It is one of the most complete tools that makes it the best option for the research. For the data collection, a specific query provided in the source mentioned above was used. As a result, information about publication numbers was downloaded. Afterwards REGPAT database was used to get the rest of the necessary information. At this step, 213591 patents were obtained.

3.3 Connection with REGPAT Database

In order to get information about CPC and IPC classes, applicants and inventors, as well as regions, data obtained from Orbit had to be merged with the *OECD REGPAT database, May 2025*. One of the biggest advantages of the REGPAT database is the fact that information about regions is processed in a highly detailed way: particular regions are linked to inventors and applicants based on addresses. The database also provides information about the patents filed under both EPO (The European Patent Office) and PCT (The Patent Cooperation Treaty). There are separate objects that will be used in the research: Applicants, Inventors, IPC codes, and Description of Regions. The data model of the REGPAT database, with only the fields that are necessary, was created using Lucidchart. The database scheme is presented in the picture:

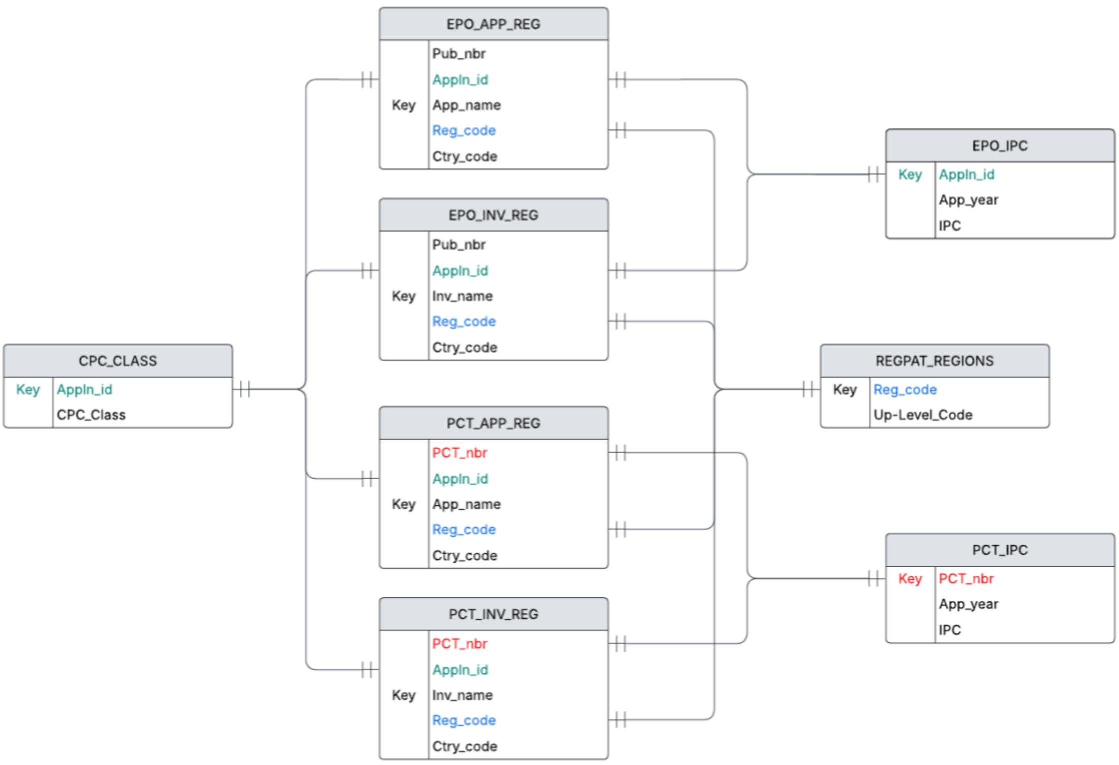

Figure 5. REGPAT Database Model

Since data about PCT and EPO patents are in different objects, the following steps were necessary for matching REGPAT with the data from Orbit:

1. Before merging all the information to create a sample for the analysis, the necessary fields were defined:
    - **Objects with the information about applicants (EPO_APP_REG, PCT_APP_REG)**: Appln_id (surrogate key necessary for creating connections with all other objects), App_name (applicant's name that will be used to identify applicants who patent the most), Reg_code (NUTS3 code of the region), Ctry_code (code of the country), PCT_nbr (used for merging with IPC classes, only for PCT patents).
    - **Objects with the information about inventors (EPO_INV_REG, PCT_INV_REG)**: Appln_id, Reg_code, Ctry_code, PCT_nbr (all follow the same logic that is relevant for applicants).
    - **Objects with lists of IPC classes (EPO_IPC, PCT_IPC)**: Appln_id (necessary to match with previous objects), IPC (list of all IPC codes related to the patent), App_year (filing year), PCT_nbr (to merge PCT patents).
    - **Object with regions (REGPAT_REGIONS)**: Reg_code (necessary to match with previous objects), Up_Level_Code (NUTS2 code that will be used for the analysis of European countries in the next steps).
    - **Object with CPC classes (CPC_CLASS)**: Appln_id (necessary to match with previous objects), CPC_Class (list of all CPC classes related to the patent).
2. In the dataset obtained from Orbit, in the column Publication Numbers, there are several numbers related to one patent. For this reason, publication numbers for EPO (starting from EP) and PCT (starting from WO) were selected and stored in separate objects. The indices of rows in the initial table were also saved for each patent to check for duplicates later. Publication numbers in Orbit and REGPAT have a scarcely different format, that is why necessary modifications were made using Python to merge all the objects in the current and next steps. After making a split between EPO and PCT patents, the following numbers were obtained:
    - 256534 publication numbers were assigned to the EPO group, but only 213591 unique row IDs were mentioned. It means that in some patents, several publication numbers start from EP. This information will be used later to check for duplicates and process missing information.

- 157168 publication numbers were assigned to the PCT group, and 146673 unique row IDs were used. Similar logic as for EPO patents will be applied to process missing information.

3. EPO patents from Orbit were merged with EPO_APP_REG and EPO_INV_REG to enrich the data with the selected fields. “Left Join” was used in order to have all the publication numbers for EPO patents, even if there is no information about applicants or inventors. Connecting the tables allowed to get the unique field Appln_id. Despite the fact that further analysis will be conducted for applicants only, using inventors was necessary: if in REGPAT there is no information about applicants, it is impossible to get Appln_id for further connections with regions, IPC, and CPC classes. Using both applicants and inventors at this stage provided access to the indicated field that was filled in at least one of the tables.
4. After creating a connection with EPO_APP_REG and EPO_INV_REG, rows in the new table were grouped based on the row index indicated in the second step. It was important to group based on this field to save the information, also about patents that were not matched with any applicants or inventors, to find out why the data is missing. It was found that there are no patents where information, at least about one of the classifications, is accessible.
5. Steps 3 and 4 were repeated for PCT patents, but with one modification: connection with IPC classes was made by using the field PCT_nbr. For the PCT group, there were also no patents where both IPC and CPC codes were missing.
6. Tables of PCT and EPO patents were combined for the further analysis.
7. Table REGPAT_REGIONS was merged with all the patents from Europe to substitute NUTS3 codes with NUTS2 codes.

After all the steps, patents with missing data were processed. It was found that there are 331419 patents in total that have all the information necessary for the analysis. Since in the beginning, for one patent, several publication numbers could be mentioned, it was decided to check if there is any overlap: information about patents could be missing in the EPO section, but it was in the PCT part instead, and vice versa. At this stage, EPO and PCT patents were processed separately, and the following reasons for missing data were found:

- For EPO patents, it was found that the current status is “Pending”. Information was checked using Google Patents, and publication numbers were selected randomly. Such patents can be defined as the recent ones, and their influence on it will be presented in the analytical part.

- For PCT patents, it was noticed that in some cases, publication numbers were filled out not in the same way as in the REGPAT database.

Patents with missing information were eliminated from further analysis, and a filter for sustainability was applied only to 331419 patents.

3.4 Sustainability-Related Patents

Since the goal of the research is to analyze the patents related both to AI and sustainability, it was decided first to get the patents related to AI and then, from these patents, select the ones that are related to sustainability. The filter for patents related to sustainability is based mostly on the article *"Measuring environmental innovation using patent data," 2015*. In the Patent Search Strategies part, IPC or CPC classes for selecting environment-related technologies were provided. In addition to that, research by *Favot et al., 2023* indicates that the Y02/Y04S Tagging scheme by EPO is commonly used to define sustainability-related patents. This scheme is based only on CPC codes. The codes provided in both methodologies cover the following technologies:

- Reduction of air pollution;
- Reduction of water pollution;
- Waste management;
- Soil improvement;
- Technologies enhancing water access;
- Generation of renewable energy;
- Energy production from non-fossil fuels;
- Technologies for climate change mitigation;
- Adaptation to climate change;
- Buildings;
- Advanced power grid systems.

If a patent contains at least one of the IPC or CPC codes indicated in the query, it is defined as sustainability-related. After applying the filter for sustainability, only 18181 patents were selected for further analysis.

3.5 Methods Used for the Analysis

After initial data processing, it was necessary to modify the fields responsible for the region, both NUTS3 codes and countries. Getting the information about the regions allows to make conclusions about the collaboration between the regions, for example, if inventors are from different places. Countries in the selected sample were already indicated as two-letter codes (ISO 3166 alpha-2 format); no further changes were needed.

Besides country codes, NUTS3 codes or alternatives for non-European countries were accessible. For the current analysis, it was decided to use NUTS2 codes instead. According to *the European Commission. Statistical Office of the European Union., 2018,* NUTS classification (Nomenclature of Territorial Units for Statistics) provides the logic for the breakdown of European territory into units. The indicated division can be used for regional statistics analysis and socio-economic analysis. NUTS2 provides information about basic regions that will be used in the current research. Patents where all the applicants are from Europe were copied to another table that is necessary for deeper analysis of NUTS2 regions.

Based on the information provided in "*History - NUTS - Nomenclature of territorial units for statistics - Eurostat," n.d.*, NUTS classification is regularly updated, and it is stable for at least three years. This factor is crucial since NUTS classification is used in statistics. Currently, the newest available data refers to 2024. This dataset will be used for the current research.

Before applying the filter for sustainability-related patents, it was decided to calculate the contribution of each country to the overall number of AI-related patents. In each row of the sample, a list of countries and smaller regions for each patent was presented. Shares were calculated in absolute values based on the following formula:

$$\boldsymbol{Absolute\ Share}\ (\boldsymbol{x_i}) = \frac{f(x_i)}{n}, \text{where}$$

$\boldsymbol{x_i} - c$ountry or NUTS2 region

$\boldsymbol{f}(\boldsymbol{x_i}) -$ frequency ofelement x_i

$\boldsymbol{n} -$ total number of elements in a list

Using the absolute share formula allows not only to check if a particular country/region contributed to a specific patent, but also to see the contribution. For example, if for patent X regions A and B participated, and list for regions presented as [A, A, A, B, B], then absolute shares for each of participants in this case will be not 1 and 1 or 0.5 and 0.5, but 0.6 for A and 0.4 for B since one of them appeared more times.

Absolute shares were calculated for each patent, and then information was grouped by country. Similar logic was applied to calculate shares for NUTS2 regions in the table of European patents only. All the grouped data were stored in separate tables for further analysis in a field for the total share. Lists of unique country/region codes, as well as their shares in each patent, were added as extra fields in the table.

After calculating all the shares, a filter for sustainability was applied. The same logic for absolute shares was applied again. Since the information about regions and shares for each patent was already available (added as new fields), grouped values were calculated again and stored in the same table as another field, indicating total share in both AI and sustainability. To check the relative focus on sustainability within each region's AI patenting activity, the percentage was calculated following the formula:

$$\boldsymbol{Relative\ Focus}\ (\boldsymbol{x_i}) = \frac{k}{K} * 100\%\text{ , where}$$

$\boldsymbol{k}$ – total share of region or country x_i in both AI and sustainability

$\boldsymbol{K}$ – total share of region or country x_i only in AI

The next step was to calculate matrices to analyze the co-occurrence for the selected attributes:

- Co-occurrence of two IPC classes in one patent. It was decided to substitute extended codes with shorter ones that contain only four symbols in order to, to concentrate not on the specific technology but on the broader definition of technological class. Usage of IPC codes allows to make a technological space analysis and check the co-citation between classes: if two technological classes were cited together in the same patent, we can talk about technological relatedness.
- Co-occurrence of two regions in one patent. For regional co-occurrence, NUTS2 codes were used.

The last part of the analysis was identifying the applicants who patent the most, both in AI and sustainability. In order to normalize the applicants' names, it was decided to use upper case for all the names. This step was necessary in order to eliminate duplicates. For the shares and descriptive statistics, the same logic as for regions and countries was applied.

3.6 Co-occurrence Analysis of the Regions

Part of the analysis was dedicated to co-occurrence for technological classes in one region and regions in one technological class. Co-occurrence matrix must be following a number of predefined rules:

- Matrix is square: it has k rows and k columns.
- i and j represent the row and column numbers, respectively.
- A co-occurrence matrix is symmetrical about the diagonal: Element with coordinates (i, j) equals to the element with coordinates (j, i).
- Elements with coordinates (i, j) and (j, i) represent the number of times regions i and j that appeared in the same patent.
- Elements on the diagonal are always equal to zero.

For the regions, it was decided to use the matrix without any modifications, while for IPC classes, normalization was used. A co-occurrence matrix is visualized as a network colored by the clusters.

3.7 The Louvain Method for Cluster Detection

As an additional step to the network representation, it was decided to detect clusters using the Louvain Method. The method was developed by *Blondel et al., 2008*. The goal of the research was to create a method that is easy to implement for community detection even in large networks. The core principle is greedy optimization: at each step, the algorithm chooses the best possible solution (optimal) that allows to get the obvious benefit in a particular step. Louvain Method works with the concept of modularity, which estimates the strength of division of a network into communities.

In order to perform the calculations and use the Louvain method, a module in Python was used, specifically **louvain_communities** function. Official documentation provides the information on how the function works *("community API — Community detection for NetworkX 2 documentation," n.d.*). The algorithm has the following steps:

1. First, network is divided into N communities, where N is the number of nodes, so each node i is assigned to its own community.
2. The algorithm shifts each node to all of its adjacent communities in an attempt to determine the maximum positive modularity gain for each node. When there are no nodes that can be moved, the step ends. Modularity gain can be computed as follows:

   $$\Delta \boldsymbol{Q} = \frac{k_{i,in}}{2m} - \gamma \frac{\sum tot * k_i}{2m^2}, \text{where}$$

   $\boldsymbol{k_{i,in}}$ − sum of the weights from node i to other nodes in community C

   $\boldsymbol{k_i}$ − sum of the weights of the links leading to node i

   $\sum \boldsymbol{tot}$ − sum of the link weights of the nodes that are included in C

   $\boldsymbol{\gamma}$ − resolution parameter

   $\boldsymbol{m}$ − size of the graph
3. The third step works with the communities that were defined in the previous step. Communities are treated as nodes, which means that now the weight of the links are defined as the sum of the links of the nodes that are connecting two communities.

Clustering the nodes using the Louvain method allows to get a better representation not only on how strong the connection of separate pairs is, but also to see if they can be split into groups.

3.8 Relatedness Analysis of IPC Classes

In order to analyze how IPC classes are connected to each other, and define what regions are specialized in particular sectors, relatedness analysis was provided. First, the matrix similar to the one for regional co-occurrence was created. Before calculating the normalized values for technological classes, RTA was calculated based on the manual provided by *OECD, 2009*. RTA stands for Revealed Technological Advantage and is used to find the relative specialization of the region: to what extent the region is active in a particular technological field compared with the global average. The following formula for RTA was used:

$$\boldsymbol{RTA_{i,j}} = \frac{\frac{x_{i,j}}{\sum_j x_{i,j}}}{\frac{\sum_i x_{i,j}}{\sum_{i,j} x_{i,j}}}, \text{where}$$

$\boldsymbol{i}$ – regions, $\boldsymbol{j}$ – technologies

$\boldsymbol{x_{i,j}}$ – number of patent applications in region i in technology j

$\sum_{\boldsymbol{j}} \boldsymbol{x_{i,j}}$ – total number of patents in region i for all technologies

$\sum_{\boldsymbol{i}} \boldsymbol{x_{i,j}}$ – total number of patents in technology j in all regions

$\sum_{\boldsymbol{i,j}} \boldsymbol{x_{i,j}}$ – total number of patents

If $RTA_{i,j} > 1$, then region i is specialized in technology j. If $RTA_{i,j} < 1$, then region i is not specialized in technology j. RTA values are necessary for the normalization of the values in the co-occurrence matrix. Based on the RTA, ubiquity was found: how many regions are specialized in a particular technology.

There are technologies that could appear more often due to being common overall. Such technologies can lead to a distorted picture of real specialization. For the normalization of the values, the formula presented below was used:

$$\boldsymbol{normalized\ value_{i,j}} = \frac{co - occurrence_{i,j}}{\max(ubiquity_i, ubiquity_j)}, \text{where}$$

$\boldsymbol{i, j}$ – technologies

$\boldsymbol{co - occurrence_{i,j}}$ – how many regions have RTA in both technologies i and j

$\boldsymbol{ubiquity_i}$ – the number of regions specialized in technology i (have $RTA > 1$)

After normalization, a heatmap can be used to represent the most outstanding pairs. To make a more detailed analysis and calculate descriptive statistics, the matrix is transformed into a long format – sorted table with 3 columns: Region i, Region j, Total co-occurrence. The created table and heatmap will provide information about collaborations.

To make a better representation of the normalized values, the network is created.

