Exploring Patent Data Analysis in AI Innovation: Integrating Advanced Methodologies and Large Language Models

AAU Business School
Spring Semester 2024

AALBORG UNIVERSITY
STUDENT REPORT

AAU Business School
Aalborg University
http://www.aau.dk

Title:
Exploring Patent Data Analysis in AI Innovation: Integrating Advanced Methodologies and Large Language Models

Theme:
Innovation and General Purpose Technology

Project Period:
Spring Semester 2024

Project Group:

Participant(s):
Christian Nielsen

Supervisor(s):
Roman Jurowetzki

Page Numbers: 59

Date of Completion:
June 3, 2024

Abstract:

The rapid advancement of artificial intelligence (AI) has led to a significant increase in related patent filings, making patent analysis a critical tool for understanding technological trends and innovation dynamics. This project explores a framework that combines various methodologies for analyzing patent data to map technological trends within AI, and combining these insight with modern AI methods.Patent counts revealed the geographical distribution of innovation, highlighting key companies and technological fields. It also supported the characterization of AI as a General Purpose Technology (GPT), as this was shown to have a wide variety of use cases across technological fields, and strongly complimenting existing technologies. In addition, distribution of AI patents across recent years showed that there has been a massive spurt of patent activity, suggesting that AI is also undergoing continuous development. Citation and network analyses mapped the relationships among patents, inventors, and organizations, identifying central actors and the structure of innovation ecosystems. Despite challenges such as a fragmented network graph and issues with repeating patents within the network, the study demonstrated the potential of combining traditional and advanced analytical techniques. The proof-of-concept Retrieval-Augmented Generation (RAG) system showcased the capability of LLMs to enhance patent analysis by providing nuanced and comprehensive responses to complex queries. Despite the flawed implementation and results of the network analysis, this project underscores the importance of a multifaceted approach to patent analysis, leveraging both traditional methodologies and advanced AI tools to gain deeper insights into technological innovation.

1 Abstract

The rapid advancement of artificial intelligence (AI) has led to a significant increase in related patent filings, making patent analysis a critical tool for understanding technological trends and innovation dynamics. This project explores a framework that combines various methodologies for analyzing patent data to map technological trends within AI, and combining these insight with modern AI methods.

Patent counts revealed the geographical distribution of innovation, highlighting key companies and technological fields. It also supported the characterization of AI as a General Purpose Technology (GPT), as this was shown to have a wide variety of use cases across technological fields, and strongly complimenting existing technologies.

In addition, distribution of AI patents across recent years showed that there has been a massive spurt of patent activity, suggesting that AI is also undergoing continuous development.

Citation and network analyses mapped the relationships among patents, inventors, and organizations, identifying central actors and the structure of innovation ecosystems.

Despite challenges such as a fragmented network graph and issues with repeating patents within the network, the study demonstrated the potential of combining traditional and advanced analytical techniques. The proof-of-concept Retrieval-Augmented Generation (RAG) system showcased the capability of LLMs to enhance patent analysis by providing nuanced and comprehensive responses to complex queries.

Despite the flawed implementation and results of the network analysis, this project underscores the importance of a multifaceted approach to patent analysis, leveraging both traditional methodologies and advanced AI tools to gain deeper insights into technological innovation.

Contents

List of Figures

List of Tables

2 Introduction

Tracking technological developments, also known as technology mapping, is one of the elements used in general technology management, and allows companies, organizations, and other entities to find a stronger link between available technological resources and objectives set by the aforementioned. Staying up to date on current developments within a field that either directly or indirectly relates to the objectives of a business, can result in significant strategic advantages in terms of further developing the market offerings, as well as resource optimization of a business [CSB00].

Innovation has therefore been studied extensively in particularly economics, with notable authors such as Joseph Schumpeter focusing directly on the purpose of innovation in businesses and markets [Sch27]. As well as Edith Penrose that explored the impacts of technological advancements on firms in terms of strategy and management [Pen95]. It is therefore not surprising that innovations can cause varying degrees of significant disruption on the markets in which they occur. This theoretical perspective, often referred to as disruptive innovation theory, as popularized by Clayton M. Christensen in 1997 with his book titled "The Innovators Dilemma", focuses on the challenges faced by companies that are financially tied to current technology, as their business has a significant potential to fail when disruption occurs [Chr97].

Due to this potentially disruptive nature of technology, it becomes vital for companies to be able to identify these innovations, ideally before they occur. Additionally, governmental organizations can utilize technology mapping in the same way as businesses. They can more directly align their investments according to their governmental goals, with the current technology fields, and further foster necessary innovation through their investment. Therefore, understanding the technological landscape in which an entity operates, generates inherent value, not just in terms of business, but society as a whole [SGdV+18]. The rapid rate of which some technologies develop, such as the recent spur of innovation with AI as well as connected technologies [RVD20], must however be fueled by some sort of underlying factor. This is what I seek to identify and analyze in the broader context of technology using the methods in literature.

At the same time there is an increasing need for implementing some types of technologies more than others, for example, in light of many countries' goals for emission neutrality which is evident in both news media [Hen24] and scientific literature [WZS+24]. Understanding and foremost identifying these underlying, somewhat pivotal factors within a technological field, becomes of great value for decision makers world wide. This begs the question as to how the current technological fields is being analyzed and processed, and how particularly pivotal technologies can be identified within the patent landscape of a given technology.

Working under the basic assumption that patented technology is more likely to be implemented sooner as opposed to theories within current scientific literature, as patented technology often has finances tied directly to them through different types of fees, leads to the direct purpose of this paper. This paper aims to explore and utilize some of the current technology mapping methods used for patent data, in order to attempt to map the current technological field. Additionally this paper will attempt to utilize Large Language Models to process these patents and analyze an single field in the context of technological innovation.

The literature review will go into more detail on the scientific literature that shaped the current methodological landscape within the field of technology mapping, patent analysis and Large Language Models. Additionally, the analysis will explore ways in which these methods within literature can be combined and thereby enhanced, in order to perform analysis on a larger scale. The final goal will be to create a proof-of-concept implementation that will focus on identifying and evaluating these underlying factors that significantly increase the innovation rate across technological fields.

Besides the written product, this paper will include python script implementations of the above, using a limited number of patents from just a single type of technology, but with the ability to upscale to a larger perspective with relatively limited code alteration.

Central to my investigationis the research question:
'How can technology mapping be utilized in conjunction with current AI and network graphing methods to extract meaningful insights into the field of technological innovation and General Purpose Technologies'

This question throws a wide net, and allows for a significant amount of creativity in terms of the proposed methodology, whilst still maintaining focus on the area of technological innovation and General Purpose Technology. This question has several sub-questions tied to it, that will need answering in order to reach a conclusion

What are the current methods and tools used for patent analysis in technology mapping?
To answer this question, I will explore already proposed and utilized methodologies within the scientific literature, and identify the dynamics within the field of technological innovation. Additionally, based on current efficiency and knowledge gaps in literature, innovative approaches will be explored for enhancing current analysis methods. Furthermore, this paper will attempt a practical use-case through the use of Large Language Models.

How do LLMs currently contribute to patent analysis and technological innovation, and what are their limitations?
As will be detailed further on in the literature review, utilizing LLMs in the context of patents and innovation studies has been done before. Understanding the limitations of their applicability is crucial for the purpose of extending any existing methodology, as failure to do so can lead to unexpected and/or unintended misunderstandings that serve as a detriment to the proposed framework.

How should this data be analyzed in order to extract insightful information within technology from the perspective of innovation?
As this should be seen as an attempt to extend or enhance current methods within the existing literature, it is once again necessary explore the current landscape of technology mapping in literature and identify what points of data have been considered meaningful for analysis.

What are the driving factors of innovation within the chosen technological field
The purpose of this question will be to explore what factors or characteristics that enable an increased innovation rate within a given technological field. Another way of phrasing this question, would be whether there are identifiable features of a technology that can be used to pinpoint hotspots of innovation in terms of the actual technological field.

Which data points can be utilized to form a connection between each patent/technology type?
In order to connect technological advancement and innovations, it is necessary to identify key data points that can provide this perspective. Scientific literature should provide us with the perspectives needed to bridge these gaps, and help identify key technologies that increase innovation rates.

What characteristics does General Purpose Technology show, and how can we identify these in patent data?
As the purpose of this paper is to identify the drivers of General Purpose Technology, one of the main objectives will be to determine whether or not the selected technology exhibits the characteristics of one.

3 Key Concepts

Before diving into the literature review, Some key concepts that will be mentioned throughout the project will be defined, in order to provide a better overview of the theoretical and technical terms mentioned going forward in this paper.

Patents:A patent is a form of intellectual property that gives its owner the legal right to exclude others from making, using, selling, and importing an invention. In exchange, the patent owner must disclose the details of the invention to the public.

Innovation Ecosystem: An innovation ecosystem is a network of interconnected organizations, including firms, universities, research institutions, and government agencies, that work together to drive technological innovation. This ecosystem supports the development, integration and commercialization of new technologies.

Text Mining: Text mining is the process of extracting useful information and patterns from large sets of textual data, often using AI related techniques.

FAISS Index: FAISS (Facebook AI Similarity Search) is a library for efficient similarity search and clustering of dense vectors. In the context of this project, FAISS is used to create an index of document embeddings to facilitate fast retrieval of relevant documents based on cosine similarity to query embeddings.

Large Language Models (LLMs): These are advanced artificial intelligence models specifically designed to understand, interpret, and generate natural human language. They are 'large' due to the significantly large amount of data they are trained on and their extensive neural network architectures.

Network Graph: A network graph is specific type of data formatting made for analyzing connections between nodes. In a network graph, you observe different nodes that can represent different entities, such as individual persons, groups, organizations, or papers, just to name a few examples. Between each of these nodes, the intent is to identify data and/or information that relate these nodes to each other, in order to facilitate analysis, such as which inventors operate most within certain fields of technology.

General Purpose Technology (GPT): The definition of General Purpose Technology is technology that can be characterized as the following:

- **Room for improvement and innovation:** The technology in question should have been iterated upon since its initial introduction, thus undergoing a continuous innovation cycle, and spurring cumulative knowledge creation around it. One example could be the invention of transformer models, as these, at least currently, constantly go through incremental research and improvement.
- **Widespread usage across technological fields:** The technology in question, should have a wide spread influence on most sectors of technology, and be widely applicable to many different types of technology.
- **Should strongly complement existing or emerging technologies:** The technology in question should have an impact on both current and emerging technologies, by introducing the need of altering and/or combing the existing as well as increasing the opportunity rate of new technologies.[Pet20]

4 Literature review

In this next section, the literature that relates to the subject of innovation, General Purpose Technology, patent literature, methods for analyzing relationships within technology will be explored and the inherent business and societal value gains explained. It will also explore some of historical development of the main technology utilized in this project, such as Large Language models, as well as related strategies that enable these to generate human-like text in a manner that is more aligned to a specific topic. To make it more transparent, this section is subdivided into individual topics that may or may not have some overlaps. Finally the potential business value, impacts and methods that can be derived from these fields of study combined will be summarized.

4.1 Theory of innovation

The following section 4.1 will briefly cover some of the main ideas within the theory of innovation, the dynamics and systems involved, and define how technological progress drives further innovation as well as generate value in businesses.

Innovation is a central concept in economic theory and business studies, fundamentally driving growth and development. Joseph Schumpeter, one of the most influential figures in innovation theory, described innovation as a process of "creative destruction," where old ways of doing things are incessantly destroyed and replaced by new ones [Sch76]. This process is central to capitalist economies, where innovation leads to new products, processes, and business models that enhance productivity and efficiency. Schumpeter's view highlights the dynamic and disruptive nature of innovation, underscoring its role in fostering economic dynamism and competitiveness.

Edith Penrose further contributed to the theory of innovation through her resource-based view of the firm. Penrose emphasized that firms grow and diversify based on their internal resources and capabilities, including their ability to innovate. This perspective underscores the role of organizational knowledge and the strategic management of resources in fostering innovation [Pen95]. According to Penrose, the growth of firms is not merely a result of external market conditions but is deeply rooted in the firm's internal resources and capabilities, which are developed and leveraged through innovation. This insight provides a strategic dimension to the theory of innovation, highlighting the importance of managerial practices and resource allocation in sustaining competitive advantage.

Innovation theory also encompasses the understanding of technological change and its underlying drivers. Richard Nelson and Sidney Winter introduced a more evolutionary theory of economic change, emphasizing routines and capabilities that firms generate over time. These routines, which include established practices and processes, influence a company's ability to innovate and adapt to changing environments [NW84]. Nelson and Winter's evolutionary perspective suggests that innovation is not a random process but is shaped by the historical and institutional contexts within which firms operate. Their work underscores the cumulative and path-dependent nature of technological change, where past innovations sets the directions for future developments.

In addition to Schumpeter, Penrose, and Nelson and Winter, several other scholars have contributed to the understanding of innovation. Chris Freeman and Keith Pavitt have emphasized the systemic nature of innovation, arguing that it is embedded in broader socio-economic systems that include interactions among firms, governments, and research institutions. This systemic perspective highlights the importance of national and regional innovation systems in fostering and sustaining innovation. Freeman's work on national innovation systems, in particular, has shown how policy frameworks and institutional settings can significantly influence the innovation process by providing incentives, infrastructure, and support for research and development [Dor88].