3.9 Descriptive Statistics

In the descriptive statistics part, measures of central tendency and variability will be presented for each variable for a better understanding of the data. Descriptive statistics will provide necessary data not only to evaluate the overall picture, but also to find valuable insights about the most outstanding parties. The following metrics will be defined both for countries and regions:

- **Mean** represents the average of the sample: the sum of all values is divided by the total number of elements. The mean value can be used to analyze the skewness of the data.
- **Standard deviation** shows the average distance from each value to the mean, which allows to see the dispersion of the dataset.
- **Minimum and maximum** represent the lowest and the highest values in a variable. Such values are used to calculate the range, as well as provide information about possible outliers.
- **Q1 (first quartile, 25th percentile)** represents the element that is located in the middle between the minimum and median. It can be said that 25% of the data is below Q1.
- **Q2 (second quartile, median)** shows the value that is located in the middle of the sorted list of values. If the number of values is odd, then the middle value is taken, and the average of two values in the middle is taken in case of even number of values. Median can provide valuable information about skewness of the distribution and outliers.
- **Q3 (third quartile, 75th percentile)** represents the element that is located in the middle between the median and maximum. It can be said that 25% of the data is above Q3.
- **IQR (interquartile range)** defines the spread of the central 50% of the values. It can be calculated as Q3 – Q1. IQR can be used to find outliers for each of the variables, that can be identified by using the formula $Q3 + 1.5 * IQR$. All values that are greater can be considered as outliers.

In the current research, the usage of mode, the most common value, was not needed, since all the shares were presented as continuous variables.

Based on the findings in the descriptive statistics, further analysis of the outstanding countries and regions will be conducted.

3.10 Data Analysis Tools

Data processing and analysis were conducted using Python programming language in Visual Studio Code. The selected tool helped to work with a lot of data and use various techniques for data processing and visualization. The following modules were selected for the current research:

- **Pandas**: a module for data manipulation and analysis *("pandas - Python Data Analysis Library," n.d.*).
- **Csv**: a module that is used to import all documents in CSV format.
- **Ast:** a module that enables working with abstract syntax grammar in Python *("ast — Abstract Syntax Trees," n.d.*).
- **Collections:** a module for working with different types of containers – objects that are used for data storage (*"collections — Container datatypes," n.d.*).
- **Matplotlib, Seaborn, Plotly:** modules necessary for data visualization (*"Matplotlib documentation — Matplotlib 3.5.3 documentation," n.d.; "seaborn: statistical data visualization — seaborn 0.13.2 documentation," n.d.*).
- **Pycountry, Geopandas:** modules used for working with geometrical data (*"pycountry," 2024; "GeoPandas 1.1.1 — GeoPandas 1.1.1+0.ge9b58ce.dirty documentation," n.d.*).
- **Itertools:** a module for efficient work with loops (*"itertools — Functions creating iterators for efficient looping," n.d.*).
- **Folium:** a module used for data manipulation, including an interactive part (*"Folium — Folium 0.20.0 documentation," n.d.*).
- **Branca:** a module that is used with folium for the HTML elements (*"Welcome to branca's documentation! — branca 0.8.1 documentation," n.d.*).
- **Numpy:** a module that is used to work with various arrays and matrices (*"NumPy," n.d.*).
- **Networkx:** a module that allows to work with complicated networks (*"NetworkX — NetworkX documentation," n.d.*).
- **Community (community_louvain):** module for allows to detect communities in networks (*"community API — Community detection for NetworkX 2 documentation," n.d.*).

As it was mentioned previously, part of the current research is regional analysis. For visualizations, it was decided to use maps – one for Europe and one for countries. Both visualizations required the use of additional files to define the coordinates and boundaries. Maps for countries were created using files in the section Countries provided by *"Natural Earth - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales," n.d*. This source provides data free of charge and is supported by NACIS (North American Cartographic Information Society).

For regional-level analysis of European countries, the data shape file provided by GISCO (Geographic Information System of the Commission) was selected. The dataset contains information for NUTS1, NUTS2, and NUTS3 levels. For the analysis, the following filter was used: LEVEL = 2.

3.11 Limitations

During the research, several limitations were identified:

- **Query for AI patents:** The first limitation is related to the query that was used in Orbit to get patents related to AI. Even though CPC codes and keywords are used to select necessary objects, the dictionary for words related to AI cannot be clearly defined due to the fact that more and more words can be related to artificial intelligence and machine learning, since new technologies are appearing. Different sources can also provide different dictionaries.
- **Missing information for several publication numbers:** Another limitation appeared during the matching between data about publication numbers and the REGPAT database. As it was mentioned in part 2.3 Connection with REGPAT Database, some values were missing because of the status "Pending", while for others the format was different, which led to the situation then no connections were established.
- **No uniform applicants' names:** The third limitation is related to the analysis of the most active and important applicants. In some cases, one applicant could have several applicants' IDs because, for instance, in one case the name was written in lower case, while in another one all the words started with capital letters. Besides that, several companies can be in one holding group, and the shares are also calculated separately.
- **Computing power limit:** The next crucial limitation in the current research is related to selecting patents that are only related to sustainability. Such a situation was caused by the fact that object EPO_IPC contains 18249394 rows, object PCT_IPC contains 17784903 rows, and object CPC_CLASS has 67963218 rows. In total, out of all tables, 103997515 rows can be found, and in each row, several IPC or CPC codes can be contained. For such big data, better equipment was necessary.

4. Presentation of the Results and Data Analysis

In this chapter, the analysis of the results will be presented. To begin with, the main descriptive statistics will be provided in tables and graphs, and then a deeper regional analysis will be conducted.

4.1 Descriptive Statistics: Number of Applications

Before analyzing particular countries or regions, it was decided to see the general tendency. In the initial dataset, the first AI-related patent was in 1980, and this is where the first value appears on the graph, which represents the number of AI-related applications per year.

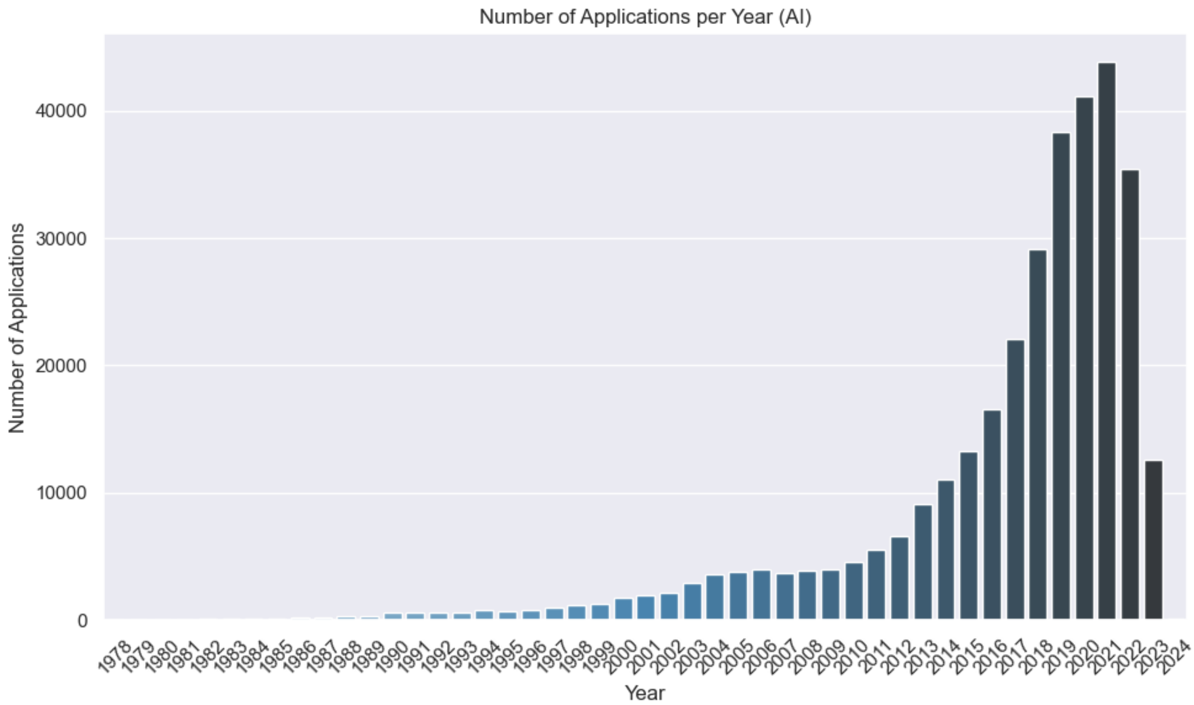

Figure 6. Trends in AI-Related Patent Applications Over Time

As can be seen from the bar chart, during the first period (from 1980 to 2003) is extremely low. During this time, overall, the patent activity was low, and the development of technologies was not as fast as it is now.

From 2004 to 2013, gradual growth can be noticed. The rising trend was still slow, but it became more noticeable due to the fact that data analytics development became more common. From 2014 to 2019, a rapid exponential growth can be seen. Such a trend can be linked not only to the increase in patent filing, but also to the widespread use of artificial intelligence, in particular deep learning and big data technologies. The peak can be noticed in 2020 and 2021, which shows that innovation activity did not stop despite the COVID-19 pandemic. From 2021 to 2022 and from 2022 to 2023, there were significant drops. In addition to that, for 2024, there are no values presented. Such a situation confirms the limitation that was noticed previously –

for many recent patent applications, the status is still “Pending”, which is why there was no possibility to link them to any data from the REGPAT database.

The graph presented above contains information only about AI-related patents. Because of one of the limitations mentioned above, conducting a similar analysis for sustainability-related patents was not possible.

A similar graph was created for patents that have both AI and sustainability factors.

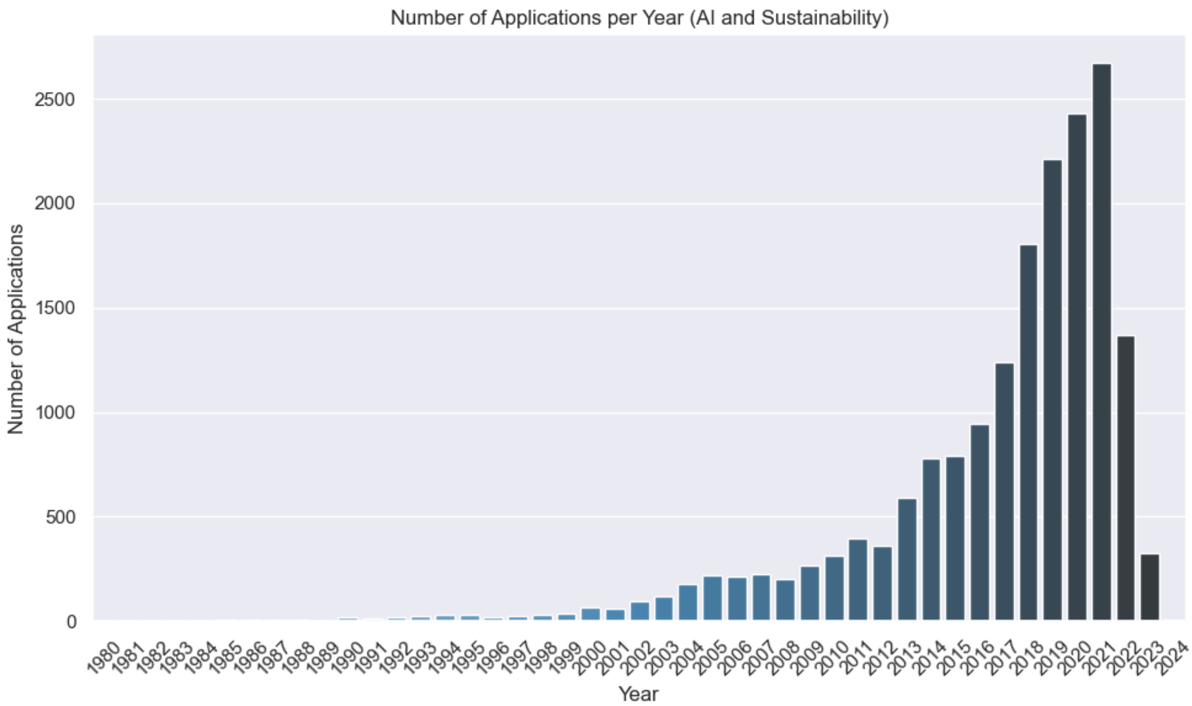

Figure 7. Trends in AI and Sustainability-Related Patent Applications Over Time

The trend that appeared in the first graph is similar to the second one. Partly, it can be explained by the fact that a filter for sustainability was applied already to AI-related patents. In addition to that, the rising trend in sustainability is also affecting the distribution. The trend explained previously in the literature review part is the same – there is a rapid growth during the last years, and overall, the trend is never going down.

4.2 Distribution of Shares: Country-Level

After analyzing the number of applications, the country-level analysis was conducted. Based on the formulas explained in the methodology, shares for all countries were calculated.

For the country-level analysis, 115 countries were selected. The first table represents the distribution of the shares for 62 countries that have patents in both AI and sustainability.

	Absolute Share in AI	Absolute Share in AI and Sustainability	Relative Focus on Sustainability, %
Mean	5335.60	293.24	8.17
Std	17059.88	899.94	8.79
Minimum	3.67	1.00	1.30
Maximum	124259.36	6331.06	50.00
Q1	46.77	2.69	3.78
Q2 (Median)	359.53	12.06	5.46
Q3	2454.84	145.25	8.11
IQR	2408.07	142.56	4.33

Table 1. Descriptive Statistics for the Country-Level Data

Based on the table with descriptive statistics, the following patterns were found in each column:

- **Absolute Share in AI:** In this column mean value is 5335.60 while the standard deviation is 17059.88, which means that the distribution of share is highly unequal. Minimum and maximum values represent the range – all the countries included in the research have a share that is in [3.67, 124259.36], which also shows a massive difference. Median value is more than 10 times lower than the mean, which supports the point about big variance and the presence of strong outliers (high absolute share). IQR value also represents a huge spread within countries in the middle 50%. Based on the values, it can be said that the data has a strong positive skewness – only a few countries have high values (leaders in AI), while the vast majority have low shares.
- **Absolute Share in AI and Sustainability:** The values of absolute shares of patents related to both AI and sustainability are significantly lower than in the case of only sustainability, but the mean and standard deviation still show that there is a widespread trend across the countries. Median of 12.06 with the maximum value of 6331.06 also shows an overall tendency of low activity in the case of sustainability. Lower numbers compared to the first case can be explained by the filter that was used to define patents in sustainability – out of 331419 patents, only 18181 were chosen, which is only 5.5%. The skewness for AI and sustainability is the same as only for AI – data is strongly positively skewed.
- **Relative Focus on Sustainability:** Overall, the values in the field that represent relative focus are lower than in the two previous cases due to the fact that here the measurement is in %, so in any circumstances, the maximum cannot exceed 100%. On average, only 8.17% of the AI patents in a country are also related to sustainability, and half of the

countries have even less than 5.46% of such patents. Data about relative focus has a moderate positive skewness.

To have a visual representation of the data presented above, it was decided to create an interactive map. Some of the regions were omitted for visualization because coordinates were not found in the database, but they were still included in the descriptive statistics and analysis. These regions are KY (Cayman Islands), SG (Republic of Singapore), HK (Hong Kong Special Administrative Region of China), BB (Barbados), MT (Republic of Malta), VG (British Virgin Islands), LI (Principality of Liechtenstein), BM (Bermuda), AI (Anguilla), MC (Principality of Monaco), MO (Macao Special Administrative Region of China). The snapshot of the map with accessible countries is presented below.

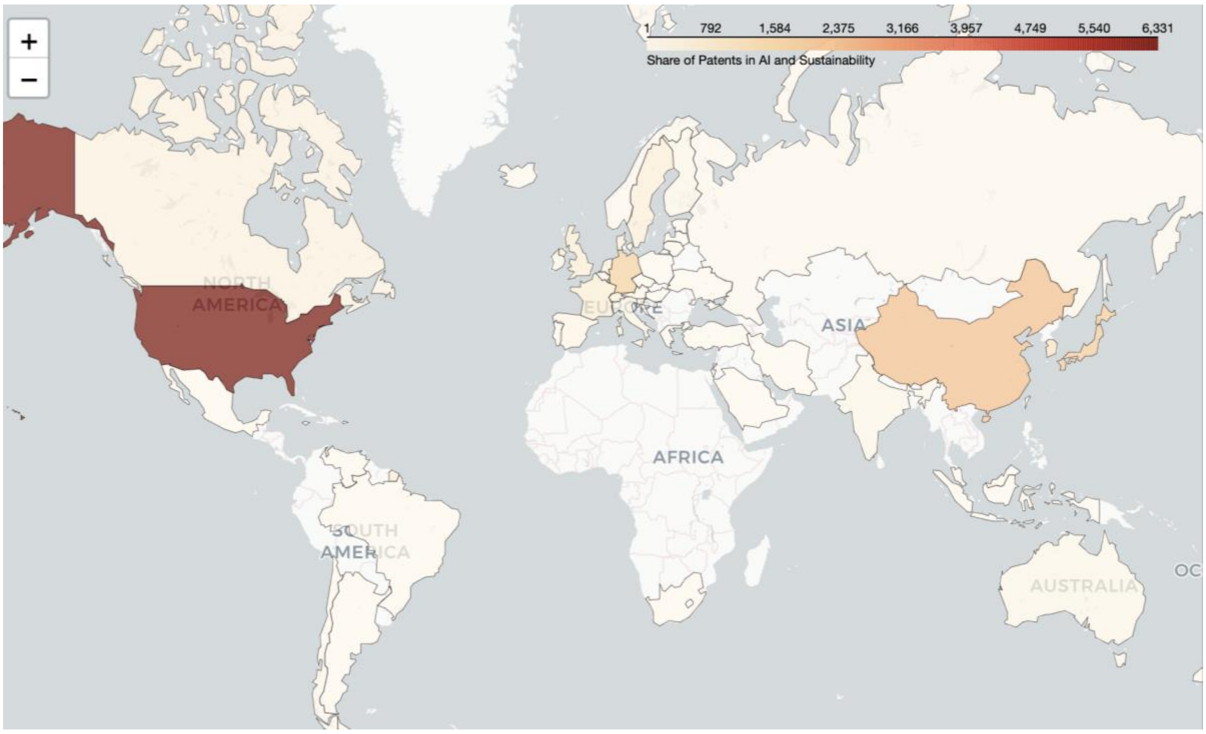

Figure 8. Map of Patenting Activity on a Country-Level

The color of the country represents the absolute share of patents in both AI and sustainability: a darker color shows a higher absolute share. As can be seen from the map, the United States of America, China, Japan, and Germany have the highest shares. These countries can be considered as outliers that caused a high standard deviation value in the absolute share of AI and sustainability. At the same time, most of the countries have a light-yellow color that represents a long tail described before – most of the countries have low shares. A deeper analysis of the leading countries will be presented in the next chapter.

As it was mentioned before, only 62 countries out of 115 have patents related to both AI and sustainability, while 53 countries have patents only with AI. Selected countries were sorted from the highest share to the lowest one. The results can be seen in the table:

Country Code	Country Name	Absolute Share in AI
AE	United Arab Emirates	121.73
CY	Cyprus	120.30
RO	Romania	40.00
QA	Qatar	40.00
SK	Slovakia	30.00

Table 2. Total Patent Share in AI for Countries Without Sustainability (Absolute Values)

According to the table, some countries still have a high absolute share in AI but do not have any patents in sustainability. Nevertheless, after comparing the data with the descriptive statistics presented in Table 1, it can be said that shares of the top 2 countries from Table 2 are located between Q1 and Q2, the second 25% of the data.

4.3 Distribution of Shares: Region-Level

A similar analysis was conducted for the European regions. For the research, NUTS2 codes were used. The focus on European countries will allow to analyze the regions in a more detailed way and find the cross-country collaborations: which NUTS2 regions collaborate the most with the countries outside of Europe. Out of 249 NUTS2 regions, 147 regions have patents that combine both sustainability and AI. Relevant descriptive statistics are presented in Table 3.

	Absolute Share in AI	Absolute Share in AI and Sustainability	Relative Focus on Sustainability, %
Mean	572.13	32.88	8.79
Std	1443.28	81.35	10.27
Minimum	1.75	0.33	0.26
Maximum	10035.84	670.26	74.11
Q1	53.88	3.00	3.15
Q2 (Median)	145.50	6.00	5.57
Q3	458.47	26.09	9.48
IQR	404.59	23.09	6.33

Table 3. Descriptive Statistics for the Region-Level Data

The descriptive statistics of all three variables lead to the following conclusions:

- **Absolute Share in AI:** The mean value is around 4 times greater than the median, and it is also higher than Q3, which means that the mean is located in the last 25% of the data with the highest values. The interquartile range of 404.59 and the standard deviation of 1443.28 indicate that the variance of the data is extremely high. At the same time, a high maximum value shows that there are several outliers with high absolute shares that have a huge contribution to the increase of the standard deviation and mean. Based on all the given information, it can be said that the data is strongly positively skewed with a long tail of low shares. Outliers in the data will be regions with shares greater than 1065.35.

- **Absolute Share in AI and Sustainability:** Interquartile range of 23.09 with Q1 equal to 3 represents that 50% of the regions have comparatively low absolute shares in comparison with the leaders. A high difference between the mean and median shows high asymmetry of the data. The maximum value of 670.26 is more than 100 times greater than the median and 20 times greater than the mean, which shows the presence of several leaders. As in the previous case, the data is strongly positively skewed – most of the regions have low absolute shares of patents related to both AI and sustainability. Outliers are greater than 60.72.
- **Relative Focus on Sustainability:** Standard deviation of 10.27 represents a moderate to high spread of the values within the dataset. 50% of the regions have a moderate focus on sustainability, from 3.15% to 9.48%. These values are close to the relative focus on the country-level. The mean value is higher than the median, that shows a positive skewness of the data. The big difference compared to the country-level analysis is the maximum value of 74.11%, while in the case of countries, it was only 50%. It represents that there is at least one region with a strong prioritization of sustainability. Outliers have a relative focus greater than 18.98.

Based on the shares calculated for the regions, an interactive map for the regional level was created.

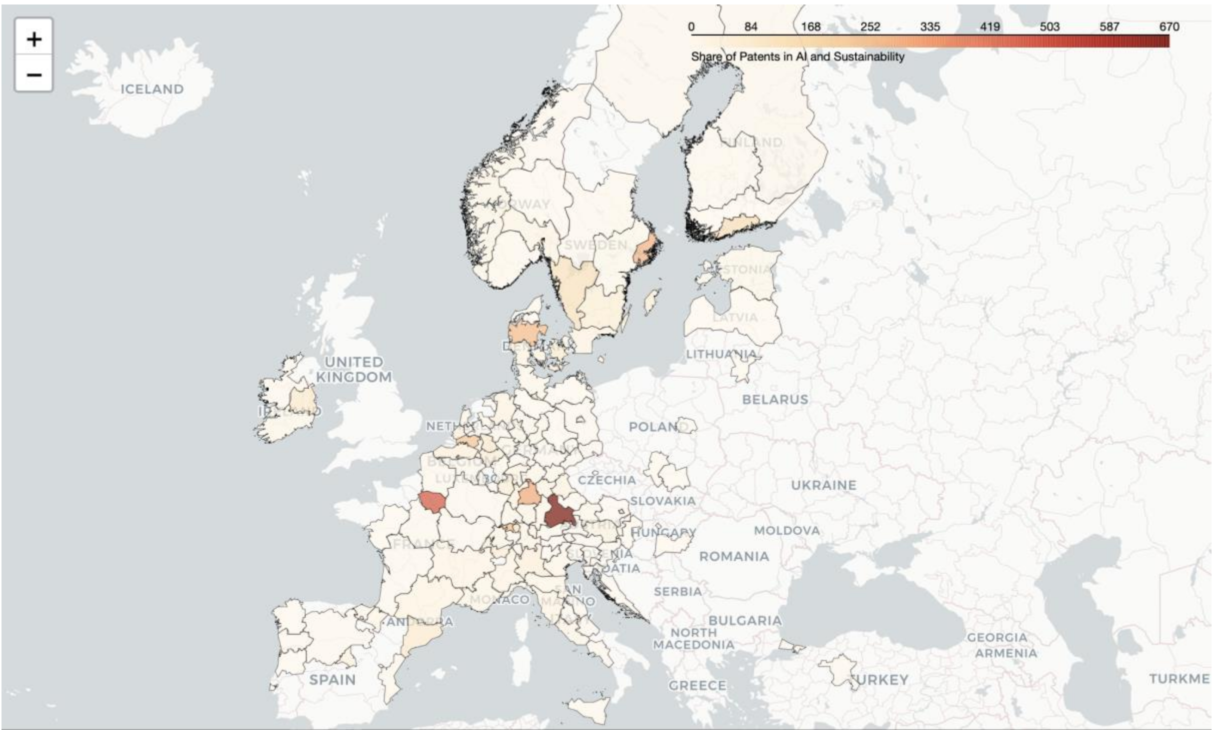

Figure 9. Map of Patenting Activity on a Region-Level

The map for the NUTS2-level follows the same principles as the one that was used for countries: darker color represents a higher share, and light color shows a small share of patents. As can be seen on the map, some regions can be considered as leaders based on the total share of AI-Sustainability patents. This list includes: DE21 (Oberbayern, Germany), FR1 (Île-de-France, France), DE11 (Stuttgart, Germany), SE11 (Stockholm, Sweden), DK04 (Midtjylland, Denmark).

The indicated NUTS2 regions have a high contribution to the descriptive statistics, and they will be analyzed in more detail in the next parts. Since focusing exclusively on absolute shares in both AI and sustainability is not enough for further conclusions, the leaders in all three aspects represented in Tables 4, 5, and 6 will be considered.

The values were sorted in three different ways: absolute share of AI-related patents, absolute share of AI and sustainability-related patents, and share of AI and sustainability-related patents from AI-related patents. Relative values were calculated for absolute shares in AI and AI-sustainability to analyze what share each region has in %. For all 3 cases top 10 regions were selected for the analysis.

Region	**Absolute Share**	**Relative Share, %**
FR1	100035.84	11.93
DE21	9525.19	11.33

NL41	7825.06	9.30
SE11	4964.44	5.90
DE11	4958.23	5.90
FI1B	3833.33	4.56
SE23	1727.76	2.05
IE06	1713.38	2.04
DE30	1707.00	1.03
DE25	1651.85	1.96

Table 4. Top 10 Regions Based on Absolute Share in AI

Region	**Absolute Share**	**Relative Share, %**
DE21	670.26	13.87
FR1	461.24	9.54
DE11	293.06	6.06
SE11	278.00	5.75
DK04	257.50	5.33
CH03	250.04	5.17
NL41	228.50	4.73
FI1B	131.00	2.71
SE23	121.00	2.50
CH04	106.78	2.21

Table 5. Top 10 Regions Based on Absolute Share in AI and Sustainability

Region	**Relative Focus, %**	**Absolute Share in AI and Sustainability**
DK02	74.11	71.50
DK04	60.23	257.50
HU22	42.86	1.00
NO02	40.00	2.00
SI03	36.36	4.00
ES22	30.48	19.00
DE94	30.04	38.00

AT34	29.71	26.00
IE023	28.57	2.00
ITI2	22.22	1.00

Table 6. Top 10 Regions Based on Relative Focus in Sustainability

According to the values that were defined as thresholds for outliers, all the regions are included in this category. Based on Tables 4 and 5, 7 regions are present in both tables, which means that they are active in patenting in AI and have a focus on patents related to both AI and sustainability. The two most active regions are FR1 and DE21 – they have high shares just in AI and in AI and sustainability, which makes these regions worth further analysis. Regions DE11 and SE11 have similar high shares in both cases – they will also be included in the analysis.

Based on Table 6, despite the fact that all the included regions have high relative focus, 8 out of 10 regions have low absolute shares in AI and sustainability. It means that even though the relative focus is high, regions are not active in patenting activities in terms of AI and sustainability. The first two regions – DK02 and DK04 will be included in the further analysis.

From the 249 NUTS2 regions included in the initial dataset, 102 regions do not have patents in both AI and sustainability. The same logic as for the country-level was applied to define the regions with the highest shares. The top 5 regions, including the information about NUTS2 code, region name, and absolute share, are presented in Table 7.

NUTS2 Code	Region Name	Absolute Share in AI
FRH	Brittany, France	300.38
CY00	Cyprus, Cyprus	114.50
TR33	Manisa, Turkey	71.00
ES61	Andalucía, Spain	65.15
ITF4	Puglia, Italy	59.00

Table 7. Total Share in AI for Regions (Without Sustainability)

Regions CY00, TR33, ES61, ITF4 that have the last four places have relatively low absolute share in comparison with all other regions included in the calculations of values in Table 3 – they are located between Q1 and Q2. At the same time, the leader, FRH (Brittany, France), has a share of 300.38, which is in the middle between Q2 and Q3, which shows a stronger focus on artificial intelligence.

4.4 Technological Classes Relatedness

In order to analyze how similar the technologies are, relatedness analysis was conducted for technological classes. Based on the RTA values and co-occurrence between two technological classes, the relatedness value was calculated. All the values have to be in the interval [0,1]. If relatedness is close to 1, then two technological classes have a strong connection, and the opposite for relatedness close to 0. Based on the values, the heatmap was created:

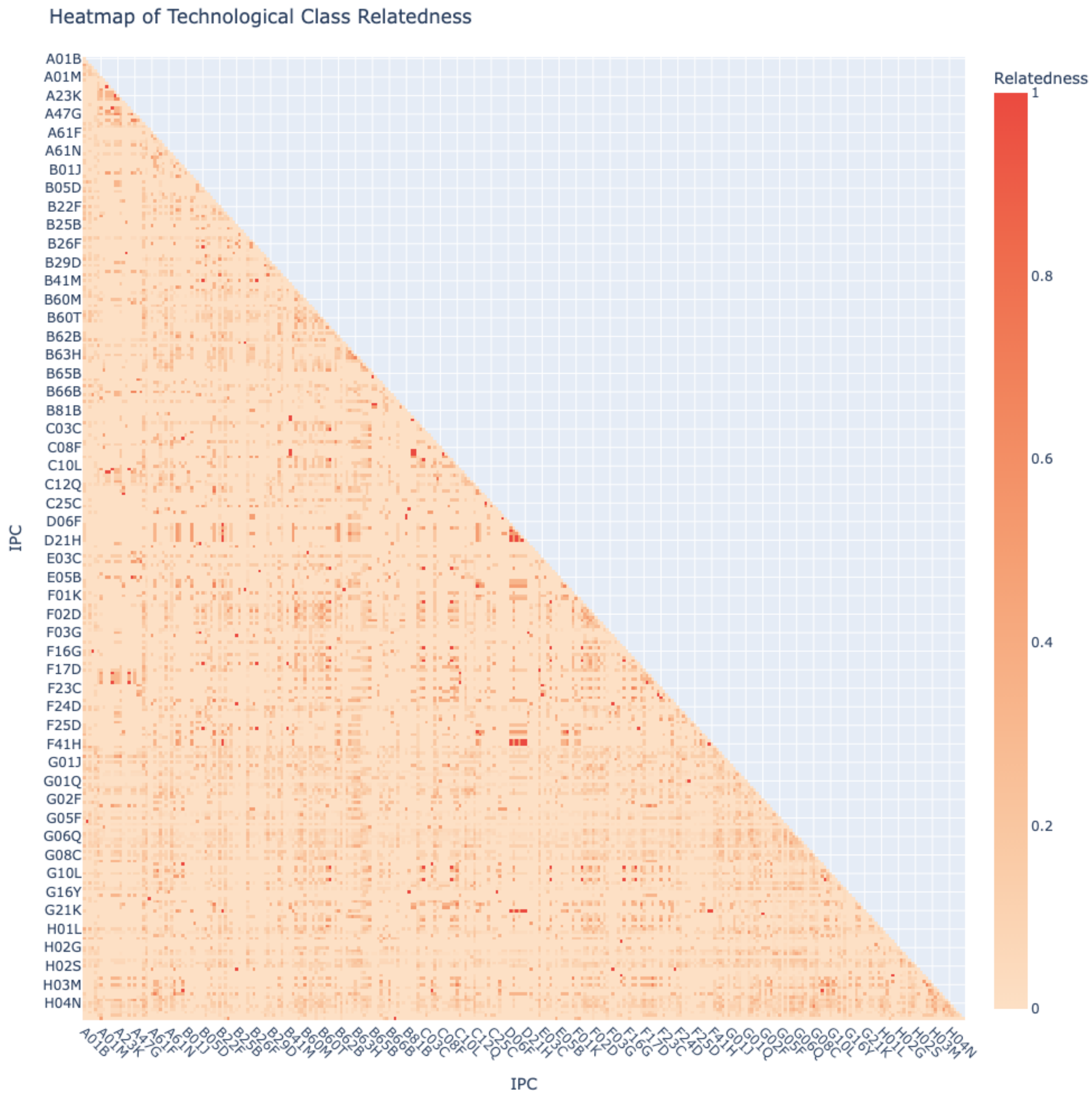

Figure 10. Heatmap for Technological Class Relatedness

In total, 14105 pairs have non-zero relatedness out of 48672 possible pairs. The visual representation with a heatmap confirms that, while for most of the pairs normalized relatedness is 0 or close to 0, there are several pairs with strong connections equal to 1. If relatedness is equal to 1, it means that the co-occurrence (number of regions with RTA in both technologies) of two technological classes is equal to the maximum of ubiquity of these two classes. It means that the technology with smaller ubiquity always appears in regions where there is the more common technology is. Such a situation shows that the two technological classes are highly related. Out of 14105 pairs, 148 have relatedness equal to 1.

To check the pairs, it was decided to find ubiquity for both technological classes and co-occurrence. In most cases, co-occurrence is equal to 1, as well as both ubiquities. For the rest of the pairs, descriptive statistics was created to select the strong pairs.

	Relatedness
Mean	0.16
Std	0.12
Minimum	0.02
Maximum	0.8
Q1	0.07
Q2 (Median)	0.13
Q3	0.2
IQR	0.13

Table 8. Descriptive Statistics for Relatedness

The mean is higher than the median by 0.03, which means that the data is slightly positively skewed. Standard deviation is equal to 0.12 and shows a moderate spread between the values. The interquartile range is from 0.07 to 0.2, which shows that 50% of the data has relatively low values and is not highly related. Based on IQR and Q3, outliers have relatedness greater than 0.4. Based on the filter for relatedness > 0.4, there are 569 outliers.

It was decided to conduct extra analysis for the outliers based on the relatedness value. First, average relatedness by the section for all pairs with relatedness in the range (0,1) was calculated. Sections are defined according to the Scheme provided in *"International Patent Classification (IPC)," n.d.*. The scheme defines sections for all patents, which are represented by the first letter of the IPC code:

- **A** – Human Necessities
- **B** – Performing Operations; Transporting
- **C** – Chemistry; Metallurgy
- **D** – Textiles; Paper
- **E** – Fixed Constructions
- **F** – Mechanical Engineering; Lightening; Heating; Weapons; Blasting
- **G** – Physics

- **H** – Electricity

The heatmap for the IPC sections is presented below:

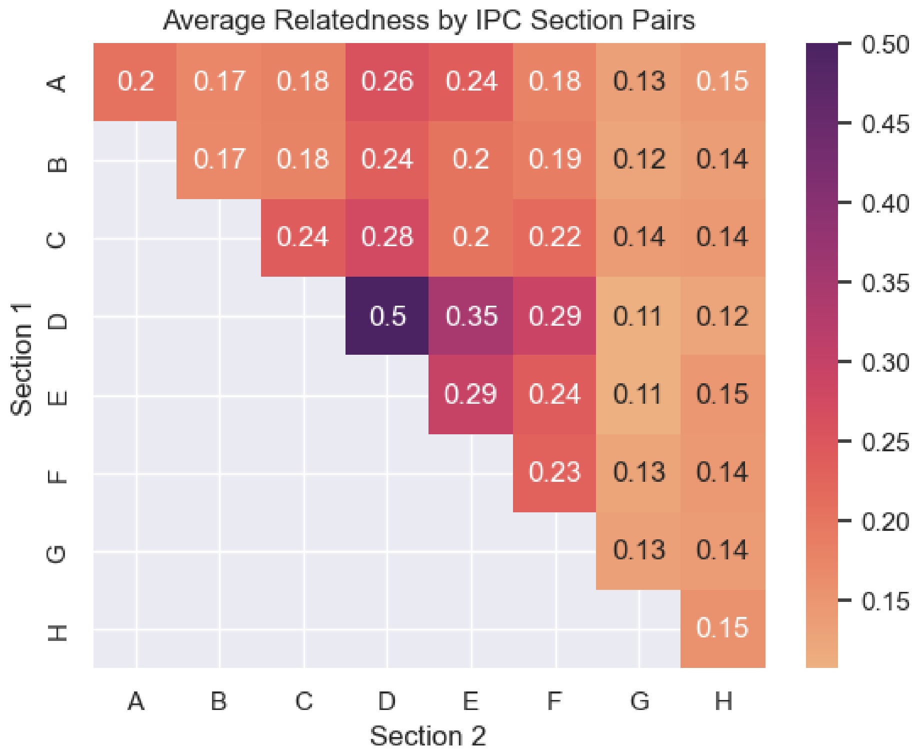

Figure 11. Heatmap for Average Relatedness (Pairs with Relatedness in (0,1))

Based on Graph 8, pairs where both classes are in section D have the highest average relatedness (0.5). Most of the pairs that have A, B, G, and H are relatively weak.