Furthermore, the theory of innovation has been further enriched by the contributions of scholars like Bengt-Åke Lundvall, who introduced the concept of interactive learning. Lundvall argued that innovation is a collective process that involves continuous interaction and feedback among various actors, including firms, universities, and government agencies. This interactive process is crucial for the creation and diffusion of knowledge, which drives technological progress and economic growth. Lundvall's work highlights the importance of networks and collaboration in innovation, suggesting that isolated efforts are less likely to succeed compared to those that leverage the collective expertise and resources of a broader innovation ecosystem [Pav95].

Overall, the theory of innovation integrates multiple perspectives, each contributing to a more comprehensive understanding of how innovation occurs and its impact on economic and business dynamics. From Schumpeter's creative destruction and Penrose's resource-based view to Nelson and Winter's evolutionary theory and the systemic approaches of Freeman, Pavitt, and Lundvall, the field of innovation studies offers rich insights into the mechanisms that drive technological progress and economic development.

4.1.1 Dynamics of innovation

The dynamics of innovation involve understanding how innovations emerge, diffuse, and impact economies and societies. The innovation process is often depicted as a series of stages, including invention, development, and diffusion. Rogers introduced the diffusion of innovations theory, which explains how, why, and at what rate new ideas and technology spread throughout society [Rog03]. Schumpeter also highlighted the role of entrepreneurs in driving innovation by identifying opportunities and taking risks to implement new ideas. He distinguished between incremental and radical innovations, with radical innovations having the potential to disrupt entire industries. Incremental innovations however happens regularly, but is generally much less disruptive in nature, thus not posing as great a threat to existing actors within an industry. [Sch80]. It is worth nothing however, that f companies within these given industries do not realign their internal resources as examined by Penrose, the company is more likely to fail. However, if the company is able shift industry along with the technological advancements, whilst still maintaining a competitive edge in the main market it first operates in, it can more easily be able to survive these disruptive waves of innovation [Pen95]

Christopher Freeman was central in conceptualizing the systemic nature of innovation, particularly through his work on national innovation systems. He argued that the interactions between various actors, such as firms, governments, and research institutions, are critical in shaping innovation processes and outcomes. He illustrated how national policies and institutional frameworks play a pivotal role in fostering an environment to increase innovation, particularly through incentive structures. These structures enable institutions and organizations to continue research that further technological development, both incrementally and radically. In essence according to Freeman, countries with well-developed innovation systems are better positioned to achieve sustained economic growth and technological advancement [Dor88]. This perspective on the concept of national innovation systems was expanded upon later in a paper by Keith Pavitt, where he explored the role of knowledge creation, learning and interactive processes as defined by Lundvall. He highlighted that innovation is not just the result of isolated activities within firms or institutions, but is deeply embedded in broader social and economic systems, a point Freeman as discussed had made earlier. Lundvall argued the point that innovation is a collective process, involving continuous interaction and feedback among several different actors in a society. His work in research has underscored the importance of knowledge sharing, collaboration, and the co-evolution of technology and institutions in driving innovation [Pav95].

In conclusion, the dynamics of innovation are complex and multifaceted, involving a range of actors and processes. The theories and models developed by Rogers, Schumpeter, Freeman, Pavitt, and Lundvall provide valuable insights into how innovation occurs and spreads, emphasizing the importance of systemic interactions and the collective nature of technological progress.

4.1.2 Economic impacts of innovation

Innovation has the potential to cause significant impacts on economic growth, societal development, and technological advancement. Schumpeter's notion of creative destruction highlights the dual nature of innovation: it can displace existing technologies and industries while creating new opportunities and markets. This process is a key driver of long-term economic growth and competitiveness [AH92]. Schumpeter's theory once again underscores how innovation acts as a catalyst for economic dynamicis by continuously reshaping the industrial landscape. As evident by literature in section 4.1 and 4.2, although this disruptive process can be challenging for established businesses, it facilitates the emergence of new enterprises and industries, fostering sustained economic development and technological progress.

Penrose's work emphasizes how innovation enables firms to exploit new opportunities and achieve sustainable growth. By leveraging their unique resources and capabilities, firms can innovate continuously, leading to incrementally improved performance and market expansion [Pen95]. This inevitably leads to higher rate of economic growth within societies that has strong innovation systems, as the available resources are utilized more efficiently across industries. The impact of innovation extends to societal influencing factors such as employment, health, and environmental sustainability. Innovations in healthcare, for instance, have significantly improved quality of life and life expectancy [CAP+17]. However, the disruptive nature of innovation can also lead to economic challenges, such as job displacement and shifts in industry demand, potentially increasing economic inequality [KPSS21]. Despite these challenges, the overall economic benefits of innovation, such as increased productivity, the creation of new industries, and sustained economic growth, highlight its essential role in driving economic progress.

In summary, innovation is a critical driver of economic growth and development. It promotes economic dynamics through creative destruction, enables firms to exploit new opportunities, enhances productivity, and creates new markets. By fostering a culture of innovation, countries and regions can improve their competitiveness, attract investment, and achieve sustainable economic growth. The economic impacts of innovation are wide-ranging and profound, underscoring the importance of fostering and supporting innovative activities across all sectors of the economy.

4.1.3 Definition and impact of General Purpose Technology in literature

The paper "General purpose technologies 'engines of growth", discusses the significant impact of General Purpose Technologies (GPTs) on technical progress and economic growth, with examples of innovations like the steam engine and semiconductors. GPTs are characterized by their widespread applicability across various sectors, potential for improvement, and capacity to generate innovational complementarity technology that enhance returns-to-scale applications.

The study highlights how GPTs serve as enabling technologies that do not just provide immediate solutions but also open new avenues for advancements across different industries. Thus functioning as a driver of innovation, whilst being innovative in of itself. This is because these technologies create a foundation for further innovations that cumulatively boost productivity across the economy. However, the paper identifies significant coordination difficulties among different sectors and temporal gaps in technology development, which hinder the optimal utilization of GPTs. It ultimately concludes that the decentralized nature of economic activities provides significant challenges in fully capitalizing on these technologies, primarily due to coordination failures and insufficient innovation incentives. [BT95]. Compared to insights gained in by the literature in the section on the economics impacts of innovation this seems less than ideal, as the economic growth would only be fueled further by incentivizing development further, unless there was a high risk that it would cause significant disruption in a key industry in the given country.

Another paper, written in 2009, presents a model of long-term economic growth driven by a sequence of GPTs. Much like earlier literature, it builds on the theory that significant inventions which occur at decreasing intervals, serve as major drivers of long-term growth. These GPTs, from early human innovations like domestication and metalworking through industrial and digital revolutions, fundamentally alter economic landscapes and significantly accelerate growth.

Central to this model is the characterization of GPTs as enabling technologies that do not just solve existing problems but open up new avenues for further innovation and application across industries. This highlights the increasingly shorter intervals between the emergence of new GPTs, reflecting an accelerating pace of technological development, driven by the accumulation of knowledge in society.

The model integrates the dynamics of GPT emergence with economic cycles, showing that new GPTs typically build on pre-existing technologies. This linkage suggests that the knowledge base and prior technologies critically shape the trajectory and impact of new GPTs. Additionally, the paper explores the economic implications of new GPTs within their lifecycle, suggesting that while they may initially disrupt output, they eventually enhance productivity and growth by accelerating innovation cycles in other areas [SSW14].

This point is further examplified by another paper that explores the impact of the internet as a GPT. The focus here was primarily on how different types of telecom services such as cellular, fixed-line, and Internet services affect firm performance across various sectors, suggesting that Internet access in particular, might function as a GPT. It found that there were significant benefits for firms in regions with better Internet access, highlighting substantial improvements in both work output and sales growth.

Results showed that firms were more productive and grew faster in countries with a better Internet infrastructure. In contrast, improvements in fixed-line and cellular services did not consistently increase the performance of a firm, and in some cases, the relationship was in fact negative. Additionally, the impact of the Internet is found to be more pronounced in small and medium enterprises (SMEs) than in large firms, suggesting that the Internet enabled SMEs to overcome traditional barriers to market expansion and efficiency gains [CQX15]. This could only further balance the economy, as it should become less reliant on a few large firms, as well as increase the amount of job offerings that exist within the markets.

4.1.4 Identifying and evaluating a General Purpose Technology

It is worth noting that the evaluation of these types of technology cannot solely be done through the characteristics of GPTs, as presented in the earlier section. In addition to these characteristics of spurring innovation, the implementation rate must be considered as a valuable indication, as discussed.

One paper explores this particular indicator through online job postings, containing technological keywords such as 3D printing, machine learning and big data. Especially focusing on how these technologies rank against each other. The ranks were based off of a variety of measures, such as gini coefficients as well as the distribution and concentration of job postings containing these keywords across industries. The analysis in the paper revealed that certain clusters of related technologies, particularly those associated with data science and machine learning, consistently rank high across all measures used in the analysis, suggesting a higher likelihood of these technologies achieving GPT status. The paper argues that the job posting-based measures correlate well with patent-based measures and can predict future technological importance, offering policymakers and managers early insights into which technologies might warrant substantial investment and strategic consideration due to their potential widespread impact [GTT23]. This notion ties into the objective of this project, as I will attempt to identify factors that enable AI and ML as a type of GPT innovation.

However, for indicating whether or not a technology can be considered a GPT, studies have found that patents may hold a significant opportunity to identify these GPTs more directly. A paper by Maryann P. Feldman and Ji Woong Yoon performed empirical testing within the field of patented rDNA technology, which at the time was considering a credible GPT. They found that the characteristics of GPTs may indeed be identifiable through an analysis of patent data, by analyzing their technological linkage within the larger scope of the technological field, when compared to patents with similar origin. The origin of these patents were evaluated through the attributes connected to these technologies, such as the country of origin and inventors [FY11].

To provide a summary, the study of innovation offers valuable insights into the mechanisms and impacts of technological change. By examining the underlying theoretical frameworks and dynamic processes of innovation, it becomes easier to understand the transformation potential of General Purpose Technologies and their important role in shaping future economic and social landscapes. In addition, the literature has provided methods that are easily implemented for identification of GPTs. They have indicated that evaluating a GPT must inherently be done in a macro perspective, meaning that it is necessary to qualitatively assess the origins and inventors of these types of innovations to determine whether it fits the characteristics.

4.2 Patent data and methods for analysis

The following section 4.2, will explore some of the methodologies of patent analysis. Patent data has become a valuable resource for various fields of study, including economics, business, and innovation management. This part of the review provides an overview of the key themes and findings in the literature on patent data, focusing on its use in innovation studies, technological forecasting, and economic analysis. The review is organized into sections covering the sources and types of patent data, methodologies for patent analysis, applications in research, and the limitations and challenges associated with using patent data

4.2.1 Patent data

Patent data is widely used in research to investigate various aspects of innovation and economic growth. Studies have used patent data to measure technological progress, track the diffusion of technologies, and analyze the determinants of innovation. Additionally, patent data helps in identifying leading innovators and emerging technological fields. Patent data is primarily sourced from patent offices such as the United States Patent and Trademark Office (USPTO), the European Patent Office (EPO), and the World Intellectual Property Organization (WIPO). These sources provide access to patent documents, including detailed information on inventions, inventors, assignees, and legal status.

Acs, Anselin, and Varga explored the use of patent data to measure regional production of new knowledge. Their study demonstrated that regions with higher patent counts tend to have more robust innovation ecosystems. They argued that patents could serve as proxies for regional innovation capacity, allowing policymakers to identify and support high-potential areas [AAV02].

Griliches paper on patent statistics from 1990, highlights the significance of patent statistics as economic indicators, offering a comprehensive survey on their use in economic research. He argues that patents, due to their detailed nature and the legal requirements for novelty and non-obviousness, serve as reliable indicators of innovation. However, he also notes the limitations of using patent counts alone, as they do not directly measure the economic value of the inventions they represent [Gri90].

Jaffe and Trajtenberg delved into the role of patents and patent citations in understanding innovation. They emphasize that patents not only protect intellectual property but also provide a significant source of data on technological developments. By analyzing citations within patents, researchers can trace the flow of knowledge and identify key innovations and inventors. This approach allows for a deeper understanding of the dynamics of technological change and the application of knowledge [Ref03]. The literature review thus far, particularly indicate the importance of the following types of patent data.

- **Bibliographic Data:** Includes information on the patent's inventor(s), application and grant dates, and International Patent Classification (IPC) codes. This data is essential for identifying trends in patenting activity across different technological fields and geographies.
- **Full-text Data:** Provides the complete text of the patent document, including claims, descriptions, and drawings. Full-text data is crucial for in-depth analysis of the technological content and novelty of the patent.
- **Citation Data:** Consists of references to prior patents and non-patent literature. Citation data is used to study knowledge flows and the impact of specific patents.
- **Legal Status Data:** Tracks the current legal status of patents, including information on renewals, expirations, and legal disputes. This data is useful for understanding the lifecycle and economic value of patents.

These types of patent data serve as a rich source of insight, if combined with the correct method. Some of which will be covered in the following section.

4.2.2 Methods and applications for Patent Data Analysis

Researchers utilize various methodologies to analyze patent data, including patent counts, citation analysis, and network analysis.

Patent counts

Counting the number of patents filed or granted over a period is a straightforward method to gauge innovative activity. However, this method has limitations as it does not account for the quality or economic impact of the patents. Despite this, patent counts are widely used as they provide a basic measure of inventive output [Gri90].

Scenario Analysis

Scenario analysis is one example that recognizes the limitations of solely quantitative or qualitative approaches, this research involves national stakeholders in workshops to create socio-economic narratives, which are then linked to technological energy forecasts. Thereby providing a richer, more contextual understanding of future scenarios. In simpler terms, it revolves around a form of storytelling, where the objective is to create more robust scenarios, by involving the experiences of actors within the given field and combining them with a more structured modelling approach [FASR15].