After selecting only outliers for the heatmap, the distribution of average values changed drastically. It was found that the highest average values correspond to pairs A-A, E-E, H-F, A-E, and A-H.

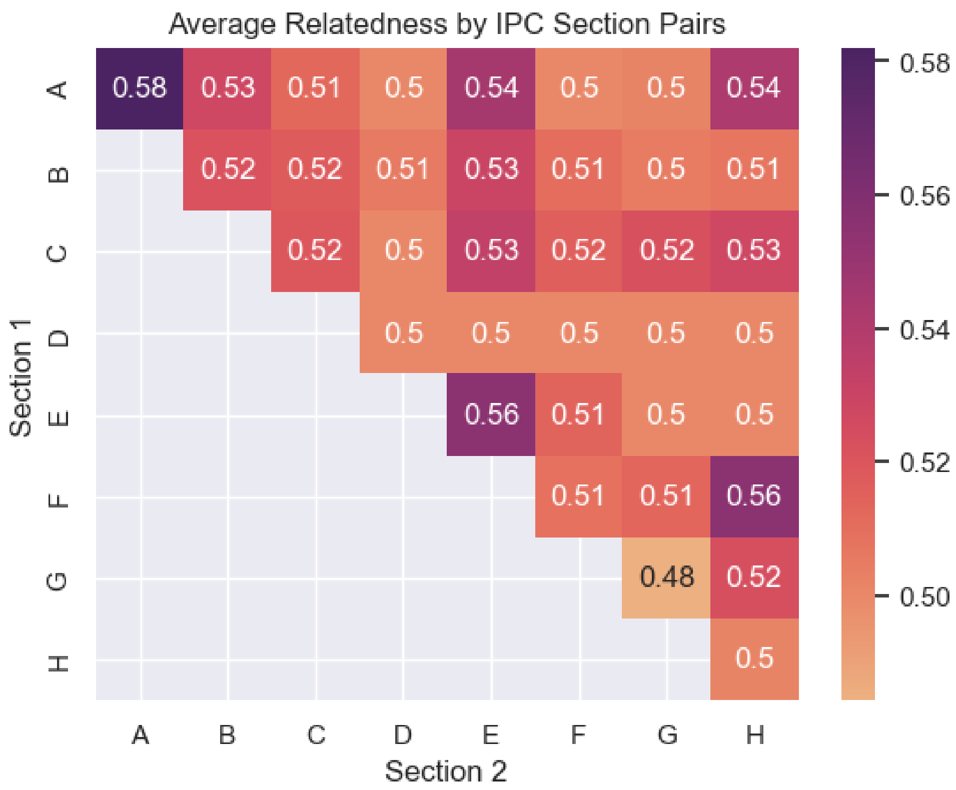

Figure 12. Heatmap for Average Relatedness (Outliers)

To analyze the most active regions for each section, it was decided to find the total number of IPC classes in each section where the regions have RTA. Results are presented on the maps.

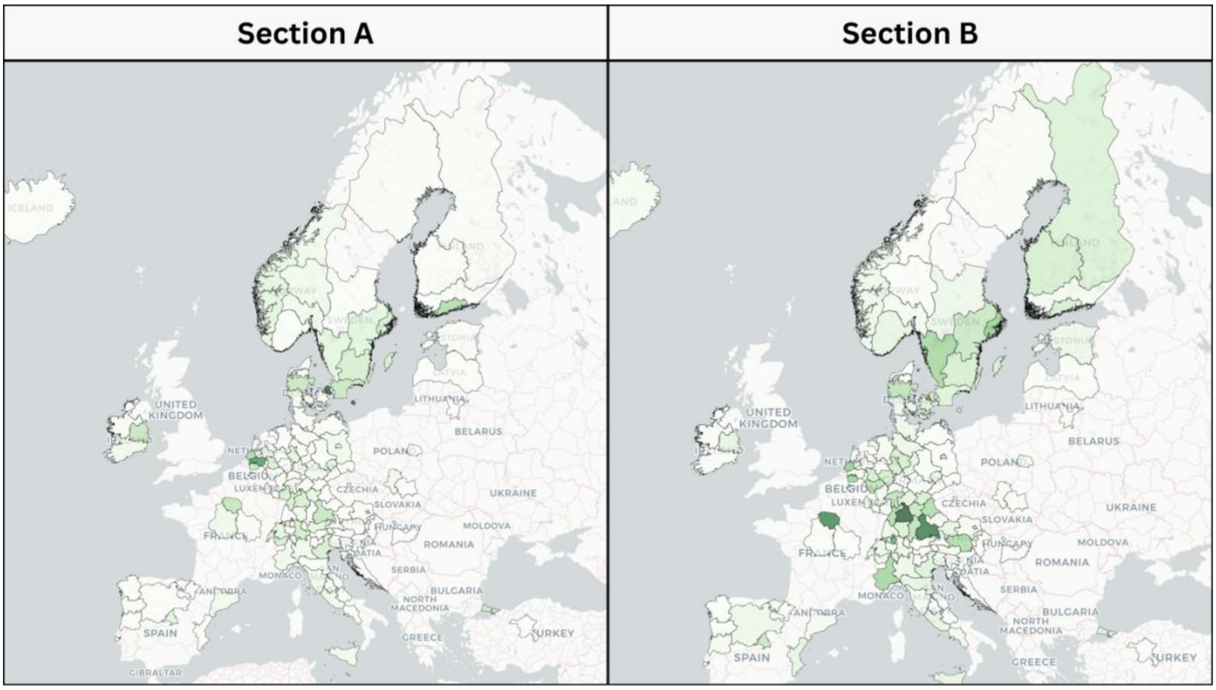

Figure 13. Regional Activity for IPC Sections A and B

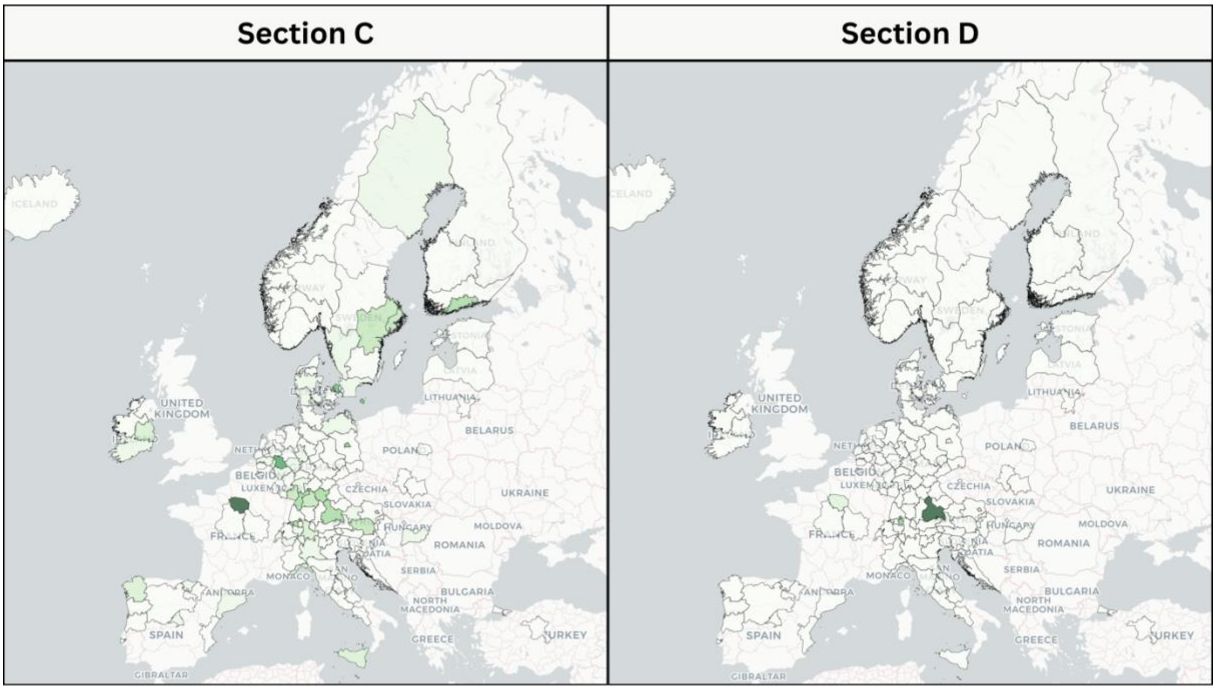

Figure 14. Regional Activity for IPC Sections C and D

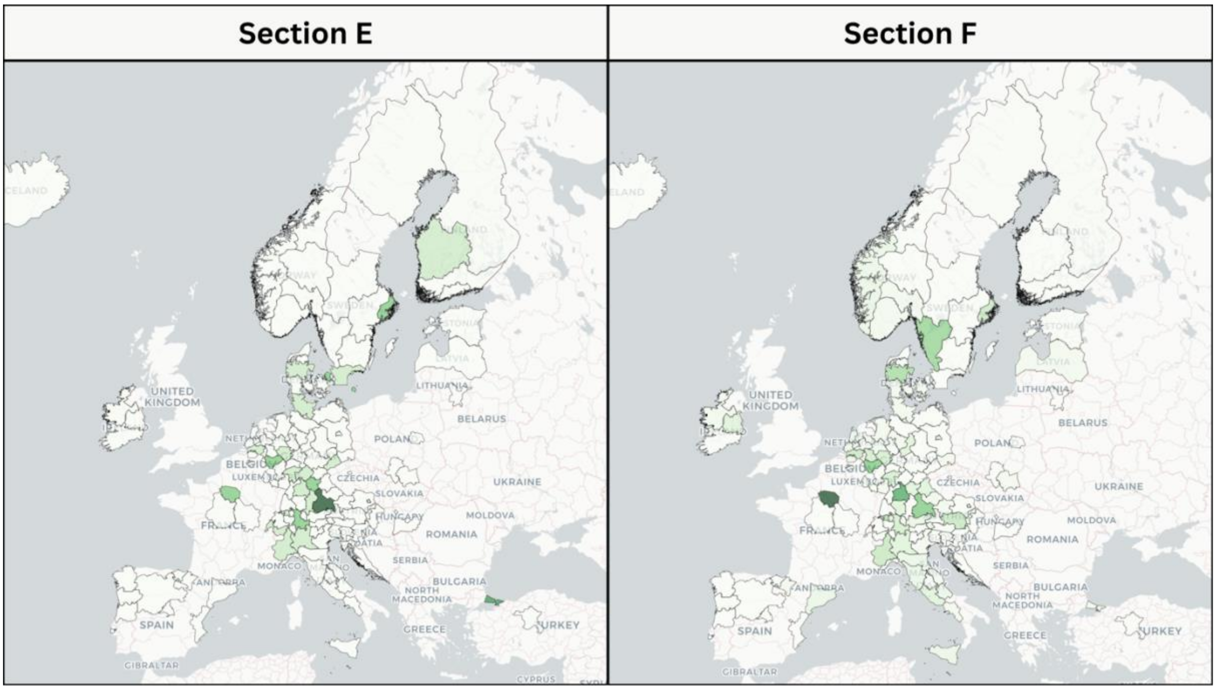

Figure 15. Regional Activity for IPC Sections E and F

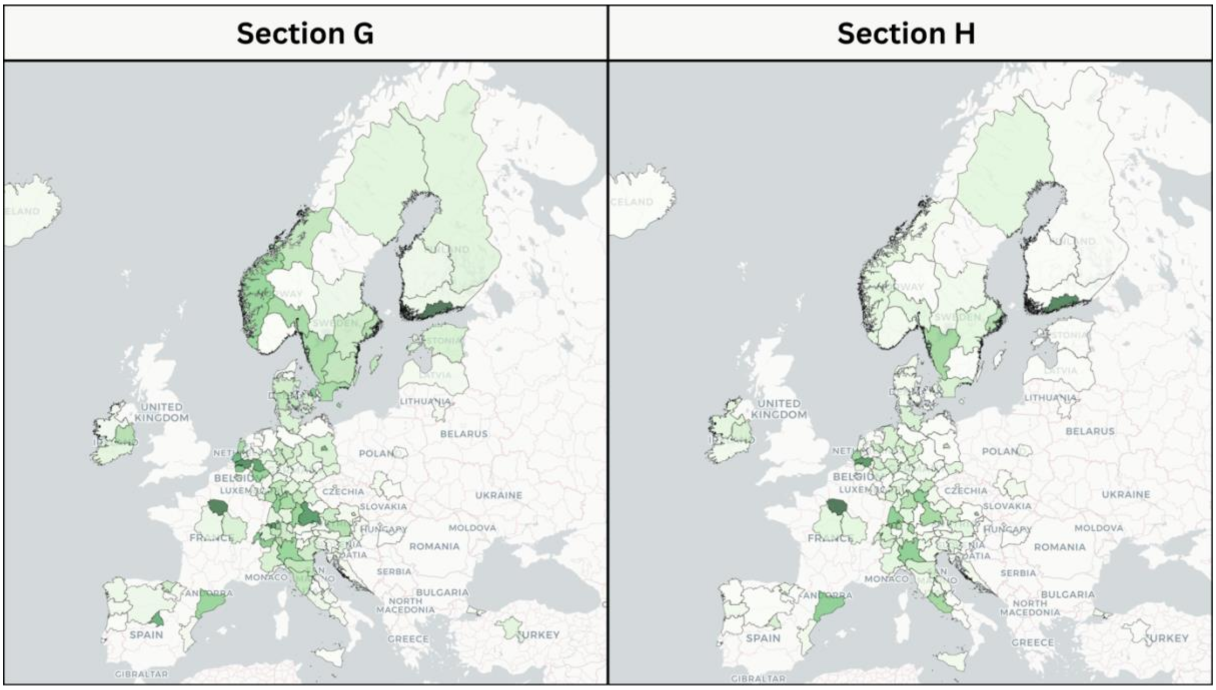

Figure 16. Regional Activity for IPC Sections G and H

Based on the maps, the most active regions for each of the sections and the number of classes with RTA were defined:

- **Section A:** DK01 (13), NL41 (10), NL36 and FI1B (5).
- **Section B:** DE11 (21), DE21 (19), FR1 (18), CH04 (14).
- **Section C:** FR1 (10), DEA1 (7), DK01 (6).
- **Section D:** DE21 (6), CH04 (3), FR1 (1).
- **Section E:** DE21 (4), TR10 (3).
- **Section:** FR1 (18), DE11 (12), DE21 and DEA2 (9).
- **Section G:** FI1B (18), FR1 (17), NL41 and CH03 (16), DE21 (15).
- **Section H:** FR1 (13), FI1B (12), NL41 (11).

Overall, a higher total number represents the breadth of specialization: regions with a higher number of technological classes in a section where they have RTA show that they cover more technologies, which means higher diversification. The findings show that regions are the most active in sections B, F, G, and H.

FR1 appears in most of the regions, which shows a technological leadership in several areas. Regions in Germany mostly specialized in transporting, performing operations, mechanical engineering, and physics. Dutch regions are strong in human necessities, physics, and electricity.

Besides the connection between the sections, network analysis was conducted to find connections between classes in different sectors. The network with all the links was not representative and readable; that is why it was decided to filter the edges by weight (relatedness). Pairs with relatedness > 0.5 were selected.

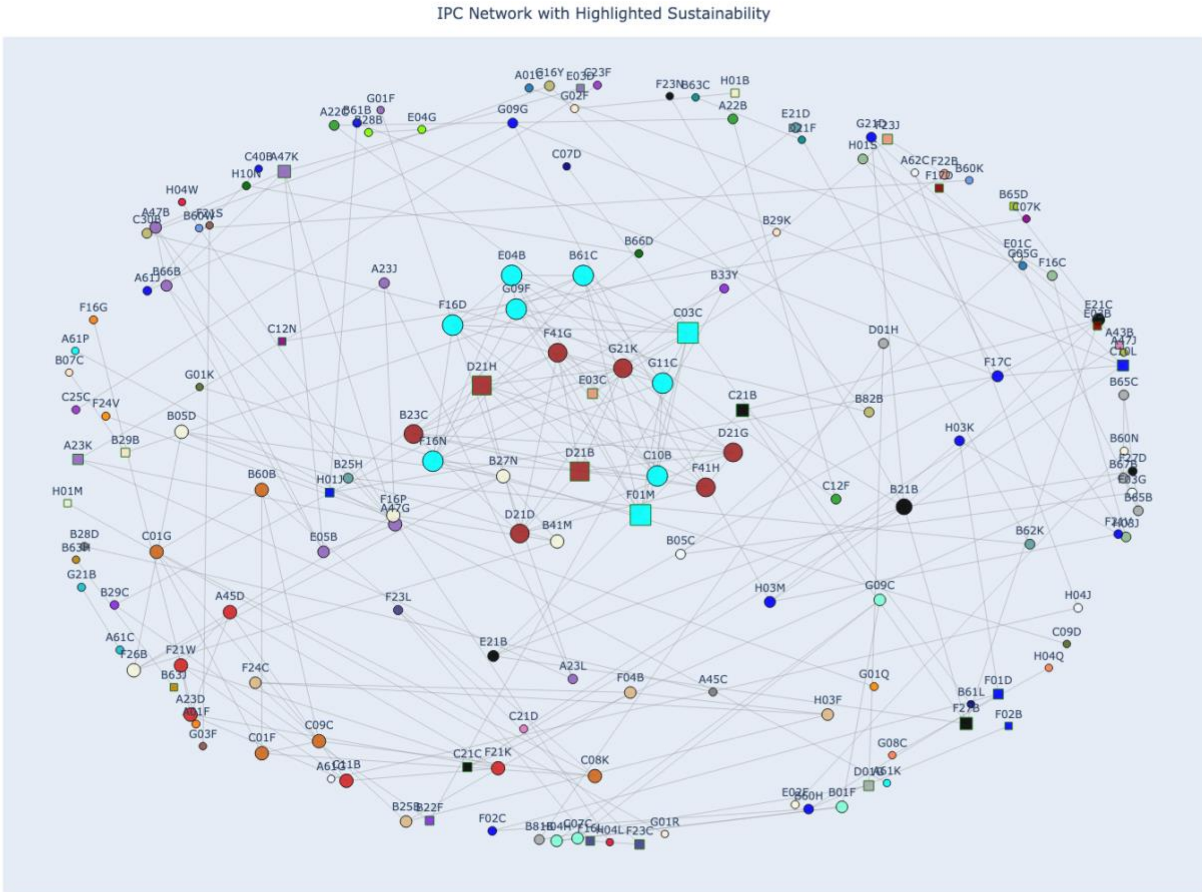

Figure 17. Network Analysis of IPC Classes

The network represents all the shortened IPC classes that have high relatedness. 46 clusters were detected by the Louvain Method of community detection. Different colors represent different clusters. The size of the nodes depends on the strength: stronger classes are bigger.

To highlight the IPC codes that can be related to sustainability, it was decided to present them as squares. Clusters represent the groups where IPC classes are closely related because they co-occur more often in patent filings. If nodes in the cluster are close to each other, it shows a stronger relationship. On the other hand, if the nodes are far from each other but still belong to one cluster, it means that the technological classes are more isolated.

There are two outstanding clusters, where most of the nodes are concentrated in one area: turquoise and dark red. Besides that, the selected cluster nodes are mostly big, which shows the overall strength of the nodes separately. Such nodes can be defined as bridges since they connect multiple IPC classes, which shows a broad application of the technologies. Sustainability-related IPC classes might be presented in most of the clusters, which highlights that sustainability is not isolated and is used with many other technologies. The possible presence of “sustainable” IPC codes as the central elements and inclusion into strong clusters reflects the connection with other technologies.

4.5 Analysis of the Most Active Applicants in Europe

Based on the information obtained after applying the filter for sustainability, shares of all the applicants in Europe were calculated. Based on the values, descriptive statistics and the ranking were created.

Initially, the total number of applicants that are contributing to the patenting activity in AI and sustainability together was 1221. After transforming all the names into uppercase, the total number of unique applicants became 1114. It means that some of them were mentioned at least twice. For example, pairs of Siemens Aktiengesellschaft and SIEMENS AKTIENGESELLSCHAFT, ROBERT BOSCH GMBH and Robert Bosch GmbH. The transformation allowed to achieve case-insensitivity. Full descriptive statistics for the updated applicants’ names are presented in the table.

	Absolute Share
Mean	4.34
Std	18.22
Minimum	0.07
Maximum	462.46
Q1	1.00
Q2 (Median)	2.00
Q3	3.00
IQR	2.00

Table 9. Descriptive Statistics for Applicants

The descriptive statistic for applicants in Europe shows high variance since the standard deviation is much higher than the mean value. A low value of interquartile range of only 2.00 represents a low variability among 50% of the applicants. At the same time, the maximum value is extremely high – 462.46. It shows that there is at least one applicant that has a high share of the market.

The top 10 applicants were selected for the representation: both the absolute share and relative share to analyze the % of the market they have.

Applicant	Absolute Share	Relative Share, %
SIEMENS AKTIENGESELLSCHAFT	462.46	9.57
VESTAS WIND SYSTEMS A/S	173.00	3.59
ROBERT BOSCH GMBH	171.00	3.54
TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)	163.00	3.37
ABB SCHWEIZ AG	149.41	3.09
SIEMENS GAMESA RENEWABLE ENERGY A/S	120.00	2.48
SIGNIFY HOLDING B.V.	97.00	2.01
SCANIA CV AB	75.00	1.55
HUSQVARNA AB	62.00	1.28
VOLVO TRUCK CORPORATION	51.00	1.06

Table 10. Top 10 Leaders in Applications

Before making the adjustments to achieve case-insensitivity, some of the applicants in the top 10 were repeated. Currently, all the names are mentioned only once. As can be seen from the table, the most active applicant in Europe is SIEMENS AKTIENGESELLSCHAFT with the absolute share of 462.46 and relative share of 9.57%. SIEMENS AKTIENGESELLSCHAFT is covering almost 10% of all the applications in AI and sustainability. The next 4 leaders are VESTAS WIND SYSTEMS A/S, ROBERT BOSCH GMBH, TELEFONAKTIEBOLAGET LM ERICSSON (PUBL), and ABB SCHWEIZ AG with the shares from 3.09 to 3.59. These applicants are included in the top 25% (above Q3). The rest of the applicants included in the top 10 have a relatively small share in comparison with SIEMENS AKTIENGESELLSCHAFT, which confirms the descriptive statistics presented earlier – most of them are located between Q1 and Q2, which means that their shares are lower than the median.

4.6 Regional Collaboration

On a regional level, it was decided to conduct an analysis of the collaborations between regions in terms of applicants. The collaboration means that two regions participated in one patent application. Based on the methodology, a co-occurrence matrix and a heatmap were created. The initial matrix has 147 regions, where all the values in the matrix are integer numbers. Since most of the pairs did not have collaborations, it was decided not to use normalization. The heatmap created for the matrix is presented below:

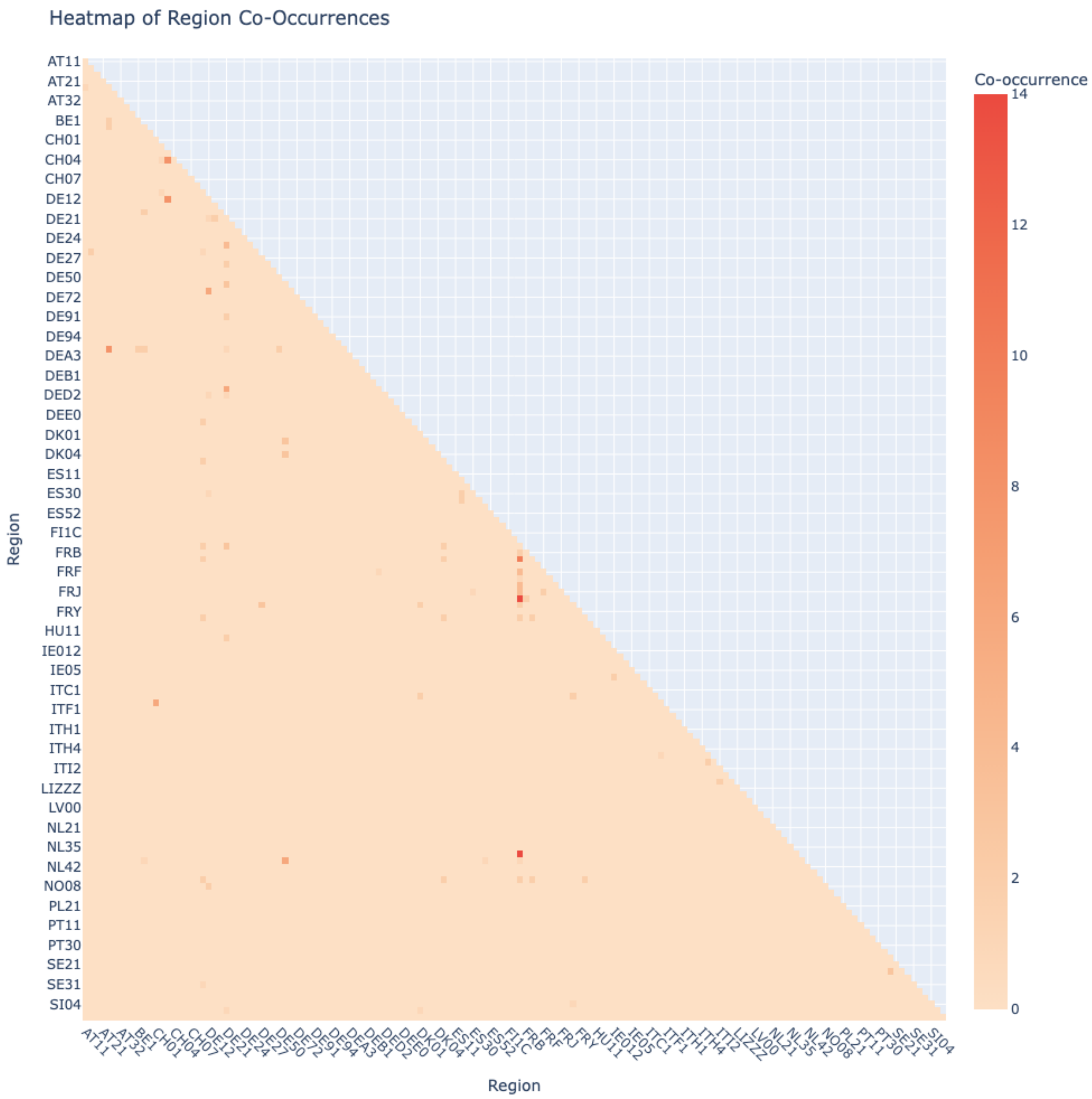

Figure 18. Heatmap for Regional Co-occurrence

As can be seen from the graph, co-occurrence for most of the regions is equal to zero. At the same time, several regions co-occurred in the same patents more often. The strongest pairs are FRK and FR1, NL36 and FR1, with the co-occurrence value of 14 in both cases. For a better analysis, descriptive statistics were created based on the long format of the matrix: a table with columns Region 1, Region 2, and Count.

The total number of possible pairs is $\frac{147*146}{2} = 10731$, because symmetrical pairs were eliminated. It was found that out of 10731 possible collaborations, only 79 are not equal to 0. Only 79 pairs were included in the descriptive statistics.

	Count
Mean	2.76
Std	2.58
Minimum	1.00
Maximum	14.00
Q1	1.50
Q2 (Median)	2.00
Q3	2.50
IQR	1.00

Table 11. Descriptive Statistics for Region Co-occurrence

Based on the numbers, it can be said that since the mean is greater than the median, the data is positively skewed. The standard deviation of 2.58, which is close to the mean, shows that the data has high variability. An interquartile range that equals 1.00 represents that half of the values are located between Q1 = 1.50 and Q3 = 2.50. In addition to that, based on the formula for outliers, it can be said that outliers have values greater than 4.00. Values that are equal to 14 are considered as extreme outliers.

To analyze the strongest collaborations, the top 10 strongest pairs are represented in the table below:

Region 1	Region 2	Count
FRK	FR1	14
NL36	FR1	14
FRC	FR1	10
AT22	DEA2	8
CH03	DE12	8
CH04	CH03	8
DE12	DE71	6
NL41	DE60	6
CH01	ITC4	6
DEC0	DE21	6

Table 12. Top 10 Strongest Collaborations of Regions

All the strongest collaborations are outliers in the initial dataset. FR1 appears 3 times in the top 3 collaborations, which shows a strong network in the region. This region was already classified as one of the most active regions in a previous part. FR1 collaborates with the other French regions and with the Netherlands. German regions collaborate with other German regions, as well as regions in Austria and the Netherlands. Switzerland collaborates with itself, Germany, and Italy. Regions that are collaborating with FR1 will be considered for further analysis.

In order to get information not only about separate pairs, but see the overall picture, a network representation was created.

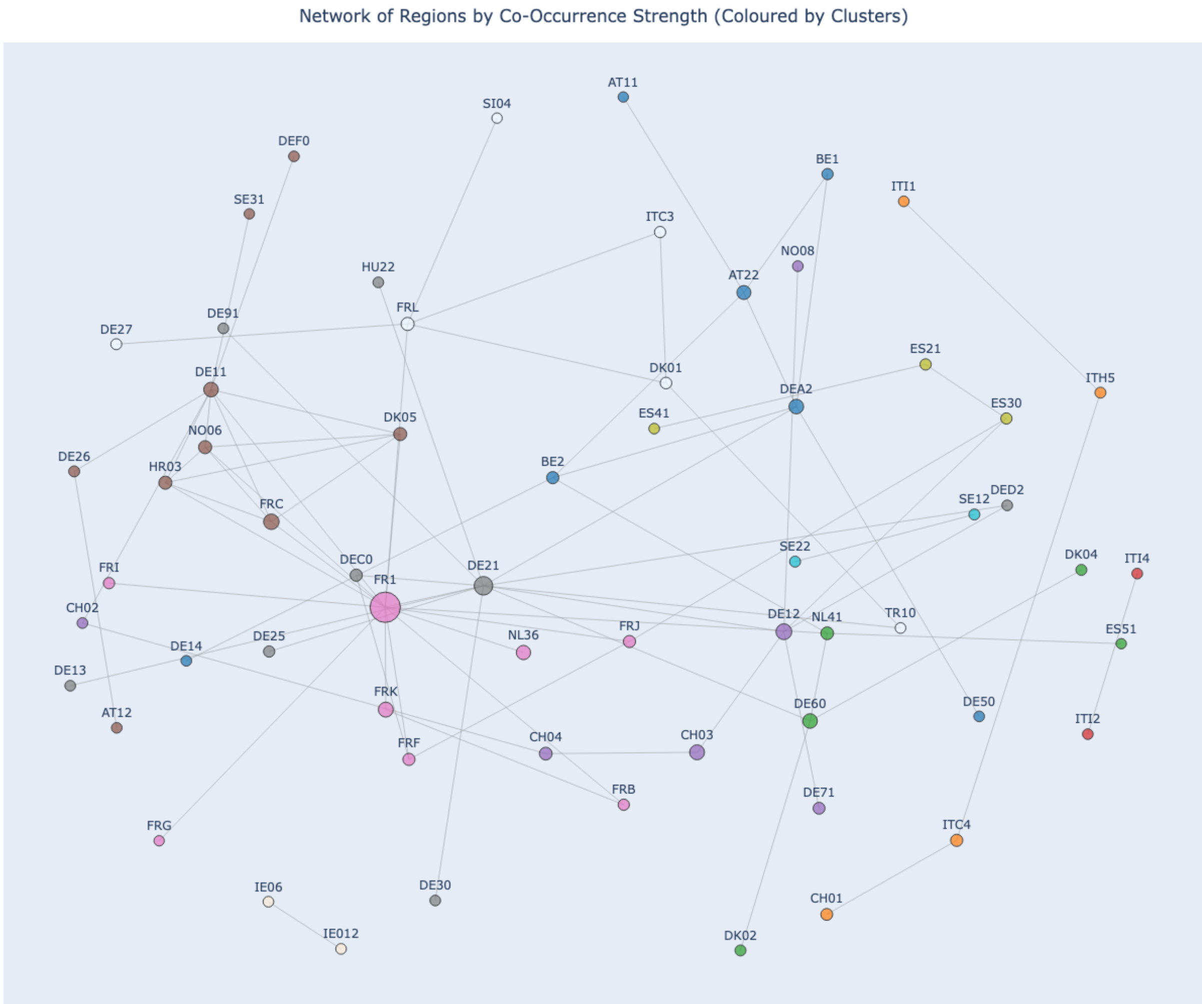

Figure 19. Network Analysis of Regional Co-occurrence

Nodes of the graph represent different NUTS2 regions, while the size of the graph shows the strength of each region. Since the number of regions that have collaborations is not high, all of them are presented on the graph. Different colors represent different clusters that were defined using the Louvain method of community detection.

Overall, 12 clusters were found. The distance between the nodes is based on the weight: the more times regional collaboration was detected, the closer the nodes are. Network representation shows that one of the strongest clusters is a pink cluster that combines FR1, FRK, FRF, FRJ, FRI, and NL36, which shows the collaboration across regions in France, as well as inter-regional collaboration of France and the Netherlands. Brown and green clusters mostly consist of regions in Germany and Nordic countries, which highlights a collaboration of the

Northern part of Europe. Grey cluster has German regions, except for one, that shows a strong collaboration across the country. There are also some separate clusters: white (IE06 and IE012), red (ITI4 and ITI2), and orange (CH01, ITC4, ITH5, ITI1). The separation from the central parts shows that the regions are isolated, and collaborations there are less frequent.

4.7 Collaboration Analysis on Country-Region Level

During the data preparation and first steps of the analysis, it was noticed that there are several cases of collaboration between European countries and other countries. It was decided to analyze such cases in order to define which European regions can be considered as "bridges" to other parts of the world.