One example of a use-case could be forecasting in the energy sector, where scenario analysis has been pivotal in exploring the implications of renewable energy adoption and the transition away from fossil fuels. For instance, stakeholders from various backgrounds—government policymakers, energy companies, and environmental groups—participate in workshops to discuss and model the impact of different energy policies and technological innovations like solar panels and biofuels [WDG20]. This means that a higher quality of knowledge input to this type of analysis, especially within the possible grasps of related technologies, is invaluable to generate more accurate scenario forecasts.

Text Mining

Another method that lies more within the realm of data science, is text mining, in which specific keywords are used to identify trends within given texts from patents, as well as scientific literature. In this approach to patent analysis, two primary methodologies are prominent: network-based patent analysis and keyword-based patent analysis. Network-based analysis focuses on the structural relationships between patents without delving into their detailed technical content. In contrast, keyword-based analysis explores the technological content of patents through specific keywords but fails to capture the interrelationships between these patents [CH14].

Technological Routes(TR)

Another paper from 2019 however, explores the technological routes (TR) methodology as an innovative approach for identifying promising research projects and technologies in technology-intensive sectors. By applying social network analysis (SNA) to patent data, the TR method constructs a technological route map using patent citations. This map helps visualize the development path of a technology, indicating which trends are most promising. Which in turn also can help identify general purpose technology indirectly through innovative trends.

Key steps of the TR methodology include:

- 1. Identification of relevant patents: Using specific IPC codes, patents pertinent to the research object were selected and cleaned for analysis.
- 2. Building and analyzing the patent network: Patents were organized and analyzed using the Gephi software, where relationships (links) are based on patent citations (backward citations).
- 3. Extraction of technological routes: By applying algorithms like SPLC (Search Path Link Count) and SPNP (Search Path Node Pair), key technological routes were identified, showing the most relevant paths and central patents within the network.

By integrating qualitative and quantitative data and utilizing comprehensive patent analysis, this method offers significant insights into technology planning and strategic decision-making. It highlights the practical application of combining SNA with patent data to enhance the competitiveness and innovation capabilities of firms in rapidly evolving sectors [LPP19].

Citation Analysis

Citation analysis examines the references made by patents to prior patents. This method is particularly useful in understanding the impact and influence of specific inventions. Hall, Jaffe, and Trajtenberg developed the NBER patent citation data file, which has become a critical resource for citation analysis. Their work demonstrated that highly cited patents are often more valuable and influential, indicating breakthroughs or significant advancements in technology [HTJ01].

Network Analysis

Network analysis also maps the relationships among patents, inventors, organizations and other entities. This approach helps in understanding the structure of innovation networks and the flow of knowledge. Trajtenberg used citation networks to study the value of innovations, showing that patents with more citations tend to be more economically valuable. This method also helps identify central actors and emerging technological trends [Tra90].

As an example, one team of researchers utilized network analysis to study regional innovation. They found that regions characterized by small-world networks, where inventors are closely interconnected, tend to exhibit higher levels of innovative activity. Their research highlighted the importance of collaborative networks and knowledge spillovers in fostering innovation [FKJ07].

Network analysis is a large scientific field in its own respect, and one crucial method that lies within this field is known as community detection. Community detection focuses on identifying clusters or groups of nodes within a network that are more densely connected to each other than to the rest of the network. This technique helps reveal the underlying structure and organization of the network, providing insights into the relationships and interactions within it. One widely used method for community detection is the Louvain method, which optimizes the modularity of the network. Modularity is a measure of the strength of the division of a network into modules (also called communities). The Louvain method is highly efficient and has been applied to large-scale networks in various fields, including social networks, biological networks, and technological networks [BGLL13].

In the context of patent networks, community detection can uncover clusters of related technologies and innovations. For instance, Chen et al. applied community detection techniques to a network of patents to identify clusters of technological fields. Their analysis revealed that patents within the same community often share similar technological features and are more likely to cite each other, suggesting a cohesive structure of innovation within those fields [CXMR07].

4.2.3 Limitations and challenges

Despite its usefulness, patent data has limitations and challenges. Not all inventions are patented, and the tendency to patent varies across industries and countries. The quality and value of patents also differ, making it difficult to use patent counts as a direct measure of innovation. Moreover, the complexity and technical language of patent documents pose challenges for data extraction and analysis.

Duguet and Kabla examined appropriation strategies and motivations behind patenting at the firm level. Their econometric analysis revealed significant variation in patenting behavior across firms, influenced by factors such as industry characteristics and firm size. They noted that while some firms patent primarily to protect their innovations, others use patents strategically to block competitors or to enhance their market position [DK98].

4.3 Development and evolution of traditional NLP into Large Language Models

Large Language Models (LLMs) represent the cutting-edge in natural language processing (NLP), built on deep learning architectures that have evolved significantly over the years. In this following section I will explore their evolution from traditional NLP methods, through the subsequent developments of Word2Vec, Sequence-to-Sequence and finally the impact of the attention mechanism.

4.3.1 Traditional methods of Natural Language Processing

Before the development of neural network-based models, NLP relied heavily on statistical methods like Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF).

- **Bag of Words (BoW):** Introduced as a basic method for text representation, BoW models text as an unordered collection of word frequencies. Each document is represented as a high-dimensional vector, where each dimension corresponds to a unique word in the vocabulary. Despite its simplicity, BoW disregards grammar, word order, and context, leading to sparse vectors that lack semantic understanding [MRS08]

- **TF-IDF:** An enhancement over BoW, TF-IDF assigns numerical weights to words based on their frequency within a document relative to their frequency across a corpus. This method improves the representation by emphasizing important words and reducing the influence of common words [Jon][MRS08]. However, TF-IDF still suffers from problems with high dimensionality and sparsity, and it does not capture semantic relationships between words.

While BoW and TF-IDF focus on individual words, N-grams consider sequences of words to capture some level of context. For instance, bigrams (n=2) consider pairs of consecutive words, while trigrams (n=3) consider triples. N-grams provide a way to model local contexts within textual data, which addresses some of the limitations of BoW and TF-IDF. By considering word sequences, N-grams can capture common phrases and local syntactic structures [BDPd+92]. However, as n increases, the dimensionality of the representation grows exponentially, leading to data sparsity and computational challenges. Moreover, N-grams struggle with long-range dependencies and semantic understanding. Despite these advancements, traditional methods like BoW, TF-IDF, and N-grams were limited in their ability to capture deeper semantic relationships and context within text. This limitation paved the way for the development of more sophisticated models that could better understand and represent language.

The introduction of neural networks into NLP marked a significant shift in how language was processed and understood. Early neural network models began to explore the use of multi-layer perceptrons (MLPs) and simple recurrent neural networks (RNNs) for tasks such as language modeling and text classification. RNNs were designed to handle sequential data by maintaining a hidden state that captures information about previous elements in the sequence. This made RNNs suitable for tasks like language modeling, where the order of words is crucial [Elm90]. However, traditional RNNs struggled with long-term dependencies due to the vanishing gradient problem. The vanishing gradient problem occurs in neural networks when gradients used to update the networks weights become extremely small, effectively preventing the network from learning long-term dependencies.

4.3.2 Word2Vec models

Developed by Mikolov and his team, Word2Vec marked a significant advancement by introducing dense vector representations of words, known as word embeddings. Word2Vec aimed to capture the semantic meaning of words in a continuous vector space, addressing the limitations of traditional methods. This shift addressed the limitations of traditional NLP methods by capturing semantic relationships between words. Word2Vec introduced two primary model architectures for learning word embeddings: Continuous Bag of Words (CBOW) and Skip-gram.

- CBOW: Predicts a target word based on its surrounding context words. This model averages the context word vectors to predict the target word.
- Skip-gram: Predicts surrounding context words given a target word. This model learns to predict the context words from the target word.

Both models use shallow neural networks to produce dense, continuous vector representations of words, capturing semantic relationships through vector mathematics [MCCD13]. Word2Vec demonstrated that semantic similarities between words could be captured effectively in a lower-dimensional space. For example, the vector representation of the word king, minus the vector for man, plus the vector for woman resulted in a vector close to the word queen, showcasing the model's ability to capture analogies. This ability to understand semantic relationships revolutionized the field and laid the foundation for more complex language models

4.3.3 Sequence-to-Sequence (Seq2Seq) Models

While Word2Vec focused on word-level embeddings, there was a need for models that could handle entire sequences of text, such as sentences or paragraphs. Sequence-to-Sequence (Seq2Seq) models addressed this need by providing a framework for translating variable-length input sequences to variable-length output sequences. Seq2Seq models typically consist of two recurrent neural networks (RNNs).

- **Encoder:** Processes the input sequence and compresses it into a fixed-length context vector.
- **Decoder:** Generates the output sequence based on the context vector.

The Seq2Seq framework significantly improved NLP tasks by enabling the transformation of one sequence into another, facilitating applications such as machine translation, text summarization, and dialogue generation. For instance, in machine translation, an English sentence could be encoded into a context vector and then decoded into a French sentence. This approach marked a significant improvement over previous methods that struggled with variable-length inputs and outputs [SVL14].

These models were further innovated by implementing Long Short-Term Memory (LSTM). LSTMs addressed the vanishing gradient problem in traditional RNNs, by integrating gating mechanisms known as the input, forget and output gates, allowing Seq2Seq models to capture long-range dependencies. LSTMs were particularly effective in mitigating the issues related to long-term dependencies, which were a significant limitation in traditional RNNs. The gating mechanisms of LSTMs allow them to retain relevant information for extended periods, making them suitable for tasks that require understanding and generating long sequences of text. This capability significantly enhanced the performance of Seq2Seq models in applications like language modeling and machine translation [HS97].

Seq2Seq models were a significant advancement in NLP, enabling more advanced tasks like translation, summarization, and dialogue generation. However, the fixed-length context vector often led to information bottlenecks, especially for long sequences. This bottleneck issue arises because the entire information of the input sequence must be compressed into a single vector, which can be challenging for lengthy or complex sequences [CvMBB14].

4.3.4 Development of modern LLMs

To address the limitations of fixed-length context vectors in Seq2Seq models, Bahdanau introduced the attention mechanism. This mechanism allows the model to focus on different parts of the input sequence when generating each word in the output sequence. This provided the model architecture two additional distinctive layers, the attention layer and the context vector. The attention layer computes a set of weights that specifies the relevance of each input word to the current word being generated. The context vector is then a weighted sum of the input words, that are dynamically computed for each output word [BCB16].

This was later innovated upon, in the paper "Attention is All You Need" by Vaswani et al., when they introduced the concept of the self-attention mechanism. This allowed the model to relate several different positions in a single sequence effectively, leading to the development of the Transformer architecture [VSP+23].

Rapid development of newer models
After the introduction of the self-attention mechanism, the development rate in field of NLP skyrocketed, and in the subsequent years, several models based on this architecture were developed by some of the leading tech companies and developers known today. BERT (Bidirectional Encoder Representations from Transformers) was first presented by Devlin and his team, where they utilized the transformer architecture to train deep bidirectional representations in the contexts of a given text. This made BERT highly effective for a wide variety of NLP tasks [DCLT19].

In the same year, the first model in the now well-known GPT [1] series, ChatGPT, demonstrated the power of unsupervised pre-training followed by fine-tuning for specific tasks. GPT3 with its 175 billion parameters showed its impressive capabilities to generate human-like text [BMR+20]. Not long after the T5 model considered all NLP tasks as text-to-text problems, which united the approaches to various tasks such as translation, summarization and question answering in a single model architecture [RSR+23].

The evolution of LLMs from traditional NLP methods to modern Transformer-based models has been marked by continuous innovation and improvements. The introduction of word embeddings, Seq2Seq models, and the attention mechanism has collectively transformed the field of NLP, enabling the development of powerful models capable of understanding and generating human language with significant efficiency. This collectively lead to the Large Language Models as they are known today.

4.3.5 Capabilities of Large Language Models

LLMs have significantly advanced the capabilities of automated systems in processing, generating, and understanding human language. This has enabled a high innovation rate in the field, as well as some prolific use-cases [GG23]. Some of the key capabilities and seen use-cases of LLMs are the following:

Text Generation and Completion: LLMs can generate coherent and contextually relevant text based on input prompts. This capability is particularly useful, for example when drafting patent claims and descriptions where specific legal and technical language is paramount for a succesful process [LH20a].

---

[1] Generative Pretrained Transformer. Not to be confused with General Purpose Technology

Semantic Textual Understanding: These models excel in understanding the underlying meanings of texts. In the context of patents, this means LLMs can accurately interpret complex technical descriptions and legal jargon, aiding in more effective information retrieval and analysis[ZWZ+].

Language Translation: LLMs, given they have been trained on more than one language, have the ability to translate directly between different languages [ERC+20]. Given the global nature of patent filings, LLMs can automatically translate patents into various languages, ensuring that innovations are accessible to international examiners and stakeholders without language barriers.

Text Summarization: LLMs can condense long documents into concise summaries, preserving key information and insights [VUB+23]. For patents, this can reduce the time needed to review lengthy documents by providing quick insights into the core invention and its claims, which tend to be lenghty process due to the detail requirements.

Sentiment Analysis: LLMs can gauge gauge the tone and implications of texts within a given context such as research papers and technology. One example of such a use-case is FinBERT, that was made to perform sentiment analysis in text regarding different types of finance [Ara19]. Although rarely applied in patent analysis, sentiment analysis can be used on papers that discuss patented technologies, potentially identifying public or market perceptions of various innovations.