For the data representation, it was decided to use a table that has 3 columns: Non-European Country (only 1 value – country code), NUTS2 regions (list of all regions that are collaborating with the particular country), Count collab (number of unique NUTS2 codes with which the country collaborates). During the first iteration, it was noticed that 2 regions, CHZZZ and IEZZZ, were included in the NUTS2 regions column. According to the document with all the changes and details about NUTS2 codes provided in *"History - NUTS - Nomenclature of territorial units for statistics - Eurostat," n.d.*, CHZZZ and IEZZZ are extra-regions. In *"Regulation - 1059/2003 - EN - EUR-LEX," n.d.*, it is stated that extraregio territory are "made up of parts of the economic territory that cannot be attached to a certain region (air-space, territorial waters and the continental shelf, territorial enclaves, in particular embassies, consulates and military bases, and deposits of oil, natural gas, etc. in international waters, outside the continental shelf, worked by resident units)". The indicated regions will also be included in the collaboration analysis. Based on the obtained results, there are several countries that are included in the analysis: Australia, Bermuda, China, India, South Korea, Russia, Singapore, the United Kingdom, Brazil, Japan, and the United States of America. To see which regions are the most active in collaboration, an appropriate table and graph were created.

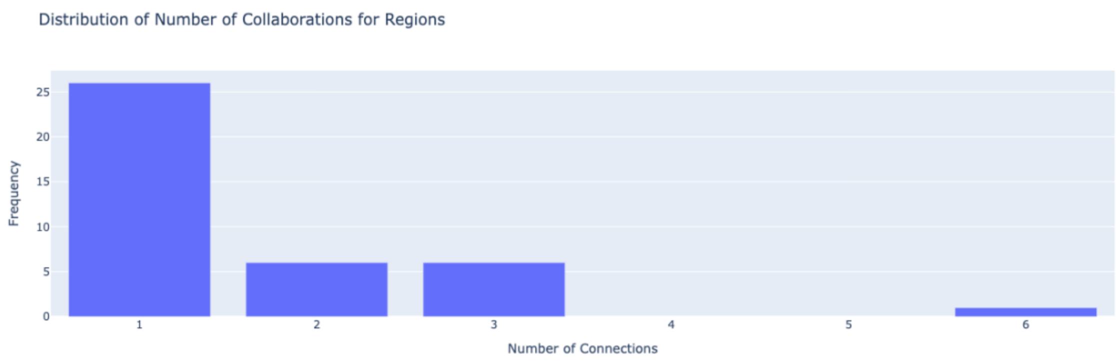

Figure 20. The Number of Connections for Regions

As can be seen from the graph, most of the regions (26) that have any connection with countries outside of Europe have access to only one country. 6 regions have connections with 2 countries, 6 regions have connections with 3 countries, and only one region collaborates with 6 countries out of 12 that were mentioned before. To see how all the regions collaborate the countries, the network was created.

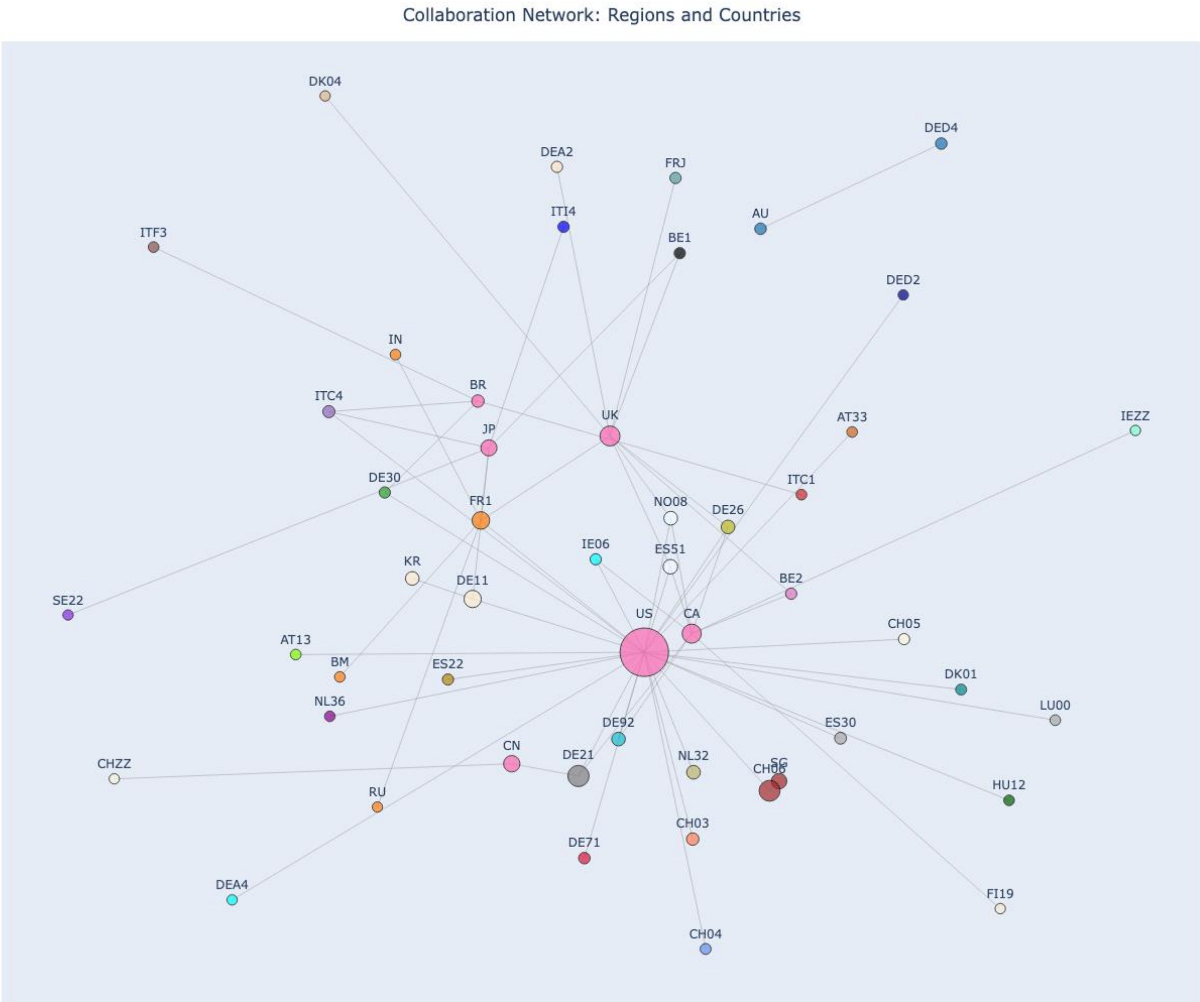

Figure 21. Collaboration of European Regions and Non-European Countries

All the European regions have their own colors. If a country collaborates only with one region, it has the same color as the region. At the same time, if a country has connections with multiple regions, it has a bright pink color. The size of the node is based on the number of collaborations it has. Bigger nodes show the hubs, both in European regions and non-European countries. The links represent only direct connections between regions and countries, since connections between European regions were analyzed separately. As can be seen from the network, the US has the highest number of connections, which shows an intense and diverse collaboration with Europe. From its centrality and size, it suggests there is a bridging function for knowledge transfer across Europe. Countries such as the UK, Japan, Brazil, Canada, and China also have multiple collaborations. Some countries are connected only to one region in Europe, which means that there is only a bilateral collaboration. There is one case where both the European region and a non-European country are connected only to each other: the DED4 region and Australia.

4.8 Regional Analysis

4.8.1 Selection of the Regions

After analyzing the patenting activities of all regions, collaborations between regions, and collaborations on the country-region level, only the most outstanding regions were selected:

- **Patenting Activity (Absolute Share in AI + Sustainability):** FR1, DE21, DE11, SE11.
- **Patenting Activity (Relative Focus on Sustainability):** DK02, DK04.
- **Collaborations on a Regional Level:** pairs FRK – FR1, NL36 – FR1, FRC – FR1.
- **Collaborations on Country-Region Level:** FR1, DE26, ITC4, NO08, DE11, DE21, ES51.

5 regions out of the mentioned ones will be analyzed in more detail. FR1 (Ile-de-France) is the one that is included in most of the categories, DE21 (Upper Bavaria) and DE11 (Stuttgart) are included in 2 categories. DK02 (Zealand) and DK04 (Central Denmark Region) have a high relative focus on sustainability, which is important in the research. In this part, the detailed analysis of the mentioned regions will be conducted.

According to the *"Visualisations - Science, technology, and innovation - Eurostat," n.d.*, all three are active in patenting.

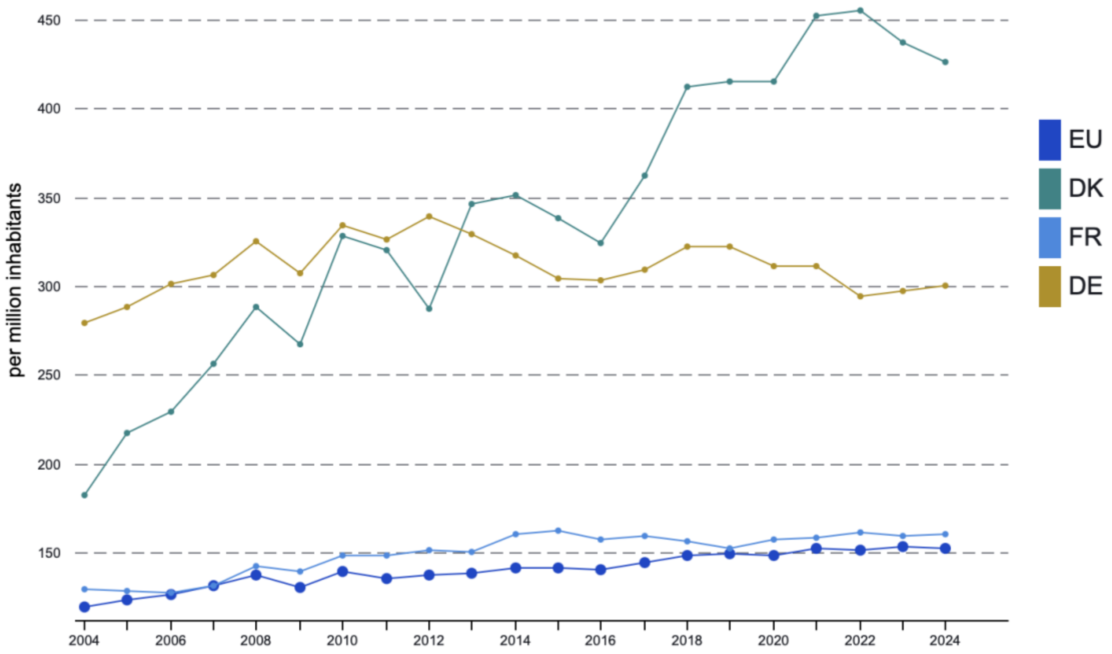

Figure 22. Number of Patent Applications to the European Patent Office

The graph shows not the total number of applications, but the number of applications per million inhabitants. As can be seen from the graph, all the countries are outperforming Europe in general. At the same time, while Île-de-France is a leader in absolute values, France has the lowest number of patents per million inhabitants at the country level. The reason for such statistics is the overall activity of the regions in countries and the population. According to *"Demography of Europe – 2024 edition - Interactive publications - Eurostat," n.d.*, in 2024 population of the selected countries was: 5961249 inhabitants in Denmark, 83456045 inhabitants in Germany, and 68467362 inhabitants in France. Denmark is the country with the fewest people, but most of the regions are relatively active. Germany is the biggest country in terms of population among the chosen ones, and it is 2 times more active than the average value in Europe. It is caused by the fact that the majority of regions in Germany are more active than regions in other countries. As for France, Ile-de-France is the only region with outstanding performance, which is why even with a high absolute share, the relative value of the number of applications per million inhabitants is lower than in the two other countries selected.

4.8.2 Ile-de-France (FR1)

The first region for the analysis is Ile-de-France (FR1). The first aspect to consider during the analysis is R&D. A high level of R&D is usually found in regions with many research universities, start-up companies, and science parks. Paris can be considered as one of the major business and university hubs, which contributes to the high level of R&D and spillover effect. One of the biggest research hubs is Paris-Saclay: it combines seven research institutes, three universities, and nine prestigious higher education institutions in France. According to *"Paris-Saclay," n.d.*, the hub has 21% of French academic research and 15% of private R&D in France since the hub combines both public and private sectors.

In 2021, Ile-de-France had the highest level of R&D expenditures, 21.7 billion euros in total. For better representation, R&D expenditures per inhabitant can be used. In this case, Ile-de-France has 1755 euros per inhabitant, while the average value for Europe is 740 euros. The amount of money spent on R&D can be used to calculate the intensity of research and development. It is a vital element in identifying the innovation capacity of the region and can be calculated as $\frac{R\&D\ expenditures}{GDP}$. The value of R&D intensity in Europe is 2.27%. Ile-de-France has an intensity of 2.83%, which is relatively high in comparison with the average for Europe, but it is still lower than for many other regions.

Another factor that is influencing the patent activity of the region is employment. High-skilled labor is considered an important aspect since it is known to affect the productivity and clustering of research and technology activity. In 2023, 123 million people from 15 to 74 years old were classified as human resources in science and technology. Ile-de-France has 2.4 million people (equated to 39.3% of the labor force) in this category, which is around 2% of the group. In 2023, Ile-de-France had 469.1 thousand people employed in the high-technology sector, which is the highest number in Europe. The relative share of R&D personnel in the region is 3.31%.

Besides the overall patenting activity of the region, the development of AI in the region is important to mention. One of the initiatives that has contributed to this aspect is Hi! PARIS – collaboration of top universities in Paris, such as Institut Polytechnique de Paris, HEC Paris, and Centre Inria de Saclay *("Hi! PARIS – Paris Artificial Intelligence for Society & Business," 2025*). It is the center that specializes in data analytics and AI. In 2024, the center was recognized as an AI Cluster by the French government. In its vision, the organization shows the importance of artificial intelligence and data science not only for business development, but also for society and sustainability in the following areas: health, energy transition, justice, and equity.

Ile-de-France is one of the strongest regions in Europe in patenting and the strongest one in France. High R&D expenditures, big research hubs, such as Paris-Saclay, a high percentage of high-skilled workers involved in science and technology, not only make the region one of the most active in patenting, but also foster collaborations of the region with other European regions and countries outside of Europe. The presence of centers specialized in AI confirms the trend in patenting activity in AI and sustainability.

4.8.3 Stuttgart (DE11)

Stuttgart is another major business and university hub. As for R&D expenditures, it has the second place after Ile-de-France, with an amount of 15.5 billion euros (3742 euros per inhabitant). Stuttgart has an R&D intensity of 6.81%, which makes the region a leader for this parameter. Relative share of R&D personnel – 4.03%. Human resources in science and technology in 2023 were 611 thousand (25.8%). 154 thousand people are employed in the high-technology sector.

According to the findings in the Global Innovation Index 2024, provided by the *World Intellectual Property Organization. et al., n.d.*, there is a strong connection between the high performance in innovations and the presence of science and technology clusters (S&T). Provided statistics show that Germany is the third place in the world with 8 clusters in total, and one of the biggest ones is Stuttgart.

Stuttgart has a high concentration of universities and companies active in AI and sustainability in general, and one of them is Robert Bosch GmbH, that is defined as one of the most outstanding applicants for AI and sustainability. In addition to that, the presence of universities active in research and innovation contributes to the regional activity. One of the research leaders is the Max Planck Institute for Intelligent Systems. The institute specializes in artificial intelligence and robotics and uses the principles in various applications (*Williams, n.d.*).

Another institution is KI-Fortschrittszentrum »Lernende Systeme und Kognitive Robotik«, which collaborates with industrial companies and other organizations for the AI development and application (*“ki-fortschrittszentrum,” n.d.*). The vision of the organization emphasizes the importance of responsible use of artificial intelligence not only for business success but also for societal benefits.

Stuttgart combines high R&D expenditures, concentration of innovative companies and institutions, with an interest in AI and sustainability, which makes the region active in patenting and innovations.

4.8.4 Upper Bavaria (DE21)

Upper Bavaria is close to Stuttgart in R&D spending – it has the third place with 13.2 billion euros. R&D expenditure per inhabitant in this case is 2807 euros. Upper Bavaria’s R&D intensity is 4.61%, which places the region in the top tier for this indicator. Relative share of R&D personnel: 3.49%. As for the human resources in science and technology, the region has 835 thousand people, and the share of the labor force is 30.6%. 250.6 thousand people are employed in high-tech sectors. Upper Bavaria has the first place in the ranking of clusters in Germany, which confirms the innovation capabilities of the region.

There are several companies in Upper Bavaria that are included in the European Digital Innovation Hubs Network. The goal of EDIH Network is to create a community to foster innovations, collaboration, digitalization, and development of different stakeholders, from SME to the public sector *(“European Digital Innovationc Hubs,” n.d.)*. One of the organizations that participates in the initiative is Bayern Innovativ (“European Commission, official website,” 2025). The agency concentrates on transferring knowledge and innovations, and it supports local companies and start-ups. The important trait of Bayern Innovativ is its focus on sustainability *(“Bayern innovativ,” n.d.)*.

Another outstanding organization that is active in innovation is Fraunhofer-Gesellschaft. It is one of the world's leaders in applied research. The organization indicates the importance of AI and new technologies for the research, as well as a commitment to sustainability through responsible development for society and environment (“Homepage Fraunhofer-Gesellschaft,” n.d.).

As it was mentioned before, the absolute leader among applicants is Siemens Aktiengesellschaft, which is located in Munich. The presence of the strongest applicant is contributing to the overall activity of the region.

A combination of high R&D expenditures and R&D intensity, the presence of strong applicants, and multiple innovation hubs make Upper Bavaria one of the strongest regions in patenting activities in AI and sustainability.

4.8.5 Central Denmark Region (DK04)

For the regions in Denmark, no information on a regional level about R&D spending per inhabitant and intensity was provided. Nevertheless, for Denmark in general, R&D expenditures per inhabitant are 1622 euros, and R&D intensity is 2.76. Compared to other regions selected for the analysis, the numbers are not high. The reason why regions are interesting for the analysis is that they are the most outstanding ones due to the high relative focus on sustainability.

According to *The geography of green innovation hubs in OECD regions, 2024*, Denmark shows a strong specialization in green technologies. The publication emphasizes that despite the fact that some small regions may not be active in patenting compared to other strong regions, they can be highly specialized. In 2021, the Central Denmark Region was 19th in the ranking for the total number of green patents, but in other aspects, it had only 92nd place.

The region has the potential to become a green innovation hub, and its success can be explained by the infrastructure adapted for R&D in green innovations. One of the examples is GreenLab. Green Lab is an industrial park that combines companies with a focus on green transition *("About GreenLab - a green and circular industrial park | GreenLab," n.d.)*. The park specializes in energy systems and energy solutions. GreenLab Academy is one of the parts of the park that is created for knowledge and experience sharing.

In of the strongest applicants that was defined in the analytical part, Siemens Gamesa Renewable Energy A/S, has its headquarters in Ikast-Brande Municipality in Central Denmark Region. The company emphasized its commitment to sustainability through the development of solutions for the wind industry.

4.8.6 Zealand (DK02)

Zealand is another region in Denmark that is focusing on sustainability. The region is included in the Greater Copenhagen (a combination of cross-border regions of Sweden and Denmark), which is considered as one of the leaders in green innovations. The area has one of the 6 cross-municipal businesses.

One of the organizations that is contributing to sustainable development is Kalundborg Symbiosis. The goal of the company is to combine industrial companies for resource sharing and foster innovations and collaboration (*Kalundborg Symbiosis, 2025*). Kalundborg Symbiosis is an example of the coexistence of sustainability and profit. According to *Valentine, 2016*, Kalundborg Symbiosis can be considered as the first industrial ecology network.

Another association that is contributing to the development of the region is Knowledge Hub Zealand, which specializes in biotech development (*Knowledge Hub Zealand, 2025*). The organization combines education and innovation for biosolutions by bringing together researchers, educational initiatives, SMEs, and start-ups.

The Zealand region in Denmark represents how collaborations between companies and a focus on sustainability can foster sustainable development and innovations.

5. Conclusion

This thesis analyzed the relationship between artificial intelligence and sustainability in terms of innovations. Emerging interest in how AI can enable sustainable development is stipulated by the deterioration of environmental, economic, and social issues.

The findings provide valuable information about important collaborations and the most outstanding regions. Currently, for both aspects, there is a rising trend: sustainability issues are becoming more and more urgent, and at the same time, AI is becoming more advanced.

First, the main definitions were presented in order to give a base for further analysis. Artificial intelligence can be used in various ways, such as neural networks, machine learning, natural language processing, and computer vision. The important feature of AI is that it can be applied in the majority of fields and industries, from production to medicine. The concept of sustainability emphasizes the importance of concentrating on future generations while meeting the needs of the present generation. Sustainable Development Goals (SDGs) represent the issues common to both developed and developing countries.

The literature review part gives an overview of how AI can be used to enable sustainability. The findings show that the presence of developed AI can foster sustainable development in many areas. AI can be used to improve lives and achieve the SDGs. New technologies help to reduce greenhouse gas emissions by process optimization, defining sources of air pollution, preventing crimes, improving the safety of citizens, improving the healthcare system through telemedicine and prevention of severe diseases, managing the transportation systems to reduce traffic and the number of car accidents. Artificial intelligence is improving over time, and it finds more and more applications. On the other hand, there are several threats that appear due to the rapid growth and development of AI. Excessive usage of data and new algorithms leads to overconsumption of resources, higher greenhouse gas emissions, as well as ethical issues, such as discrimination against particular groups of people due to data bias, the problem of data protection, and information privacy. Mitigation of the risks can be achieved by tackling the problem from a different perspective: raising people's awareness about the sustainability issues and overconsumption of resources, companies should evaluate the impact of their activities, and the regulations must be created at the governmental level to provide a legal base. The combination of digital and green transition is defined as the twin transition, and the key success factors were defined in the literature review part.

The theoretical part highlights the importance of collaboration at the intra-region and inter-region levels. It shows a positive effect on innovation efficiency and the spillover effect. Science and technology parks are also considered as important factors that can influence the innovative capacity. The findings were vital, since the regional collaborations were analyzed in the practical part.

It was defined that patents can be used to measure innovation. They present information about the technicalities of the invention, legal information, business-relevant information, and information relevant for public policy. Such detailed data can be valuable to define patent value, analyze trends for specific technologies, find the patenting activity of applicants and inventors, as well as for specific regions and countries. Moreover, patent information, specifically classification codes, can be a valuable source for insights not only about the separate trends in technologies, but also regarding the potential relationships between technological classes.

In the second part, the methodology for the practical part was presented. Python programming language was chosen to work with the data from different sources. All the Python modules and their purposes were defined. First, data was extracted from the Orbit database, and then it was combined with the REGPAT database to get the complete picture and all the necessary information about regions, classifications, and applicants. The chapter explained patent classifications and region classifications that were used during the analysis. Methodology for descriptive statistics, co-occurrence of regions in the same patent, relatedness of the technological classes, and applicants' evaluation was presented. For visual representation of the findings, tables, heatmaps, networks, maps, and bar charts were used. For co-occurrence and relatedness analysis, networks were used, including cluster identification using the Louvain Method. Several limitations appeared in the practical part: missing information, no normalized data about applicants, and no uniform dictionary that can be used for the search of AI-related patents.

The practical part includes both quantitative and qualitative analysis. First, quantitative information obtained after data processing was analyzed. Based on it, the regions for the deeper analysis were selected where qualitative data was considered.

The analysis shows that the number of patents in AI and both in AI and sustainability is increasing every year. Descriptive statistics for patents related to AI and for patents related both to AI and sustainability have shown that in all cases, data is positively skewed, which means that the average performance across countries and regions is relatively low, but the is a presence of outliers, or active parties. Similar results were obtained for applicants: there are several strong applicants, such as Siemens Aktiengesellschaft, Vestas Wind Systems A/S, and Robert Bosch GmbH. The connection between the most active regions and active applicants was found: applicants are mostly located in the regions with high patenting activity.

Regional collaboration analysis has shown that most of the regions do not have many collaborations, except for FR1, which can be considered as a leader in this aspect. Analysis of the network and identification of the cluster allowed to get more insights. Overall, 12 clusters were defined, where the strongest ones were France and the Netherlands, Germany and Northern regions, cluster of German regions. Besides that, there were some distant, isolated clusters of two to four regions.

Analysis of the collaborations of European regions with countries outside of Europe has shown similar results – most of the regions do not collaborate at all, while several regions have good connections with the countries. The country that has the highest number of "bridges" to Europe is the United States of America. Ile-de-France has six connections with six different countries, which makes the region one of the strongest collaboration hubs.

Analysis of the relatedness of technological classes gave valuable insights into which of the technological sections mostly appear in the same regions. Revealed Technological Advantage analysis helped to define regions that are specialized in particular technological classes. Based on the values, it was found that overall regions are more active in sections B (Performing Operations; Transporting), F (Mechanical Engineering; Lighting; Heating; Weapons; Blasting), G (Physics), and H (Electricity). The network analysis, including cluster identification, was provided for pairs that have RTA > 0.5. It has been found that there are two strong clusters, where most of the classes are concentrated in one are, which shows a strong connection. It was found that multiple IPC codes related to sustainability are not excluded from the representation, and some of them have central positions. The presence of "sustainable" IPC classes in most of the clusters shows that the technologies are not isolated and are supported by other fields.

Qualitative research shows that innovation intensity in the region can be defined by several factors: R&D expenditures, high-skilled workers, presence of strong educational institutions and research centers, as well as hubs that bring together researchers, start-ups, and universities. Regions selected for the deeper analysis have different peculiarities that make them exceptional:

- **Ile-de-France** has high R&D expenditures and many high-skilled workers, as well as Hi! PARIS collaboration of universities and research hubs. In comparison with other countries, France has only one region with a high patenting activity.
- **Stuttgart** and **Upper Bavaria** are well-developed regions in Germany. Both of the regions are the biggest science and technology clusters. A high number of research universities and high-tech companies that are interested in both AI and sustainability, as well as the presence of the European Digital Innovation Hubs Network participants, have a positive impact on patenting activities in AI and sustainability.

- **Central Denmark Region** and **Zealand,** which are located in Denmark, can be examples of how small regions with relatively low patenting activity in general can have a high specialization in sustainability. Different organizations, such as GreenLab, Siemens Gamesa Renewable Energy A/S, and Kalundborg Symbiosis, show how the support of green technologies and collaboration can be combined to create a better future and foster innovations.

Provided results can be beneficial for multiple fields. First, the findings can be taken into account for innovation strategies and funding, since strong and weak regions, regional clusters, and technological clusters were defined. Collaborations of European regions, as well as European regions and non-European countries, highlight already existing networks and show the potential areas of knowledge spillover and technology diffusion, since collaborations are beneficial for both aspects. Analysis of the strong regions provides insights on what can be improved or implemented in the weaker regions to be more competitive and become stronger, not only in terms of innovations in general, but also in twin transition. Analysis of RTA in different sectors provided valuable insights about strong regions. Based on this information, potential regions for collaborations or possible competitors can be defined.

References

1. About CPC | Cooperative Patent Classification [WWW Document], n.d. URL https://www.cooperativepatentclassification.org/about
2. About GreenLab - a green and circular industrial park | GreenLab [WWW Document], n.d.. GreenLab. URL https://www.greenlab.dk/about/
3. ast — Abstract Syntax Trees [WWW Document], n.d. . Python Documentation. URL https://docs.python.org/3/library/ast.html
4. Bayern innovativ [WWW Document], n.d. URL https://www.bayern-innovativ.de/en/
5. Biggi, G., Iori, M., Mazzei, J., Mina, A., 2025. Green intelligence: the AI content of green technologies. Eurasian Economic Review. https://doi.org/10.1007/s40821-024-00288-1
6. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics Theory and Experiment 2008, P10008. https://doi.org/10.1088/1742-5468/2008/10/p10008
7. Cambridge English Dictionary: Meanings & Definitions, 2025.
8. collections — Container datatypes [WWW Document], n.d. . Python Documentation. URL https://docs.python.org/3/library/collections.html
9. Demography of Europe – 2024 edition - Interactive publications - Eurostat [WWW Document], 2024. Eurostat. URL https://ec.europa.eu/eurostat/web/interactive-publications/demography-2024
10. European Commission, official website [WWW Document], 2025. . European Commission. URL https://commission.europa.eu/index_en
11. European Commission. Statistical Office of the European Union., 2018. Regions in the European Union: nomenclature of territorial units for statistics (NUTS): 2024 edition. Publications Office, LU.
12. European Digital Innovation Hubs [WWW Document], n.d. . Shaping Europe's Digital Future. URL https://digital-strategy.ec.europa.eu/en/policies/edihs
13. Faggian, A., Marzucchi, A., Montresor, S., 2024. Regions facing the 'twin transition': combining regional green and digital innovations. Regional Studies 1–7. https://doi.org/10.1080/00343404.2024.2398555
14. Fan, F., Lian, H., Wang, S., 2019. Can regional collaborative innovation improve innovation efficiency? An empirical study of Chinese cities. Growth and Change 51, 440–463. https://doi.org/10.1111/grow.12346
15. Favot, M., Vesnic, L., Priore, R., Bincoletto, A., Morea, F., 2023. Green patents and green codes: How different methodologies lead to different results. Resources Conservation & Recycling Advances 18, 200132. https://doi.org/10.1016/j.rcradv.2023.200132

16. Feng, F., Li, J., Zhang, F., Sun, J., 2024. The impact of artificial intelligence on green innovation efficiency: Moderating role of dynamic capability. International Review of Economics & Finance 96, 103649. https://doi.org/10.1016/j.iref.2024.103649
17. Folium — Folium 0.20.0 documentation [WWW Document], n.d. URL https://python-visualization.github.io/folium/latest/
18. Foresight, n.d. Aggravating resource scarcity | Knowledge for policy [WWW Document]. URL https://knowledge4policy.ec.europa.eu/aggravating-resource-scarcity_en
19. GeoPandas 1.1.1 — GeoPandas 1.1.1+0.ge9b58ce.dirty documentation [WWW Document], n.d. URL https://geopandas.org/en/stable/
20. Global economy under pressure could slow to 2.3%, signals UN Trade and Development [WWW Document], 2025b. . UN Trade and Development (UNCTAD). URL https://unctad.org/news/global-economy-under-pressure-could-slow-23-signals-un-trade-and-development
21. Hazir, C., LeSage, J.P., Autant-Bernard, C., 2014. The Role of R&D Collaboration Networks on Regional Innovation Performance. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2507284
22. Hi! PARIS – Paris Artificial Intelligence for Society & Business [WWW Document], 2025. URL https://www.hi-paris.fr/
23. High-Level Expert Group on Artificial Intelligence, 2019. A definition of AI: Main capabilities and scientific disciplines: Definition developed for the purpose of the AI HLEG's deliverables. The European Commission.
24. History - NUTS - Nomenclature of territorial units for statistics - Eurostat [WWW Document], n.d. . Eurostat. URL https://ec.europa.eu/eurostat/web/nuts/history
25. Homepage Fraunhofer-Gesellschaft [WWW Document], n.d. . Copyright. URL https://www.fraunhofer.de/en.html
26. International Patent Classification (IPC) [WWW Document], n.d. . Classification-ipc. URL https://www.wipo.int/en/web/classification-ipc
27. itertools — Functions creating iterators for efficient looping [WWW Document], n.d. . Python Documentation. URL https://docs.python.org/3/library/itertools.html
28. Kalundborg Symbiosis, 2025. About us - Kalundborg Symbiose [WWW Document]. Kalundborg Symbiosis. URL https://www.symbiosis.dk/en/om-os/
29. Key findings: Exploring the SDGs through patents [WWW Document], 2024. Patent-analytics. URL https://www.wipo.int/web/patent-analytics/key-findings-exploring-the-sdgs-through-patents

30. Khan, A., 2024. Artificial Intelligence: A Guide for Everyone. Springer Nature Switzerland, Cham. https://doi.org/10.1007/978-3-031-56713-1
31. ki-fortschrittszentrum [WWW Document], n.d. URL https://www.ki-fortschrittszentrum.de/
32. Knowledge Hub Zealand, 2025. Home - knowledgehubzealand.dk [WWW Document]. Knowledgehubzealand.dk. URL https://knowledgehubzealand.dk/en/
33. Mailänder, L., 2016. Topic 8: IPC and CPC Basics. WIPO.
34. Martinelli, A., Mina, A., Moggi, M., 2020. The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution. Industrial and Corporate Change 30, 161–188. https://doi.org/10.1093/icc/dtaa060
35. Matplotlib documentation — Matplotlib 3.5.3 documentation [WWW Document], n.d. URL https://matplotlib.org/3.5.3/index.html
36. McKinsey Global Institute, 2017. A FUTURE THAT WORKS: AUTOMATION, EMPLOYMENT, AND PRODUCTIVITY. McKinsey & Company.
37. McKinsey Global Institute, 2018. SMART CITIES: DIGITAL SOLUTIONS FOR A MORE LIVABLE FUTURE. McKinsey & Company.
38. Measuring environmental innovation using patent data, 2015. OECD Environment Working Papers. https://doi.org/10.1787/5js009kf48xw-en
39. Montoro-Sánchez, A., Ortiz-de-Urbina-Criado, M., Mora-Valentín, E.M., 2011. Effects of knowledge spillovers on innovation and collaboration in science and technology parks. Journal of Knowledge Management 15, 948–970. https://doi.org/10.1108/13673271111179307
40. Muench, S., Stoermer, E., Jensen, K. et al., 2022. Towards a green & digital future : key requirements for successful twin transitions in the European Union. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/977331
41. Nagaoka, S., Motohashi, K., Goto, A., 2010. Patent Statistics as an innovation indicator, in: Handbook of the Economics of Innovation. pp. 1083–1127. https://doi.org/10.1016/s0169-7218(10)02009-5
42. Natural Earth - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales [WWW Document], n.d. URL https://www.naturalearthdata.com/
43. NetworkX — NetworkX documentation [WWW Document], n.d. URL https://networkx.org/
44. NumPy [WWW Document], n.d. URL https://numpy.org/
45. OECD, 2009. OECD Patent Statistics Manual. OECD.