Question Answering: LLMs can be fine-tuned to develop question-answering systems that provide direct answers to specific questions posed by users. One example use-case of this can be found in the medical field, where Med-PaLM2 had been fine-tuned to answer medical questions from patients, based on a vast dataset of medical information combined with some additional methods to ensure a refined response [STG+23]. A similar approach could be used, were details about patent ownership, prior art, and technology scope can be fed into a model. This can significantly speed up research phases during patent landscaping and competitive analysis.

4.3.6 Methods used in conjunction with Large Language Models

Since these models became more popular, they have been heavily researched not only in scientific literature, but likewise in the AI open source community. This has resulted in several methods that enhance the capabilities of LLMs significantly, as well as reducing the workload related to utilizing these in more or less every given context.

One such method is Retrieval-Augmented Generation (RAG), which is a method that injects new contextual information into the prompt of an LLM, in order to provide the model with more recent information. This reduces the rate at which models need to be updated, as the information they contain in their neural networks become obsolete. Instead information retrieval systems can be designed around the model, to inject new relevant information that the model can then directly analyze and respond to, via the capabilities described in the previous section [GXG+24].
Few-shot and zero-shot learning are two other methods that enhance LLMs' flexibility and efficiency. Few-shot learning allows models to generalize from a small number of examples, significantly reducing the amount of labeled data needed for training on new tasks. Zero-shot learning, on the other hand, enables models to perform tasks they were not trained on by utilzing their understanding of related tasks. These methods have proven particularly useful in adapting LLMs to new domains with minimal data [BMR+20].
Prompt engineering is another technique that involves carefully engineering input prompts to guide the behavior of LLMs more effectively. By designing prompts that provide clear instructions and context, users can generate more accurate and relevant responses from the models. This approach makes use of the inherent capabilities of LLMs without requiring changes to their underlying architecture, making it a practical and powerful method for optimizing model performance in specific applications [RM21].

4.3.7 Applications of Large Language Models in the context of patents

The introduction of Large Language Models (LLMs) has revolutionized numerous fields by leveraging advanced natural language processing capabilities to interpret and generate human-like text. In the realm of intellectual property (IP), particularly in patent analysis, LLMs have emerged as powerful tools that streamline various processes, enhance accuracy, and provide strategic insights. This section explores the diverse applications of LLMs in the context of patents, highlighting their potential to transform how patents are classified, searched, analyzed, and managed. The integration of LLMs into patent analysis not only augments traditional methodologies but also paves the way for innovative approaches to handling the ever-growing corpus of patent literature.

Large Language Models (LLMs) are adept at categorizing text into predefined categories, which is essential in the patent domain where accurate classification can direct patents to the appropriate examination channels and facilitate efficient prior art searches. Using LLMs, patents can be automatically classified according to the Cooperative Patent Classification (CPC) system or the International Patent Classification (IPC) system based on their textual content. This automation not only speeds up the classification process but also enhances consistency in how patents are categorized across different jurisdictions [LH20b].

The vast and growing volume of patent documentation makes manual search and retrieval challenging. LLMs transform patent search by understanding complex queries in natural language and retrieving relevant documents with high precision. Unlike traditional keyword-based search engines, LLMs can interpret the searcher's intent and contextual nuances, allowing them to return results that are contextually related rather than merely textually similar. This capability is particularly beneficial for identifying non-obvious prior art during patent examinations or freedom-to-operate analyses [KLLL20].

LLMs can analyze large datasets of patent texts to identify emerging trends and patterns in technology. By processing the language used across thousands of patent documents, LLMs can detect shifts in terminology and the emergence of new concepts before they become mainstream. This application is invaluable for strategic planning and innovation management, as it provides early warnings of technology shifts and potential areas of technological convergence [KL].

Through semantic analysis, LLMs can uncover relationships between different patents that may not be evident through traditional citation analysis. By understanding the content of the patents at a deeper level, LLMs can suggest potential technology linkages and collaborations that could spur innovative developments. This aspect is crucial for companies looking to innovate through partnerships and for academic institutions aiming to translate research into practical applications [SYAN20].

Patent filing and litigation involve a significant amount of legal documentation and compliance with specific regulatory standards. LLMs can be trained to recognize and generate legally precise text, aiding attorneys and patent agents in drafting claims and descriptions that meet stringent legal requirements. Additionally, LLMs can monitor changes in patent law and regulations across different regions, providing timely updates to stakeholders to ensure that patent filings remain compliant [AA].

The integration of Large Language Models into patent analysis represents a significant advancement in how intellectual property can be managed and leveraged. As these models continue to evolve, their impact on the efficiency, accuracy, and strategic insight of patent-related activities is expected to grow, transforming the landscape of innovation management. This section of the literature review highlights the practical implications of LLMs in the dynamic field of patent analysis. This provides a peek into a future where AI and patents intersect to drive forward the next wave of technological innovation.

4.3.8 Limitations of Large Language Models

Despite their remarkable advancements, Large Language Models (LLMs) still face several critical challenges and limitations that researchers and practitioners must address. A primary concern revolves around the interpretability of these models. Unlike traditional machine learning algorithms, LLMs often operate as "black boxes," making it difficult to understand the reasoning behind their outputs. This opacity can pose significant issues in high-stakes fields such as healthcare and legal applications, where transparency is crucial for trust and accountability [Rud19].

Moreover, LLMs are not immune to the biases present in the data they are trained on. These models can inadvertently learn and perpetuate societal biases, leading to unfair and discriminatory outcomes. For instance, biases related to gender, race, and culture have been observed in word embeddings and other language models, meaning it is necessarily an ongoing efforts to detect and mitigate these biases to ensure fairness and equity [BCZ+16].

The computational resources required to train and deploy LLMs represent another significant limitation. Models like GPT-3, with its 175 billion parameters, demand extensive computational power and infrastructure, which translates to high costs and environmental impacts. This resource intensity limits the accessibility of cutting-edge LLMs to well-funded organizations and raises concerns about the sustainability of such models [SGM19].

LLMs also struggle with maintaining context over extended passages, often producing outputs that are contextually appropriate but factually incorrect. This limitation is especially critical in domains that require precise and reliable information, such as scientific research and legal documentation. Enhancing the ability of these models to retain and utilize long-term context remains an ongoing area of research [BGMMS21].

The performance of LLMs is heavily dependent on the quality and diversity of their training data. If the data is limited, outdated, or lacks diversity, the resulting models may exhibit poor generalization and reliability. Ensuring comprehensive and high-quality datasets is essential but challenging, highlighting the need for continuous data curation and updating [DSM+21].

Ethical and privacy concerns also arise with the deployment of LLMs. These models can generate harmful content, including misinformation, hate speech, and inappropriate material. Furthermore, the use of sensitive data in training these models raises significant privacy issues. Establishing robust ethical guidelines and regulatory frameworks is crucial to ensure the responsible use of LLMs and protect user privacy [FC22].

Scalability is another challenge that LLMs face. While they have demonstrated impressive capabilities in various NLP tasks, extending these models to new domains or languages often requires substantial data and computational resources. Additionally, fine-tuning LLMs for specific tasks can be resource-intensive and may not always yield significant improvements, underscoring the need for more efficient and scalable approaches [CKG+19].

Finally, despite their proficiency in generating human-like text, LLMs have some limitations in terms of true understanding and reasoning capabilities. They rely heavily on patterns in the data rather than genuine comprehension. This limitation becomes evident in tasks requiring deep understanding, logical reasoning, and common-sense knowledge, where LLMs often fall short compared to human expertise [Mar20].

Addressing these limitations involves ongoing research focused on improving model interpretability, reducing bias, enhancing computational efficiency, and developing ethical frameworks for AI deployment. Hybrid models that combine LLMs with symbolic reasoning, transfer learning techniques, and domain-specific adaptations are also being explored to create more robust and versatile language models. As the field progresses, overcoming these challenges will be crucial for fully harnessing the potential of LLMs in various applications.

4.4 Technology mapping using LLMs

The following section 4.3 will explore the field of technology mapping, specifically within the already explored topic of GPTs, as well as the potential for analyzing patent data using LLMs. The current scientific literature surrounding technology mapping seems to focus on the development of indicators for each type of technology. This makes sense, as each technological field comes with its own challenges, goals and thereby measurements of success in achieving these.

In the article ‘Mapping general purpose technologies with patent data’, Sergio Petralia specifically focuses on one of two archetypes of technology, being General Purpose Technology (GPT) which is characterized by continuous development and widespread usage across several technological sectors, by serving as a complement of already existing and emerging technologies as a result of innovation [Pet20]. This notion he borrowed from another paper published in 1995, where it was concluded that the characteristics similarly make GPTs valuable in terms of growth potential across different sectors, as they can provide significant advantages in terms of industry and/or technology sector outputs [BT95]. The implication of GPTs as a concept however is still argued, as the measurement of their characteristics and the usefulness of the concept, as it is not easily identifiable due to the inherent interdependence of different technological fields [LCB05].

However, building upon the current understanding of GPTs and the role across technological sectors, it becomes important to explore innovative methodologies that can further enhance the analysis and interpretation of these data landscapes that are contained within patents. Whilst Sergio grouped patents into economically relevant categories, in order to identify what he elegantly described as the 'GP-ness measurements', translated to 'General Purpose-ness', this did not provide a way to directly identify a specific innovation/technology as a GPT, but instead the overall occurrence and/or centralities of GPTs within these economic groups [Pet20].

This is where the possibility of extending the methodology presented lies, as the recent developments within the field of natural language processing (NLP), namely in shape of Large Language Models (LLM) have shown significant opportunities to perform large scale in-depth analysis of any given data, as long as the data pipeline work [WTB+22]. In addition, the surrounding methods for improving performance for downstream tasks such as data analysis, has had significant improvements by utilizing the so-called context windows (or token limits) of these models to inject relevant information to the query in a method known as Retrieval-Augmented Generation (RAG), as presented earlier. This method can be seen as a high-level form of prompt engineering, where the prompt is dynamically influenced by the query, as well as the subsequent data-retrieval in order to alter the model behavior. This has been shown to reduce the need for fine-tuning, which is costly both in terms of data collection and cleaning, as well as computation power and time required to correctly monitor and test these models before deployment [LT24].

Combining methods within patent analysis, specifically network analysis to identify central technologies, and combing this with a qualitative analysis performed by an LLM could potentially allow for a stronger identification process of these GPTs. By utilizing the inherent NLP potential of LLMs, I should be able to extract more relational information from patents by applying RAG to inject contextual information into a well-engineered prompts. This will provide the model with not only the instructions on how to perform the analysis, but also the relevant information to perform it. In addition, this should allow for a more in-depth analysis of their possible implications on technology in a more holistic manner. This approach aligns both with the continuous evolution of technology mapping methods, and may also provide a more nuanced understanding of this field that more traditional methods have so far overlooked.

4.5 Summary of literature review

To summarize the literature review of this paper, the following topics have been explored.

As detailed in section 4.1, the theory and impact of innovation has been heavily researched for several years. The potential ability to disrupt and transform a market by introducing a brand new form of technology, is a threat that most if not all firms face in one way or another. Therefore, having the ability to identify these potential threats remain a significant advantage for firms, especially those that are heavily invested in the current landscape. This is why understanding the driving forces of innovation and potential disruption becomes important not only for researchers, but for companies alike. I explored one of these driving forces that is a result technology itself, as a concept known as General Purpose Technology (GPT). This conceptual type of technology has the capability to function as a catalyst for further innovation across several if not all technological sectors, whilst also undergoing continuous development itself. It therefore also has the ability to cause major disruptions on several markets at the same time, and depending on the technology, permanently change industries and create new ones. This means that identifying and better understanding these technologies can help further the research as well as the strategic advantages of firms that either utilize or are directly affected by them. These technologies however have a characterization that doesn't lend itself to analysis within a single given field, but instead one must analyze all technological fields on a macro-level. Instead, however, it should be possible to identify what drives these types of technology by analyzing the patents within the field on a micro-level.

In section 4.2, I also explored the field of patent data and some of the methods that are used to analyze patents in the context of innovation and technology. As discussed in literature, patents contain vast amounts of data related to technologies, which enables research into the production of knowledge and research of new technology. It also allows an insight into the patterns that connect innovation pathways and thereby the flow of knowledge, by providing citations both to scientific literature and other patented technology that enabled a new technological development. This provides an framework for understanding the process of identifying innovation clusters that could hold underlying factors that drive the characteristics of GPTs through network analysis, as well as identify key inventors and how knowledge is applied. However, the literature also revealed that relying solely on quantitative methods have certain limitations, as it does not provide an understanding of how a particular technology may be able to cause an impact on other technology as well as markets. For this purpose, the method must also utilize a qualitative approach where by full-text data becomes relevant to extract meaningful insights into the potential impact of a given technology.

This is where a potentially significant use-case for Large Language Models can be found, as detailed in section 4.3. These models that are effectively a significant advancements to more traditional Natural Language Processing, which provide us with the possibility of extending the framework of patent network analysis with a qualitative dimension. As patent full-text data is often quite technical in its language, the inherent understanding of human language that LLMs hold, and their ability to summarize text can be used to perform high-level analysis on several types of technology. By creating a system for retrieval-augmented generation (RAG) for the patent data, combined with effective prompt engineering, the analysis can be steered in a direction set by the developer. This means that it becomes feasible to provide the LLM with a contextual framework for analyzing the patents and derive better insights from full-text data from patents, as well as creating a framework for answering technological questions from the LLM

By creating this combined framework of quantitative network analysis and subsequent qualtitative analysis via LLMs, relevant actors can potentially identify the driving forces of GPTs and analyze their individual impact on entire technological sectors. This could potentially aid in identifying knowledge gaps and new ways of combining technologies that have not yet been seen. The value that can be generated for companies and governments by providing a more fulfilling technology mapping methods is significant, in terms of both the impact on investment and research of key technological advancements.

5 Theory and Methodology

In the following section, I will present the core methodology for the proposed framework, and explain each components theory.

5.1 Objective of the project

As presented in the literature review, General Purpose Technology exhibit distinct characteristics. These characteristics is what enables the ability to function as a catalyst for innovation across technological fields, whilst remaining under continuous development itself. These characteristics however, also mean that the techniques and innovation within the field, that function as driving forces of a GPT cannot be evaluated through purely quantitative methods, but should be supplemented by qualitative evaluation to assess their impact on the GPT as a whole. Therefore the first objective of this project will be to determine whether or not AI can be considered a GPT through quantitative analysis of patent data. The next objective will be to identify central patents within the field, using methods within network analysis. Finally, I will attempt to implement a question-answering system using RAG-methods, in order to enable an LLM to subsequently analyze the most central patents in a more qualitative manner.

As explored by Sergio Petralia, one of the exihibited characteristics of a GPT, is the widespread usage across technological fields, which enables the opportunity to identify possible candidate GPTs through patent analysis [Pet20]. This subsequently should enable the assumption that one of the drivers of these GPTs should lie somewhere in the center of technological innovation clusters, which likewise can be identified through patent networks. In more direct terms, the driving force that should be identifiable, should be the most utilized techniques used in the more central patents. By identifying these clusters of innovation, and subsequently identifying which technological patents are central to these clusters, the driving forces of GPTs can be indicated as per the aforementioned characteristic. In addition, certain data variables of technology patents can be used to identify not only the innovation clusters, but likewise their origin, as detailed in the end of section 4.2 [FY11].

Identifying the origins of patent filings, can reveal which global regions experience the highest level of innovation within the field, and enable the analysis of what drives innovation within these regions.

Analyzing the occurrence of patent filings across the patent timelines can also help determine the continuous development, and subsequently be used to qualitatively determine roughly when the technology became a GPT.

In addition, we can analyze the patent holders or inventors, and subsequently explore which industries they belong to, in order to determine the spread of AI across industries. This will ultimately help determine whether AI and ML can be considered a General Purpose Technology, which is more or less essential to the direct objective of the project, which as mentioned earlier is to identify the driving forces of these within the technology itself.

5.2 Database Selection

For the purpose of this study, data was sourced from the PATSTAT database, specifically the 2022 autumn version, which is a comprehensive resource maintained by the European Patent Office. PATSTAT is globally recognized for its extensive coverage of patent data, offering detailed bibliographic and legal event information on patent applications from over 100 patent offices. This database is instrumental for researchers requiring historical and statistical patent information, making it an ideal choice for investigating the technological domain of machine learning and artificial intelligence [Off]. The fact that only the 2022 version is utilized, does mean that it will not be possible to analyze the most recent patented technology, but it will suffice for an experimental approach.

International Patent Classification (IPC) The International Patent Classification (IPC) system, under which patent documents are classified, facilitates the orderly arrangement of a vast number of patents into a manageable system of distinct technology sectors. The IPC provides a hierarchical system of language-independent symbols, each denoting a specific category of technology. This system enables users to locate patent information across multiple languages and legal jurisdictions, ensuring comprehensive and systematic access to global patent information [Ada01].

Rationale for Focusing on Specific IPC Codes
For this study, the focus was specifically directed towards IPC codes relevant to machine learning and AI. The use of predefined IPC codes leverages an established and internationally recognized classification system to ensure precision and relevance in selecting patents. The decision to limit the analysis to these IPC codes was driven by several key considerations:

Relevance and Precision: By concentrating on specific IPC codes, the study targets patents directly relevant to the fields of machine learning and AI. This approach minimizes the inclusion of unrelated patents that could dilute the analysis, ensuring a high degree of precision in the data collected. On the other hand however, a limitation occurs on background of the characteristics presented in section 4.1.3. Specifically the widespread usage across different technological fields. It is worth noting however, that the patent abstracts should contain textual information, that can be used to qualitatively assess which industry a patent belongs to. Additionally, identifying central clusters of innovation, and subsequently evaluating these central patents qualitatively through LLMs, should still prove to be a strong indication of their potential to be drivers of AI as a General Purpose Technology.

Technical Implementation: The targeted IPC codes represent the cutting edge of machine learning and AI technologies. Focusing on these areas allows for a detailed examination of the technical advancements and trends within this specific sector, providing insights into the evolution of these technologies and their application across various industries. In terms of the LLM implementation, this will complement the IPC-based selection by providing a deeper analysis of the patent texts, by possibly identifying nuances of technology that IPC codes alone might miss. This should enhance the overall robustness and depth of the analysis, enabling the identification of central innovation clusters and their potential impact across various fields.

Manageability: Machine learning and AI are vast fields with applications spanning numerous sectors. Narrowing the scope to specific IPC codes makes the dataset more manageable, allowing for a more focused and in-depth analysis of this field in particular. However, as this project serves as an experimental proof-of-concept, scaling the final framework across different sectors should only be a matter of time and available processing power.

Comparative Analysis: By using specific IPC codes, the research facilitates comparative analysis over time and across different technological sub-fields. This is essential for understanding the landscape of patent filings and innovations in AI and machine learning, providing a clearer picture of technological and regional innovation trends. This allows further analysis on the origins of innovation, which could provide insight into the significant inventors, companies and other relevant entities within the given sector.

Keyword-Based Selection To complement the IPC-based classification, keywords related to machine learning and AI were employed to refine the dataset further. This dual approach ensures that the patents analyzed are not only categorized under relevant IPC codes but are also directly associated with machine learning and AI through their textual descriptions. Keywords such as "machine learning' and 'artificial intelligence' were used to filter the patents, ensuring the inclusion of the most pertinent and impactful innovations in the field. This also further extends the initial scope, to hopefully capture some of the linkages to technologies in other fields, which maybe further help indicate significant centroids in these innovation systems.

This meticulous approach to data collection—combining the structured hierarchy of IPC codes with precise keyword filters—provides a robust dataset tailored for a comprehensive analysis of machine learning and AI innovations, their evolution, and their technical and commercial implications.

5.3 Data selection

Patent data contains a wide variety of information related to the individual intellectual property that is being sought protection for. This includes data such as the applicant(s), inventor(s), patent number, title, abstract as well as a list of claims that very specifically explains exactly what intellectual property that is being patented. Utilizing various methods within network graphing, natural language processing and database management, this data can be used to extract information that exactly depicts what topic or technological sub-field each patent fits under. In this case, the following data was extracted from the patents for further processing to conduct a thorough analysis of machine learning and artificial intelligence (AI) innovations. These datasets are crucial for understanding various aspects of patent activities, such as literature citations, inventor details, and bibliographic data. Some variables, such as application ID will repeat across the datasets, due to the inherent nature of relational databases, which is the case of PATSTAT, where some variables are used as identifiers in a particular table. Each dataset serves a distinct purpose in the study, allowing for a multi-faceted examination of the patents.

Patent Literature Dataset:

- appln_id: This is an identifier variable that is unique to each specific patent, thus allowing cross-referencing different tables for relevant data related to a specific patent application
- cited_pat_publn_id: Identifier for a cited patent publication
- pat_publn_id: Identfier for a citing patent publication

Purpose of patent literature data: This dataset provides data on citations between patents, offering insights into the influence and relevance of specific patents within the field of AI and machine learning. Analyzing citations helps identify key technologies and inventions that have shaped the landscape of AI, illustrating both the knowledge flow and technological dependencies within the field.

Inventor Dataset:

- appln_id: Unique identifier
- person_name: Name of an inventor of a given patent

- person_address: Address of the inventor in question.
- person_ctry_code: Country code for the inventor in question

Purpose of inventor data: This dataset captures details about the inventors, including their names, addresses, and country codes. Such information is essential for geographical and demographic analyses of innovation, revealing where major AI and machine learning innovations are being developed and by whom.

Bibliographic Dataset:

- appln_id: Unique identifier.
- appln_title: Title of patent application
- appln_abstract: Abstract explaining the precise scope of the invention.
- appln_filing_date: The date on which the application was filed.
- publn_date: The date on which the patent was published.

Purpose of bibliographic data: The bibliographic details provide fundamental data about each patent application, including titles, abstracts, filing dates, and publication dates. This dataset is key to understanding the content and scope of inventions, enabling content analysis for technology scouting and trend analysis. It also facilitates tracking the evolution of AI technologies over time.

Why These Datasets?
The selection of these specific datasets was driven by the need to comprehensively understand the dynamics of patenting in AI and machine learning. By examining legal, citation, inventor, and bibliographic data, the study aims to construct a detailed picture of how AI technologies are evolving, how they are interconnected, and what their impact might be on future technological developments. This holistic approach ensures a robust analysis, capable of supporting strategic decisions in policy making and business strategy within the high-tech sector of artificial intelligence and machine learning. Furthermore, I will perform a short exploratory data analysis of these datasets for indication of AI as a GPT through the given characteristics.

5.4 Approach to network analysis

The analysis will begin by constructing a patent network that enables an analysis of the relationships between patents to illustrate how technologies interconnect and influence one another. This network should enable the identification of clusters of technology and the key technologies that serve as bridges in the technological landscape. The selected data elements are essential for constructing a multidimensional network analysis that not only links patents by citations but also by shared technology categories and inventors.

Constructing the network

- Node Creation: Each patent application (appln_id) serves as a node. Nodes are tagged with bibliographic data (title, abstract) to contextualize each patent within the network.
- Edge Formation: Connections (edges) between nodes are established based on:
  - Citations: Direct citations (cited_pat_publn_id and citing_publn_id) create a link, reflecting the technological lineage and influence among innovations.
  - Inventor Connections: Patents sharing common inventors suggest a thematic or collaborative link, offering insights into innovation clusters.

Analytical Techniques:

- **Centrality Analysis:** Identify key patents (nodes) based on their eigenvector centrality in the network. A high eigenvector centrality suggests that a node is well connected to other very central nodes. This is where it should be possible to identify the most influential technologies that may serve as foundational driver for the General Purpose Technology.

- **Community Detection:** In order to also identify thematic keywords within technological groups, I will utilize well-known algorithms. Specifically the Louvain detection algorithm. This can be used to detect the communities within the network, which can reveal clusters of related technologies or fields that are densely interconnected.

Network Centrality Analysis

The centrality analysis will be the initial indication for identification of potential GPT drivers. Whilst they can each have their own meaning and implication on the analysis, collectively they can be compared for greater insights. Calculating centralities will also function as a filter for the LLM implementation, as running all patents through an API can result in certain token limitations due to the sheer amount of these.

- **Degree Centrality:** Degree centrality measures the number of edges connected to a node. In a directed network like our patent citation graph, I consider both in-degree (number of incoming edges) and out-degree (number of outgoing edges) centrality. In-degree centrality in a patent network highlights patents that are frequently cited by other patents, indicating their influence and foundational nature within a particular technology domain. Conversely, out-degree centrality identifies patents that cite many other patents, suggesting they are derived from existing technologies and may represent transitional technologies or enhancements [BJT23].
- **Betweenness Centrality:** Betweenness centrality measures the extent to which a node lies on the shortest path between other nodes in the network. This metric reflects the ability of a node to act as a bridge within the network. Patents with high betweenness centrality can control the flow of information in the network and may be crucial in connecting different technological areas. Such patents are potentially critical for the diffusion of new ideas and can be pivotal in technology integration processes [Fre77].
- **Eigenvector Centrality:** Eigenvector centrality considers not only the number of connections a node has but also the quality of these connections. It assigns relative scores to all nodes in the network based on the principle that connections to high-scoring nodes contribute more to the score of the node in question. In the patent network, eigenvector centrality helps identify patents that are influential not just through their direct connections but also through their indirect influences over the network. This is particularly useful in recognizing patents that may not have a high volume of direct citations but are influential due to their strategic position within a network of highly influential patents [Bon72].

These centrality measures collectively enable the identification of significant patents that shape the technology landscape. By analyzing patents that score high across these metrics, researchers and organizations can pinpoint influential innovations that serve as key nodes within the broader network, guiding strategic decisions in research development, technology acquisition, and innovation policy. Additionally, visualizing these centrality measures using network plots can provide intuitive insights into the structural importance of each patent within the network. [BJT23]. However, due to the comprehensive size of this particular network on account of the hundreds of thousands patents I will be utilizing, a visualization of these do not provide any meaningful insight, and instead I will rely on descriptive statistics of the network to perform the analysis.

This detailed approach not only specifies the mathematical computation of centrality measures but also clarifies their practical implications in identifying and understanding the role of critical patents within the technological ecosystem.

Community Detection

As presented earlier community detection in network graphs is a powerful analytical tool for uncovering the modular structure of networks, where nodes are more densely connected within the same community than with nodes in other communities. This method is particularly useful in patent network analysis as it helps to identify clusters of patents that represent closely related technological domains or thematic areas. Community detection algorithms partition the network into communities or clusters, where, in this case, patents within the same community are more closely related to each other than to those in different communities. This method provides insights into how innovations are grouped and interrelated, potentially uncovering the structural patterns of technology development [For10]. Below are details on the implementation and application of community detection:

- **Algorithm - Louvain method:** The Louvain method is an efficient and widely-used community detection algorithm that optimizes the modularity of a network locally. Modularity is a scale of the density of links inside communities as compared to links between communities. The algorithm iteratively groups nodes that achieve the highest modularity until it cannot be increased further [GHT+18]. In the context of patent networks, the Louvain method can be used to detect clusters of patents that may represent emerging technologies, specific research focuses, or closely related sets of innovations. This clustering helps in understanding the landscape of technological advancements and the relationships between different innovations.