46. OECD, 2025. OECD REGPAT DATABASE. EPO & PCT Patent Applications at Regional Level. OECD.
47. Overview - GISCO - Eurostat [WWW Document], n.d. . Eurostat. URL https://ec.europa.eu/eurostat/web/gisco
48. pandas - Python Data Analysis Library [WWW Document], n.d. URL https://pandas.pydata.org/
49. Paris-Saclay [WWW Document], n.d. . Paris-Saclay. URL https://welcometoparissaclay.com/en/
50. Purvis, B., Mao, Y., Robinson, D., 2018. Three pillars of sustainability: in search of conceptual origins. Sustainability Science 14, 681–695. https://doi.org/10.1007/s11625-018-0627-5
51. pycountry [WWW Document], 2024. . PyPI. URL https://pypi.org/project/pycountry/
52. Regulation - 1059/2003 - EN - EUR-LEX [WWW Document], n.d. URL https://eur-lex.europa.eu/eli/reg/2003/1059/oj
53. Regulation - EU - 2024/1689 - EN - EUR-LEX [WWW Document], n.d. URL http://data.europa.eu/eli/reg/2024/1689/oj
54. Researchers and innovators invited to shape Europe's AI Strategy in Science [WWW Document], 2025. Shaping Europe's Digital Future. URL https://digital-strategy.ec.europa.eu/en/news/researchers-and-innovators-invited-shape-europes-ai-strategy-science
55. Ritchie, H., Rosado, P., Roser, M., 2023. CO_2 and Greenhouse Gas Emissions [WWW Document]. Our World in Data. URL https://ourworldindata.org/co2-and-greenhouse-gas-emissions
56. seaborn: statistical data visualization — seaborn 0.13.2 documentation [WWW Document], n.d. URL https://seaborn.pydata.org/
57. Smart cities [WWW Document], n.d. . European Commission. URL https://commission.europa.eu/eu-regional-and-urban-development/topics/cities-and-urban-development/city-initiatives/smart-cities_en
58. Sobolieva, T.O., Holionko, N.G., Batenko, L.P., Reshetniak, T.I., 2021. Global technology trends through patent data analysis. IOP Conference Series Materials Science and Engineering 1037, 012059. https://doi.org/10.1088/1757-899x/1037/1/012059
59. Stahl, B.C., 2021. Ethical issues of AI, in: SpringerBriefs in Research and Innovation Governance. pp. 35–53. https://doi.org/10.1007/978-3-030-69978-9_4
60. Statista, 2025. AI market size worldwide from 2020-2031 [WWW Document]. Statista. URL https://www.statista.com/forecasts/1474143/global-ai-market-size

61. Tao, M., 2024. Digital brains, green gains: Artificial intelligence's path to sustainable transformation. Journal of Environmental Management 370, 122679. https://doi.org/10.1016/j.jenvman.2024.122679
62. THE 17 GOALS | Sustainable Development [WWW Document], n.d. URL https://sdgs.un.org/goals
63. The geography of green innovation hubs in OECD regions, 2024. , OECD Local Economic and Employment Development (LEED) Working Papers. https://doi.org/10.1787/c48ad2b1-en
64. United Nations Environment Programme (2024) Artificial Intelligence (AI) end-to-end: The Environmental Impact of the Full AI Lifecycle Needs to be Comprehensively Assessed - Issue Note. [WWW Document], n.d. URL https://wedocs.unep.org/20.500.11822/46288
65. Valentine, S.V., 2016. Kalundborg Symbiosis: fostering progressive innovation in environmental networks. Journal of Cleaner Production 118, 65–77. https://doi.org/10.1016/j.jclepro.2016.01.061
66. Visualisations - Science, technology, and innovation - Eurostat [WWW Document], n.d. . Eurostat. URL https://ec.europa.eu/eurostat/web/science-technology-innovation/visualisations
67. Welcome to branca's documentation! — branca 0.8.1 documentation [WWW Document], n.d. URL https://python-visualization.github.io/branca/
68. Williams, J., n.d. About us [WWW Document]. Max Planck Institute for Intelligent Systems. URL https://is.mpg.de/en/about
69. WIPO, 2019. Background paper for WIPO Technology Trends: Artificial Intelligence. WIPO.
70. WIPO, 2025. Guide to the International Patent Classification. WIPO.
71. World Intellectual Property Organization, 2021. WIPO Guide to Using Patent Information.
72. World Intellectual Property Organization., Dutta, Soumitra., Lanvin, Bruno., Rivera León, Lorena., Wunsch-Vincent, Sacha., n.d. Global Innovation Index 2024 :: Innovation in the face of uncertainty. World Intellectual Property Organization,. https://doi.org/10.34667/TIND.50062
73. Yadav, P., Tudela, L.A.M., Marco-Lajara, B., 2024. The role of AI in assessing and achieving the Sustainable Development Goals (SDGs), in: Advances in Finance, Accounting, and Economics Book Series. pp. 1–17. https://doi.org/10.4018/979-8-3693-3282-5.ch001

Appendix

Appendix 1. Data Processing and Preparation

1.1 Selectin EPO and PCT Patents

```
#lists of publication numbers for the patents splitted into EPO and PCT
#different tables and formats will be used for matching with REGPAT database
pub_nbr_wo = []
pub_nbr_ep = []
tab_id_wo = []
tab_id_ep = []
for i in range(num_patents):
    obj = df['Publication numbers'].iloc[i]
    num = obj.split('\n')
    num_sort = sorted(set([p for p in num if (p.startswith('EP') or p.startswith('WO'))]))
    for j in num_sort:
        if (j[:2] == 'EP'):
            pub_nbr_ep.append(j[2:])
            tab_id_ep.append(i)
        else:
            pub_nbr_wo.append(j.replace('/', ''))
            tab_id_wo.append(i)
```

1.2 Merging Two Tables

```
df_epo = pd.merge(epo_patents, epo_app_reg, on="pub_nbr", how="left")
df_epo = df_epo.rename(columns={'appln_id': 'app_appln_id', 'reg_code': 'app_reg_code', 'ctry_code':
'app_ctry_code',
                                'app_name': 'app_name'})
```

1.3 Grouping The Table

```
grouped_df_epo = df_epo_notna.groupby('block_row').agg({
    'pub_nbr': lambda x: list(set(code for sublist in x.dropna() for code in (sublist if isinstance(sublist, list) else
[sublist]))),
    'appln_id': 'first',
    'app_reg_code': lambda x: list(x.dropna()), #keeping all the regions with duplicates
    'app_ctry_code': lambda x: list(x.dropna()),
    'inv_reg_code': lambda x: list(x.dropna()),
    'inv_ctry_code': lambda x: list(x.dropna()),
    'app_name': lambda x: list(x.dropna()),
    'app_year': 'first',
    'IPC': lambda x: list(set(code for sublist in x.dropna() for code in (sublist if isinstance(sublist, list) else
[sublist]))),
```

```
    'CPC_Class': lambda x: list(set(code for sublist in x.dropna() for code in (sublist if isinstance(sublist, list) else [sublist])))
}).reset_index()
```

Appendix 2. Data Analysis

2.1 Graph for Distribution of AI Patents

```
applications_per_year = all_concat['app_year'].value_counts().reset_index()
applications_per_year.columns = ['year', 'count']
applications_per_year = applications_per_year.sort_values('year')

applications_per_year['year'] = applications_per_year['year'].astype(int)

sns.set_theme(style="darkgrid")
plt.figure(figsize=(10, 6))
sns.barplot(data=applications_per_year, x='year', y='count', palette='Blues_d')

plt.title('Number of Applications per Year (AI)')
plt.xlabel('Year')
plt.ylabel('Number of Applications')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

2.2 Calculating Shares for Applicants' Countries for Country-Level

```
app_ctry = []
app_ctry_share = []
inv_ctry = []
inv_ctry_share = []
applicant = []
applicant_share = []

for i in range(all_concat.shape[0]):
   app_codes = all_concat['app_ctry_code'].iloc[i]
   inv_codes = all_concat['inv_ctry_code'].iloc[i]
   appl_names = all_concat['app_name'].iloc[i]

   # Frequency analysis for up_reg_codes_app
   counter_app = Counter(app_codes)
   total_app = len(app_codes)

   counter_inv = Counter(inv_codes)
   total_inv = len(inv_codes)
```

```
    counter_applicant = Counter(appl_names)
    total_applicant = len(appl_names)

    unique_regions_app = list(counter_app.keys())
    reg_share_app = [count / total_app for count in counter_app.values()]

    unique_regions_inv = list(counter_inv.keys())
    reg_share_inv = [count / total_inv for count in counter_inv.values()]

    unique_appl = list(counter_applicant.keys())
    appl_share = [count / total_applicant for count in counter_applicant.values()]

    app_ctry.append(unique_regions_app)
    app_ctry_share.append(reg_share_app)
    inv_ctry.append(unique_regions_inv)
    inv_ctry_share.append(reg_share_inv)
    applicant.append(unique_appl)
    applicant_share.append(appl_share)
contry_share_totals = defaultdict(float)

for regions, shares in zip(app_ctry, app_ctry_share):
    for region, share in zip(regions, shares):
        contry_share_totals[region] += share


country_share_df = pd.DataFrame(
    list(contry_share_totals.items()),
    columns=["Country", "Total_share"])
```

2.3 Calculating Absolute and Relative Shares for AI and Sustainability

```
countries_AI_sust = df_countries[df_countries['Total_share_AI_and_sust'].notnull()]
countries_AI_sust['share_sustainability'] = (countries_AI_sust['Total_share_AI_and_sust'] /
countries_AI_sust['Total_share_AI'])*100

countries_AI_sust[['Total_share_AI', 'Total_share_AI_and_sust', 'share_sustainability']] =
countries_AI_sust[['Total_share_AI',
                                        'Total_share_AI_and_sust', 'share_sustainability']].round(2)

total_AI = countries_AI_sust['Total_share_AI'].sum()
total_sust = countries_AI_sust['Total_share_AI_and_sust'].sum()
```

```
countries_AI_sust['Relative_Share_AI'] = countries_AI_sust['Total_share_AI']/total_AI*100
countries_AI_sust['Relative_Share_AI_Sust'] = countries_AI_sust['Total_share_AI_and_sust']/total_sust*100

countries_AI_sust.describe()
```

2.4 Map Visualization for Country-Level

```
countries = gpd.read_file("ne_110m_admin_0_countries.shp")
ctry = countries.rename(columns={"ISO_A2": "Country"})

merged = ctry.merge(countries_AI_sust, on="Country", how="right")

m = folium.Map(location=[50, 10], zoom_start=5, tiles="cartodbpositron", width='100%', height='800px')

colormap = linear.OrRd_09.scale(
    merged["Total_share_AI_and_sust"].min(),
    merged["Total_share_AI_and_sust"].max()
)
colormap.caption = "Share of Patents in AI and Sustainability"
colormap.add_to(m)

folium.GeoJson(
    merged.to_crs(epsg=4326,
    style_function=lambda feature: {
        'fillColor': colormap(feature['properties']['Total_share_AI_and_sust']),
        'color': 'black',
        'weight': 0.3,
        'fillOpacity': 0.7,
    },
    tooltip=GeoJsonTooltip(
        fields=["Country", "SOVEREIGNT", "Total_share_AI_and_sust", "share_sustainability"],
        aliases=[
            "Country Code:", "Country Name:",
            "Applications in AI and Sustainability:",
            "Share of AI+Sustainability from all AI patents (%):"
        ],
        sticky=True
    )
).add_to(m)

m
```

2.5 Calculating Shares for Applicants

```
applicant_share_total = defaultdict(float)

for applicants, shares in zip(europe_filtered_df['applicants'], europe_filtered_df['applicant_share']):
    for applicant, share in zip(applicants, shares):
        applicant_upper = applicant.upper()
        applicant_share_total[applicant_upper] += share

applicant_share_total_df = pd.DataFrame(
    list(applicant_share_total.items()),
    columns=["Applicant", "Total_share"]
)

applicant_share_total_df = applicant_share_total_df.sort_values(
    by="Total_share", ascending=False
).reset_index(drop=True)

total_app_share = applicant_share_total_df['Total_share'].sum()
applicant_share_total_df['Relative_share,%'] = applicant_share_total_df['Total_share']/total_app_share*100
applicant_share_total_df.describe()
```

2.6 Co-occurrence of Two Regions in the Same Patent

```
co_occurrence_counter = Counter()

for regions in europe_filtered_df['nuts2_app']:
    if not regions or len(regions) < 1:
        continue
    unique_regions = sorted(set(regions))
    for r1, r2 in combinations(unique_regions, 2):

        co_occurrence_counter[(r1, r2)] += 1
        co_occurrence_counter[(r2, r1)] += 1
    for r in unique_regions:
        co_occurrence_counter[(r, r)] += 0

all_regions = sorted(set([r for pair in co_occurrence_counter.keys() for r in pair]))

matrix_reg = pd.DataFrame(0, index=all_regions, columns=all_regions)

for (r1, r2), count in co_occurrence_counter.items():
    matrix_reg.at[r1, r2] = count
```

2.7 Heatmap for Regional Co-occurrence

```
# Create a copy of the matrix
matrix_lower_reg = matrix_reg.copy()

# Mask the upper triangle (excluding the diagonal)
mask = np.triu(np.ones(matrix_lower_reg.shape), k=1).astype(bool)
matrix_lower_reg = matrix_lower_reg.mask(mask)

# Plot with Plotly
fig = px.imshow(
    matrix_lower_reg,
    labels=dict(x="Region", y="Region", color="Co-occurrence"),
    x=matrix_lower_reg.columns,
    y=matrix_lower_reg.index,
    color_continuous_scale='Peach',
    aspect="auto"
)

fig.update_layout(
    title="Heatmap of Region Co-Occurrences",
    xaxis_tickangle=45,
    autosize=True,
    width=1000,
    height=1000
)

fig.show()
```

2.8 Converting Matrix into Long Format

```
co_occurrences_long_reg = matrix_reg.stack().reset_index()
co_occurrences_long_reg.columns = ['Region1', 'Region2', 'Count']

co_occurrences_long_reg = co_occurrences_long_reg[co_occurrences_long_reg['Region1'] !=
co_occurrences_long_reg['Region2']]

top_collaborations_reg = co_occurrences_long_reg.sort_values(by='Count', ascending=False)

top_collaborations_reg['SortedPair'] = top_collaborations_reg.apply(
    lambda row: tuple(sorted([row['Region1'], row['Region2']])), axis=1
)
top_collaborations_reg = top_collaborations_reg.drop_duplicates(subset='SortedPair').drop(columns='SortedPair')
```

2.9 Network for Regional Co-occurrence

```
G = nx.Graph()

threshold = 0

for r1 in matrix_reg.index:
    for r2 in matrix_reg.columns:
        weight = matrix_reg.loc[r1, r2]
        if r1 != r2 and weight > threshold:
            G.add_edge(r1, r2, weight=weight)

pos = nx.spring_layout(G, weight='weight', k=1.5, iterations=100, seed=42)

edge_x, edge_y = [], []
for edge in G.edges(data=True):
    x0, y0 = pos[edge[0]]
    x1, y1 = pos[edge[1]]
    edge_x += [x0, x1, None]
    edge_y += [y0, y1, None]

edge_trace = go.Scatter(
    x=edge_x, y=edge_y,
    line=dict(width=0.5, color='#999'),
    hoverinfo='none',
    mode='lines'
)

partition = community_louvain.best_partition(G, weight='weight')

# Find number of clusters
num_clusters = len(set(partition.values()))
print(f"Detected {num_clusters} clusters")


colors = list(mcolors.TABLEAU_COLORS.values()) + list(mcolors.CSS4_COLORS.values())
cluster_colors = {cluster: colors[i % len(colors)] for i, cluster in enumerate(set(partition.values()))}

node_x, node_y, node_text, node_color, node_size = [], [], [], [], []
node_strength = dict(G.degree(weight='weight'))
for node in G.nodes():
    x, y = pos[node]
```

```
    node_x.append(x)
    node_y.append(y)
    node_text.append(f"{node}")
    node_color.append(cluster_colors[partition[node]])

    size = 10 + (node_strength[node] * 0.3)
    node_size.append(size)

node_trace = go.Scatter(
    x=node_x, y=node_y,
    mode='markers+text',
    text=node_text,
    textposition="top center",
    hoverinfo='text',
    marker=dict(
        size=node_size,
        color=node_color,
        line=dict(width=1, color='black')
    )
)

fig = go.Figure(data=[edge_trace, node_trace],
    layout=go.Layout(
        title="Network of Regions by Co-Occurrence Strength (Coloured by Clusters)",
        title_x=0.5,
        width = 1200,
        height=1000,
        showlegend=False,
        hovermode='closest',
        margin=dict(b=5, l=5, r=5, t=40),
        xaxis=dict(showgrid=False, zeroline=False, showticklabels=False),
        yaxis=dict(showgrid=False, zeroline=False, showticklabels=False)
    )
)

fig.show()
```

2.10 Relatedness of Technological Classes

```
def extract_first_4(ipc_list):
    if isinstance(ipc_list, list):
        return [code[:4] for code in ipc_list if isinstance(code, str)]
```

```
    return []

europe_filtered_df['IPC_short'] = europe_filtered_df['IPC'].apply(extract_first_4)
ipc_codes = []
nuts2_codes = []

for i in range(ipc_analysis.shape[0]):
    nuts_list = list(set(ipc_analysis['nuts2_app'][i]))
    ipc_list = list(set(ipc_analysis['IPC_short'][i]))
    for nuts in nuts_list:
        for ipc in ipc_list:
            ipc_codes.append(ipc)
            nuts2_codes.append(nuts)

app = [1]*len(nuts2_codes)
```

2.11 RTA Computation

```
def compute_rta(matrix):
    row_totals = matrix.sum(axis=1).values.reshape(-1, 1)
    col_totals = matrix.sum(axis=0).values.reshape(1, -1)
    grand_total = matrix.values.sum()

    rta = (matrix.values / row_totals) / (col_totals / grand_total)
    return pd.DataFrame(rta, index=matrix.index, columns=matrix.columns)

rta_matrix = compute_rta(matrix)
```

2.12 Relatedness Matrix

```
co_occurrence = rta_binary.T.dot(rta_binary)

np.fill_diagonal(co_occurrence.values, 0)

ubiquity = rta_binary.sum(axis=0)

relatedness_matrix = pd.DataFrame(index=co_occurrence.index, columns=co_occurrence.columns, dtype=float)

for i in co_occurrence.index:
    for j in co_occurrence.columns:
        max_ubiq = max(ubiquity[i], ubiquity[j])
        if max_ubiq > 0:
            relatedness_matrix.loc[i, j] = co_occurrence.loc[i, j] / max_ubiq
```

```
    else:
        relatedness_matrix.loc[i, j] = 0.0
```

2.13 Exceptions with Relatedness = 1

```
ubiq=[]
cooc = []
for p in range(148):
    row = long_df[long_df['value'] == 1.0].iloc[p]
    i, j = row['IPC1'], row['IPC2']

    # Now check values
    cooc.append(co_occurrence.loc[i, j])
    ubiq.append([rta_binary[i].sum(),rta_binary[j].sum()])
indices = [i for i, x in enumerate(cooc) if x > 1]

exceptions = long_df.iloc[indices]
exceptions['cooccurrence'] = [cooc[i] for i in indices]
exceptions['ubiquity'] = [ubiq[i] for i in indices]
```

2.14 Heatmap for Average Relatedness for Sectors

```
non_1_relatedness['Section 1'] = non_1_relatedness['IPC1'].str[0]
non_1_relatedness['Section 2'] = non_1_relatedness['IPC2'].str[0]
section_pairs = non_1_relatedness.groupby(['Section 1', 'Section 2'])['value'].mean().reset_index()
pivot = section_pairs.pivot(index='Section 1', columns='Section 2', values='value')
sns.heatmap(pivot, annot=True, cmap='flare')
plt.title("Average Relatedness by IPC Section Pairs")
plt.show()
```