- **Analysis of Detected Communities:** Analyzing the results from the Louvain method provides a map of the technological landscape captured by the patent data. Each community detected will contain some central patented technologies, which can be extracted for the qualitative analysis.

Community detection can not only reveal the clusters of related technologies but also helps in predicting technology trends by observing the evolution of these clusters over time. It allows researchers and strategists to identify core areas of technological development and potential areas for investment or research focus. By understanding the community structure, organizations can better navigate the technology landscape, aligning their research and development, as well as innovation strategies with the most active or promising areas of technological development.

5.5 Utilizing Large Language Models (LLMs)

To deepen the analysis, integration of Large Language Models will be used to process and analyze the textual data associated with the most central patents. The technical aspects of this approach will be explored in further detail in the section on technical implementation. This approach leverages the capabilities of LLMs to perform semantic analysis for labeling a sample of patents based on pre-defined categories. In addition I will be utilizing a RAG-based approach to generate a question-answering system, which should ideally help enhance the understanding of technological content and connections.

The final RAG implementation will likewise be tested using a range of metrics such as precision, recall and mean reciprocal rank. Precision is the measure of the percentage of retrieved patents that are relevant to the question and recall is a measure the for the percentage of relevant documents that are retrieved. These definitions can be a bit hard to intuitively understand. In less confusing terms, if the system retrieves a non-relevant document for a query, this would be considered a false positive, as it is retrieving the wrong patent for a given query because it has been wrongly evaluated as being relevant to the query. Oppositely, if the system fails to fetch a patent that is in fact relevant to the query, this would be considered a false negative. This means that high values of precision and recall both correlate with a well-functioning system, as they indicate that relevant documents are being retrieved correctly [MRS08].

	Relevant	**Nonrelevant**
Retrieved	true positives	false positives
Not retrieved	false negatives	true negatives

Table 1: Confusion matrix for information retrieval

Mean reciprocal rank (MRR) on the other hand, is an average measurement done across several test queries. In this case it basically creates a ranked list of patents for each query, and then denotes the rank of the first relevant patent in this list. This rank is then used to divide 1, so that if the rank is k, the reciprocal rank is 1/k. Finally, the average of these reciprocal ranks are calculated which provides the final score. A higher MRR indicates that the system is consistently able to retrieve relevant documents across different queries [VT00].

Cosine similarity scores between the query and retrieved documents will also be evaluated, as these indicate how semantically similar the two are, and therefore also functions as an indicator as to how relevant the retrieved documents actually are to the query. Cosine similarity are measurements made between the embedding vectors of the query and the retrieved texts in a multi-dimensional space, and provides an output between zero and one, with one indicating that the vectors, or in this case patents are completely identical [MCCD13].

Implementation steps

- 1. Question answering: Use an LLM to answer questions about specific inventors, specific types of technology, through indirect text extraction strategies. More specifically, the proof of concept LLM should be able to answer at least simple questions, and ideally provide comprehensive answers, such as the most occuring type of technology, given a type of AI application.
- 2. Semantic Linking: Assuming the first step succeeds, it should be possible to apply LLMs reasoning power to identify semantic links between patents beyond direct citations. Without going into the specific technical details at this point, this involves applying a way for injecting contextual similarities from the technology descriptions, into the context window of an LLM.
- 3. Evaluation of RAG implementation: Utilizing the proposed measures, I will evaluate the systems ability to fetch relevant documents across several queries that will vary in terms of specificity, as it it important to evaluate the systems ability to answer question related to a specific type of technology within the field, as well as the field in a general manner.

5.6 Combining the two methods

Next I will attempt to combine the insights from the network graph analysis and LLM analysis to create a comprehensive representation of this particular technological landscape. This will fuel the subsequent discussion of the results, as combining these insights, ideally should provide for a more holistic view of the technological field. From a theoretical perspective, the output of this project will be to generate an initial methodical framework that can be utilized to identify GPTs within any given technological field. From a more technical perspective, the ideal output will be to enable the users to use tools like Gephi for network visualization or a custom interactive dashboard that allow exploration of different layers of data, such as legal status, technological linkage, and innovation trends. There is also the possibility of enabling an LLM to utilize the patent abstracts to answer questions related to a specific type of technology within the field, by utilzing a method known as RAG, as detailed in section 4.3.6.

From a business, or perhaps more importantly, from a managerial perspective, the final output will have the purpose to empower business managers and other relevant decision makers to more easily create the insight they need to generate value through the exploitation of the technological fields in which they operate. In more direct terms, having the ability to ask an AI questions pertaining to their specific invention and/or technology, and how they might enhance them using what already exists. This can fuel generation of insights created from forecasting methods presented in the literature review, such as scenario analysis.

From the academic perspective, this framework, aside from the potentially contributing to studies of identifying the underlying drivers of GPTs, should also help indirectly identifying potential knowledge gaps in the different technological fields. Since GPTs are known to function as a catalyst for innovation, there will by definition, inherently be knowledge gaps surrounding them waiting to be explored further.

6 Technical implementation

The technical implementation of the project is structured to handle the complexity and scale of patent data analysis, integrating several components including database management, network analysis, natural language processing, and a possible user interface for the opportunity of easy interaction and more comprehensive testing. This section outlines the process and technologies involved in each step. It should be seen as the operationable steps based on the proposed methodology, thus some information will be repeated from that.

6.1 Database Setup and Management

Objective: Efficiently manage and query large volumes of patent data.

Database Selection: As mentioned earlier, this project will make use of the PATSTAT database, which provides extensive patent data across multiple jurisdictions. The database includes legal status, bibliographic details and citation information essential for this study.

PATSTAT as presented earlier, is what is known as a relational database, containing several related tables which in turn each contain several variables pertaining to specific applications. These tables are all interconnected through different variables, such as application ID as well as other variables depending on the specific table. Whilst it would in the case of generating a question-answering system, perhaps be more fruitful to include even more variables than those chosen, the particular scope of this project remains mainly on identifying GPTs, with the question answering being more of an additional proof of concept application. Including further variables to the network analysis is more likely to only dilute the analysis, and make it more difficult to comprehend the final results. Instead, as is the case with many types of AI based system, it would be better to aim it at a specific use-case as have been done with this analysis framework.

Data Extraction: Develop scripts using Python and SQL (Structured Query Language) to extract data based on selected IPC codes relevant to AI and machine learning. This includes filtering data by specific keywords to refine the dataset further.

As mentioned above, relational databases are interconnected through specific variables such as ID's. Additionally these database utilize so-called indices which allow for more efficient data queries. The difference between correct and incorrect use of these indices, can manifest itself primarily in the database processing time spent to fetch data from the database[2] This can result in less than an ideal situation, where just one of the generated datasets could take 2240 hours to complete. In contrast, utilizing these table indices enabled the same dataset to be generated in one or two hours, as the database processor does not have to look up all entries in the table, but can instead focus on the indexes of the queried application IDs.

In order to utilize of these indexes efficiently, as well as limiting the data queries to the scope of the project, it was chosen to first make use of a two-step approach to identify and query the data needed. The first of these steps was text mining of the application abstracts and titles, searching for keywords such as "artificial intelligence" and 'machine learning'. The application ID's of these queries where then saved for the final query setup. In order to ensure that all possible technology patents within the field was fetched, IPC codes were also used for filtering for relevant application IDs where where then subsequently fetched and added to the ones fetched from text mining. Collectively this resulted in 448 thousand unique application ID that were then used for the queries for each dataset.

Data Cleaning: Perform data preprocessing to handle inconsistencies, missing values, and data normalization. This step ensures that the data used in network construction and analysis is accurate and reliable.

The data cleaning in this case, mostly consisted of filtering out patents that were not connected any other patent through citations. The main argument for this decision is two-fold.
First and most importantly, if a patent did not cite another patent, or in turn was not cited by another patent, it likely did not contribute to the technological ecosystem in any meaningful way, when considering the scope of the project being on interconnected technology, as per the definition of GPTs.
Secondly, to generate a network graph containing hundreds of thousands of nodes with citation edges between them, would result in not only inconsistent graph construction, but many of these patent nodes would be isolated from the rest of the system anyhow, thus not contributing to the network analysis and possibly diluting the results.

---

[2] In less technical terms, without the use of the index for each table, each query performed (in this case a little over 4480 per dataset), could take upwards of 30 minutes

6.2 Network Construction and Analysis

Objective: Construct the patent network graph and perform analytical procedures to extract insights.
This involves setting each patent as a node, and the edges as the citations between them. The rest of the variables will be attributes connected to each patent, which collectively will enable analysis of the patent. Once a patent has been identified as a possible driver of the General Purpose Technology, this will allow for more in-depth analysis of technological innovation pathways.

- **Node Representation:** Each patent in the dataset is represented as a node in the network graph. Nodes are the fundamental elements that depict individual patents.
- **Edge Representation:** Edges between the nodes represent citations among patents. An edge from node A to node B indicates that patent A cites patent B. This directional aspect of the edges helps in understanding the influence of one patent on another.
- **Attribute Association:** Each node is enriched with attributes derived from the patent data, such as application ID, inventor details, publication date and other bibliographic information. These attributes facilitate multifaceted analysis, such as technological focus, and industrial, geographic and temporal distributions of patent activities.

Centrality Measures: To identify key patents and influential nodes within the patent network, I will calculate a centrality metric. The chosen metric in this case will be eigenvector centrality, which provides a significant perspective on the network's structure and the importance of each individual patent in relation to others. By analyzing this centrality measures, the aim is to uncover patents that are pivotal in the development and diffusion of technology, which are potential drivers for AI as a General Purpose Technology

Community Detection: To explore the structure of the patent network further, community detection algorithms will be applied, specifically the Louvain method for modularity optimization. This algorithm is chosen for its efficiency and ability to uncover hierarchical community structures within large networks. The Louvain method optimizes the modularity of the network, grouping patents into clusters based on the density of connections within these specific clusters compared to connections with the rest of the network. Identifying these clusters will allow us to detect groups of closely related patents, which can provide insights into emerging technological trends and possibly some of the individual drivers of this type of general-purpose technology.
By identifying these communities, it becomes easier to better understand how different technologies are related and how they evolve over time. Clusters of patents that show strong interconnections may indicate emerging fields or underlying technologies with broad applications.

Descriptive statistics of the network: To provide a comprehensive overview of the patent network, various descriptive statistics will be extracted and analyzed. These statistics will help create an understanding of the basic properties and structure of the network, as well as confirming AI to fit with the characteristics of General Purpose Technology. Measuring the total number of nodes will give us the overall size of the network, indicating the number of patents included in our analysis. Analyzing the distribution of patents by their publication years and the countries of origin will provide insights into temporal trends and geographical patterns in technological development. This information can highlight periods and regions of intense innovation. Finally, a summary will be presented of the patents with the highest centrality measures. These patents are considered the most influential and are likely candidates for GPT drivers due to their pivotal role in the network.

In summary, by examining these descriptive statistics, the aim is to attempt to contextualize the findings further and provide a foundation for identifying patents that have the potential to be considered the driving forces of general-purpose technology. This initial overview will set the stage for more detailed analysis and discussions in subsequent sections.

6.3 Integration of Large Language Models

The objective of integrating Large Language Models (LLMs) into our patent analysis is to enhance the assessment of central patents identified through network analysis. By leveraging advanced Natural Language Processing (NLP) techniques, I aim to gain deeper insights that can help determine the potential impact of these patents as the driver for AI and ML as general-purpose technologies.

To perform this advanced semantic analysis, I will utilize pre-trained Large Language Models, in this case through API calls to the gpt-3.5-turbo-0125 model from OpenAI. This model was chosen for their robust capabilities in understanding complex language, extracting meaningful information, and providing semantic interpretations of textual data.

As mentioned in earlier sections, in this project, I will also implement the Retrieval-Augmented Generation (RAG) approach to process and analyze patent data as a proof of concept. RAG combines the strengths of retrieval-based methods and generation-based methods to provide accurate and contextually relevant answers. This section details the steps and tools involved in integrating and utilizing the RAG approach for our patent analysis.

To start, patent data is loaded into a dataframe, focusing on the patent abstracts and their associated metadata, such as filing dates, application ID, inventors and titles. The meta data will be used to enrich the abstracts, before they are then to be processed into embeddings. Additionally, we will further enrich the abstracts using an LLM to generate specific keywords for each abstract. These will enable a retrieval system to fetch the most relevant patents based on semantic similarity, which can then be provided to a large language model as contextual information for answering queries.

Embeddings are a crucial component of the RAG approach, as they transform text data into numerical representations that can be efficiently processed. A pre-trained transformer model is used, specifically all-mpnet-base-v2 from the sentence transformers library [RG20], to generate embeddings for the patent abstracts. These embeddings capture the semantic meaning of the text, allowing for more effective retrieval and generation processes. These embeddings will then be entered into a vector index known as FAISS, which essentially functions as a database for these embedding vectors.

FAISS (Facebook AI Similarity Search) is a library developed by Facebook AI Research to enable efficient similarity search and clustering of dense vectors. By creating a FAISS index, the relevant documents based on the generated embeddings can quickly be retrieved. This index allows us to perform similarity searches across patents efficiently, ensuring that the most relevant patent abstracts are retrieved for any given query [DGD+24]. This is what enables the utilization of a RAG-based approach for the LLM implementation. RAG leverages both retrieval and generation.

I utilize semantic similarity scores through the earlier mentioned embedding vectors to generate a response based on the retrieved context and a prompt engineered to guide the answers. However, the number of patents retrieved will be limited to a max of five, as this insures that the context window limit within the selected model is not exceeded. Going beyond that limit might exclude otherwise relevant information, resulting in less accurate responses. Utilizing a model with a higher context window, might prove capable of containing higher numbers of patents without information loss[JHY+24], but due the testing being performed through API calls, this method will suffice. Collectively, this approach should enhance the model's ability to provide accurate and contextually relevant answers, which provided that it proves capable of responding in a relevant manner, should make it a powerful tool for patent analysis.

The steps of this entire process can be summarized here

- Enrich patent abstracts: The abstracts will be enriched with further metadata, enabling even more semantic links to be made to the user query. This should ideally allow the model to answer questions about specific patented technology, including the inventor and time of publication.
- Generate text embeddings: The enriched patent abstracts need be converted into embeddings, which are essentially vector representations of text. This is used for the next step
- Retrieving Relevant Documents: Using the FAISS index, the RagRetriever model retrieves multiple documents that are semantically similar to the query.
- Generating a response: The retrieved information is then used to generate a comprehensive and contextually relevant answer to the query.

Semantic Analysis: Furthermore, the LLM will be employed to perform specific labeling of a patents' content and context. The model will identify and extract key AI techniques within the field from the patent documents. This will help in pinpointing the core technological concepts and innovations described in each patent. This can highlight potential overlaps and collaborations between different technologies, further aiding in the identification of which specific type(s) of technology that enable AI and ML as a GPT.

Additionally, I will utilize the calculated centrality measures from the network analysis as a threshold filter in order to reduce the total amount of token usage on the API calls. This still results in a sample containing about 300 patent abstracts that will be labeled according to pre-defined categories of technology within AI and ML. Furthermore, I will also have the LLM label the patents based on the industry they belong to. The pre-defined categories will be determined through a exploratory data analysis of companies that hold the most patents, in addition to the patents geographical distributions that can reveal which main countries and thereby industries that utilize AI.

Combination with Network Analysis: The combination of centrality metrics and semantic analysis allows us to not only identify key patents but also understand their contextual relevance. This dual approach ensures that patents identified as central indicate their strength as drivers of this particular GPT. The insights gained from the LLMs will be used to conduct qualitative assessments of the most central patents. By examining the the LLM responses and semantic similarities between queries and retrieved patents, I can better gauge each patent's potential impact, which is a critical factors in identifying driver candidates.

The integration of Large Language Models with traditional network analysis techniques provides a powerful approach to patent analysis. By combining quantitative centrality measures with qualitative semantic insights, I should ideally be achieve a more comprehensive understanding of the patent landscape. This enhanced methodology will not only be able to determine whether AI can be considered a GPT, but also identify key patents, and subsequently evaluate their technological cores to determine whether they are central to this general-purpose technology, ultimately contributing to a more informed and nuanced assessment of technological innovation within this field.

6.4 Visualization and User Interface

I will attempt to develop a dashboard to visualize the patent network and provide user-friendly access to complex data insights on the technology landscape in question. This aims to bridge the gap between sophisticated data analysis techniques and their practical application for users who may not have extensive technical expertise.

The framework library Streamlit forms the core of this initiative. This dashboard should allow users to interact with the RAG system seamlessly, exploring patent data and retrieving relevant information through an intuitive interface. By incorporating interactive features, users can input queries and receive real-time responses from the RAG system, making the interaction smooth and informative.

The dashboard will also include visual tools to represent the distribution of patents across different technological fields, geographical regions, and time periods, helping users quickly understand key trends and insights. To make sure that the interface stays responsive and efficient, it will also have optimized backend processes to ensure the system can handle complex queries and data retrieval operations efficiently. By developing a user-friendly and efficient dashboard, the project aims to make the complex insights derived from patent analysis accessible to a broader audience. This not only enhances the practical application of the research but also provides a valuable tool for stakeholders in the field of AI and patent analysis.

6.5 Testing and Deployment

It will be necessary to ensure the reliability and effectiveness of the RAG system before and after deployment.

Integration Testing: The first step involves thorough integration testing to verify that all components of the RAG system work together seamlessly. This includes testing the interaction between the data retrieval mechanisms, the RAG model, and the user interface. Integration testing will identify and address any compatibility issues or data flow interruptions, ensuring that the system operates as a cohesive unit.

Deployment: The deployment strategy will be designed for scalability and accessibility. The system will initially be deployed as a proof of concept (PoC) using the university's cloud resource service, UCloud. UCloud provides substantial computational resources similar to those of commercial platforms like AWS or Azure but without the associated costs. This will facilitate the deployment of the Streamlit-based dashboard and the handling of API calls for LLM interactions, ensuring the system can operate efficiently with available resources

By following these steps, the technical implementation of this project aims to provide robust support for analyzing technological developments using patent data. The combination of thorough testing, strategic deployment, and ongoing maintenance will ensure the system delivers valuable insights into the dynamic field of AI and machine learning.

7 Patent Analysis

I will begin the patent analysis with a simple descriptive analysis of some of the key variables in the available datasets. This includes identifying the distributions of inventors, countries and patent filing years to aid the analysis of the network, as well as determine whether or not AI and ML can be considered a General Purpose Technology. Once this is done, I will move on to analyzing the network itself.

7.1 Variable Distributions

First the distributions of key variables within the data will be visualized and analyzed, as this will provide an initial insight into the field. Patent filing years will indicate innovation activity across the temporal dimension, thus indicating the characteristic of undergoing continuous development. The entities within the data that have the highest counts of patents will, through qualitative assessment, provide an indication of the characteristics of implementation rates across industries. Finally, the countries that have the highest rates of implementation will be explored to identify geographical innovation hot spots within the field, and subsequently undergo a short evaluation to identify possible reasons for their patent activities.

7.1.1 Patent filing years

Let us first explore the distribution of patent filing years. As one of the characteristics of General Purpose Technology is to undergo continuous development, this will help us to better understand the temporal dimension of innovation in this field, as this can help indicate the adoption rate of AI and ML in industry. Additionally, I will attempt to supplement the analysis with further sources to confirm the observations from the distribution.

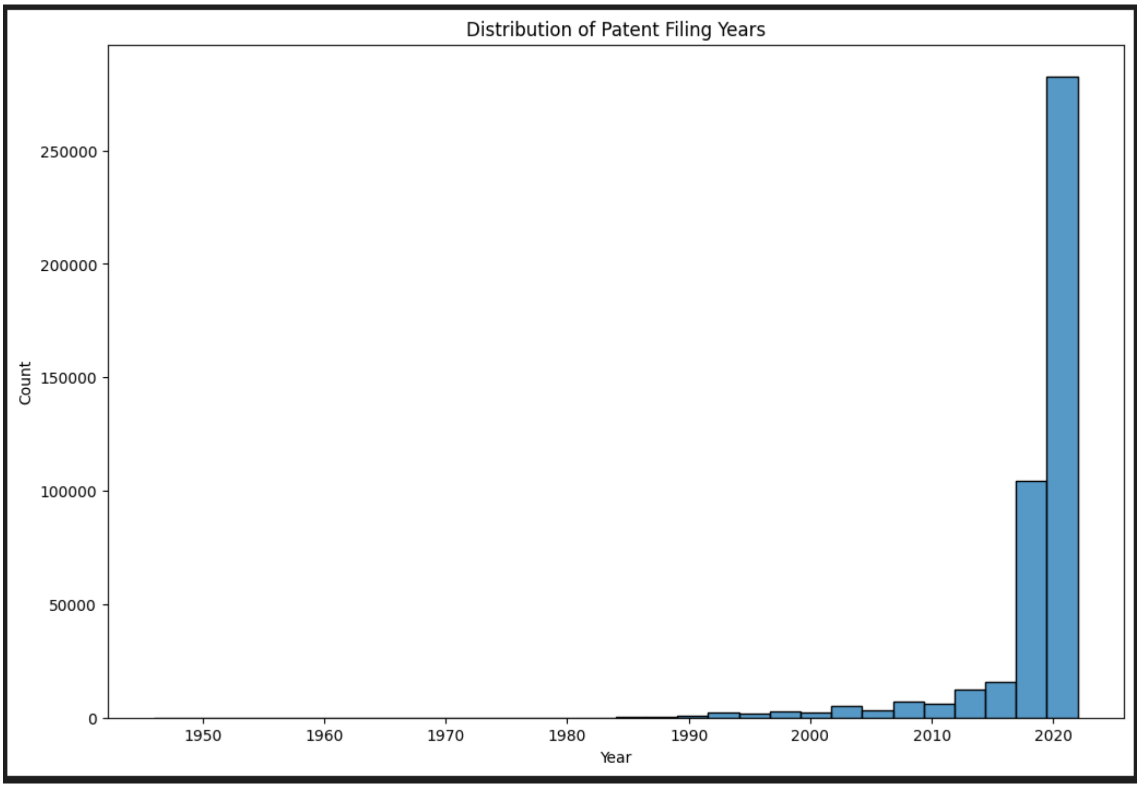

Figure 1: Distribution of Patent Filing Years

Figure 1 illustrates the distribution of patent filing years for our targeted field. It reveals a few key trends and observations. The most striking feature of this distribution is the exponential increase in the number of patents filed in recent years, particularly post-2010. A report from WIPO on technology trends, suggest that this surge is likely a result from rapid acceleration in research and development activities in AI and ML technologies, driven by advancements in computational power, availability of large datasets, and improvements in algorithms [Ref].

Prior to 2000, the number of patent filings in AI and ML was relatively low. During this time, research was seemingly confined to academic explorations, with limited practical applications and commercial interest, which is the reason there is a lower occurence of patent filings [HKMSZ24]. However, the years between 2000-2010 indicate an initial transition as it can be noticed that there is a gradual increase in filed patents. This is most likely due to an growing interest in the field by tech firms, as the rate of innovation increases and more practical applications of ML and AI are found for commercial purposes [Gue24].

The most significant growth occurs after 2010, with a large increase in patent filings. By this time, AI and ML technologies had matured significantly, leading to more robust and scalable applications. There was widespread adoption of AI and ML across various industries such as healthcare, finance, transportation, and entertainment, driving innovation and patent activity, as both public and private investments in AI and ML rapidly increased, furthering the research efforts in the field (kilde).

7.1.2 Patent holders

Next, I will explore some of the key companies and industries that exists within our datasets, as this provides an additional indication of AI technology being a general purpose technology. For this purpose, the distribution of the 15 companies that hold the highest number of patents within this field will be examined visually, and then the goal will be to subsequently explore which type of industry these belong to.

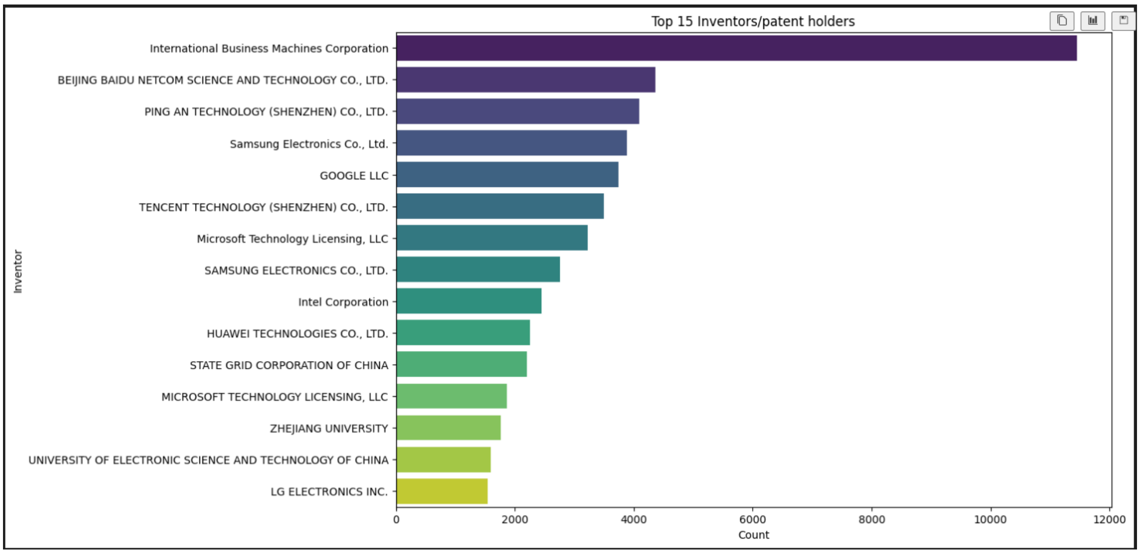

Figure 2: The top 15 patent holders in the field

From figure 2, I can deduct a few trends as far as patenting behavior goes. In total in the inventor dataset, there is 908.206 individual patents, meaning that the highest percentage of patents held is International Business Machines Corporation (IBM), which around 1 percent of the total amount of patents. This indicates that the technology has a relatively high spread across several firms, as each of the rest hold less than one percent of the technology patents.

If I look into each individual company I observe on figure 2, I also see a relatively diverse representation of markets for both services and products.

IBM is a global technology and consulting company, that is known for its contributions to AI, particularly through the Watson AI platform. This enables individuals to more easily perform AI-related work such as prompt engineering, Machine Learning Operations (MLOps), fine-tuning as well as hosting foundation models for these purposes [Mac24]. Evidently from the analysis, it has been a major player in terms of patent filings for AI systems, which also speaks to their influence on industry standards and technological advancement.

Beijing Baidu Netcom Science and Technology is a leading internet product and service provider in China, and is sometimes referred to as the "Google of China". They utilize AI to power their search engines, smart devices and other various internet services. Whilst not holding as many patents as IBM, based on the utilized filtering criteria as presented earlier, they still remain a major player in the field. Their relatively high ranking indicate a robust research effort, which drives the companies competitive edge in this highly dynamic market [Bai24].

Ping AN Technology is a subsidiary of Ping AN Insurance, and they utilize AI for fintech, healthcare as well as other initiatives that enable a higher efficiency rate as well as improving the customer experiences in terms of insurance and finance [Gro24]. This indicates yet another industry in which AI has taken hold, as they shift towards a tech-driven future.

Samsung Electronics is a well-known tech giant within the field of consumer electronics and information technology, that produces smart phones, TVs and other devices. They utilize AI to enhance their users experiences through intelligent features in their products, which helps maintain their strategic edge on the market [Sam24].

Google is the company known for running the well-known search engine, as well as various other internet-based services and products. By some, they are considered a pioneer in AI research, having made significant advancements within the fields of machine learning, natural language processing (NLP), and computer vision [NH20]. AI technology lies at the heart of the search algorithms they use, which enables them to deliver highly relevant search results and improve user experience. Additionally, Google's AI innovations power its voice-activated assistant, Google Assistant, and enhance the capabilities of its cloud platform, Google Cloud AI [Goo24].

Tencent Technology is one of the key players in the markets of social media and gaming. The AI research they perform, focuses on increasing the user engagements on the companies owned platforms, as well as advancing the utilization of AI in gaming. Tencent leverages AI to personalize user experiences, enhance content recommendation algorithms, and improve real-time communication features. In gaming, Tencent employs AI to create more immersive and interactive environments, utilizing machine learning to develop smarter NPCs (non-player characters) and to enhance game mechanics and player matchmaking [Ten24].

Microsoft Corporation is one of the worlds leading technology companies. Microsoft is renowned for its wide range of software products, including the Windows operating system, Office productivity suite, and Azure cloud computing services. Microsoft has positioned itself as a leader in AI through extensive research and development in machine learning, natural language processing (NLP), and computer vision. AI is deeply integrated into many of Microsoft's products and services. Azure AI provides cloud-based AI services that empower businesses to build intelligent applications. Microsoft Office 365 utilizes AI to enhance productivity features, such as real-time language translation in Word and smart email sorting in Outlook. Additionally, Microsoft's AI research has led to the development of Cortana, a virtual assistant that helps users manage tasks through voice commands [Mic24].

Intel Corporation is a global leader in semiconductor manufacturing and technology. Intel has been at the forefront of developing AI hardware and software solutions, focusing on processors optimized for AI workloads and AI accelerators. Their AI technologies support data centers, edge devices, and autonomous systems. Intel's AI products, such as the Intel Nervana Neural Network Processor are designed to accelerate machine learning and AI applications specifically. Intel also invests in AI research through collaborations with academic institutions and industry partners, aiming to advance AI capabilities and develop innovative solutions. The company is actively involved in various AI applications, including autonomous driving, healthcare, and smart cities, leveraging its robust technological infrastructure to drive advancements [Int24].

Huawei Technologies is a global leader in telecommunications equipment and consumer electronics. Huawei invests heavily in AI for telecommunications infrastructure, consumer electronics, and cloud services. Their AI research focuses on network optimization, smart devices, and IoT. Huawei's AI capabilities are integrated into its smartphones, smart home devices, and enterprise solutions, enhancing performance and user experience [Ent24].

The State Grid Corporation of China is one of the largest utility companies in the world. State Grid uses AI to manage power grids, improve efficiency, and integrate renewable energy sources. Their AI solutions help optimize energy distribution, enhance grid stability, and reduce operational costs. State Grid's AI initiatives also focus on predictive maintenance and fault detection, ensuring reliable and efficient power deliver[oESoC23]y [Rik22].

The University of Electronic Science and Technology of China (UESTC) specializes in electronic engineering and computer science. UESTC significantly contributes to AI research in areas such as signal processing, telecommunications, and automation. The university's research focuses on developing advanced AI algorithms and applications, aiming to solve complex problems and drive technological innovation [oESoC23].

LG Electronics is a South Korean multinational corporation that invests in AI for smart home appliances, mobile devices, and other consumer electronics. LG's AI research focuses on enhancing product functionality and user experience through intelligent features. The company's AI research spans across several key areas, such as warehouse optimization, demand and sales forecasting and several more [Res24].

The dominance of technology companies and research institutions in this graph further indicate the role of AI and ML across various sectors, being significant indicators of its position as a General Purpose Technology. These organizations are leading the way in developing and deploying AI technologies, reflecting the strategic importance of AI and ML in driving innovation, efficiency, and competitive advantage. The significant investment in AI research by both private companies and academic institutions highlights the interdisciplinary nature of AI advancements and their broad applicability across different industries.

7.1.3 Geographical distribution

Next, a short overview of the geographical hotspots will be analyzed. The following figure below shows the distribution of patent activity:

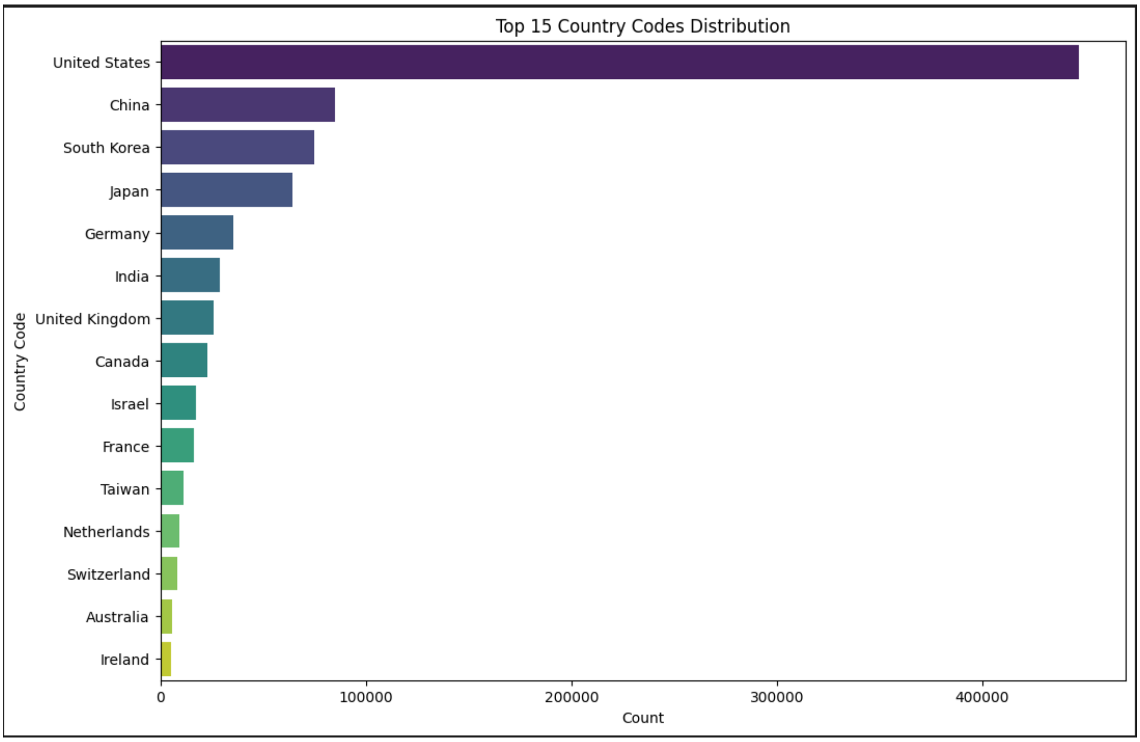

Figure 3: Top 15 Country Nodes Distribution

Figure 3 shows that The United States (US) leads significantly, followed by China (CN), South Korea (KR), and Japan (JP). This distribution likely reflects a combination of historical, economic, and policy-driven factors that influence the innovation landscape globally. As done with the distribution of companies, we will explore some of the underlying factors that contribute to this visualization.

The United States shows a highly dominant position within the field, which can likely be attributed to a few key factors. The first of which is a historically strong innovation ecosystem, which is a result of heavy investments in research and development of new technologies [Gri90]. In addition, the US has some of well-established IP-protection laws that encourage an increased rate of patent filings, as the intellectual property of inventors are better protected (kilde). Most importantly, the US also contain several global technology leaders, such as IBM, Google and Microsoft, all of which have been shown to output a significant amount of patents as some of the top 15 companies in the earlier distribution analysis. Finally, heavy funding from both government and military have historically driven the innovation rates for various defense purposes significantly in areas such as AI, ML and cyber security [JTH93].

China coming in at second place on patent activity is noteworthy, as it indicates a strategic focus on technology and innovation. Govermental policies likely lies at the heart of this development, as China has implemented policies that encourage innovation through tax incentives as well as direct funding for research and development [BM14]. Therefore, investments into higher education and research institutions have likely had a significant impact on the capabilities of China to innovate. Additionally, an geographically aggressive IP protection strategy have been employed by several chinese tech companies, such as Tencent and Alibaba [Li09].

South Korea holds a strong position in AI and ML patent filings is primarily driven by its technological advancements and the influence of major conglomerates like Samsung, LG, and Hyundai. These companies invest heavily in AI and ML research and development, leading to a high number of patent filings [EK02]. The South Korean government also provides substantial support through grants and research incentives [BM14]. South Korea's global leadership in electronics and semiconductor industries, sectors that are highly patent-intensive in AI and ML, further enhances its patent behaviour [FV02].

Japan likewise, has shown a significant commitment to the research of AI and ML technology, which is shown by a relatively large number of patent filings. The country has a strong industrial base in fields such as electronics, automotive and robotics, which each generate a significant amount of patents within AI and ML [Sco04]. In addition, a culture of continuous improvement, drives collaboration between the government and companies in research activities, further boosting the number of patent filings [FV02].

Germany has an innovation landscape within AI and ML that is primarily formed by a strong industrial focus, as they are known for their excellent engineering capabilities within the automotive, machinery and chemical industries. Once again, in a combination with a collaborative effort between institutions and the government through funding and incentives, this places Germany as one of the highest patent counts within the data [BM14].

India has seemingly has had a growing role in the field of AI and ML due to the efforts that have been made into the pharmaceutical and IT industries. They remain a major player in these particular industries [Sco04], which have seemingly lead to a significant number of AI-related patent filings . In addition, the increase AI research efforts can also be attributed to govermental initiatives such as "Make in India", which helps foster innovation and in turn increase patenting behaviour [BM14].

The United Kingdom has a strong research base and historical innovation culture that has contributed to its AI and ML patent filings. Leading universities such as Oxford and Cambridge are hubs of AI and ML innovation. The UK is also a leader in financial services technology, also known as fintech as well as biotechnology, both of which are sectors that increasingly integrate AI and ML. Like the other countries, the government utilizes programs and incentives for research and development, which in turn increase the number of patent filings in these areas [EK02].

Canada much like the other countries, has an innovation ecosystem that benefits from strong collaboration between academic institutions and related industries. This is further fueled by a government support system that increase the research incentives in fields such as AI and ML. Collaboration between universities and industry drives innovation, supported by governmental programs like tax incentives [BM14].

Israel likewise contains high-tech industries and a startup ecosystem, which also makes it one of the leaders in AI and ML innovation. Israels ecosystem, particularly in AI and ML technologies, leads to high patent activity. Innovations from military research and development often translate into civilian patents, enhancing Israel's position in AI and ML [FV02].

France also has a strong industrial base supported by governmental incentives, which is reflected in a likewise relatively high number of patents within the field. Their significant efforts into aerospace and defense technology, combined with govermental initiatives is what drives innovation in this country, thus resulting in a high level of patent activity [EK02].

Taiwan likewise can attribute much of its patent filing activities to governmental incentives, as well as a few particularly strong industries, such as the production of semiconductors. Much of the knowledge creation reflected in its patents, come from government funded incentives for research and development [BM14].

The Netherlands also benefits from a strong research infrastructure, as well as the presence of global companies that utilize AI and ML technology. Research institutes, universities and companies like Philips and Shell, contribute significantly to the patent filing activity in the country [MR99].

Switzerland has a strong pharmaceutical and biotech industry, that drives the number of patents in this country. Companies such as Novartis and Roche are significant contributors, and in addition this innovation is further boosted by strong IP policies as well as innovation ecosystem [EK02].

Australia similarly benefits from a strong innovation ecosystem, fueled by strong collaboration networks between academia and industry related to AI and ML. Also here, governmental programs and incentives provide support for these activities, which also increases the innovation rate and subsequent patent activity [FV02].

Finally, **Ireland** has a rather unique role in this picture, as it serves as a hub for multinational companies such as Facebook, Google and Apple, which inherently increases the number of patent filings in this country. This is in particular incentivized through strong tax policies that serves as an attraction to major companies and research centers [FV02].

In summary, it appears that innovation and patent activity is fueled by several significant factors outside of the technology itself. These include major collaboration networks between academic institutions, governments and industry. Academia seemingly plays the role of producing the knowledge needed for innovation, whilst industry takes the role of finding usable applications for this innovation, as well as investing through collaborative efforts into research and development of new technology. Finally governments primarily play the role of incentivizing the innovation, thus increasing the rate of innovation across industries. This however comes as no surprise, as these insights have been generated for decades through academic papers that explored innovation factors. What can however be deducted from this perspective, is once again the rate of adoption of AI and ML technology across not only technological sectors, but likewise regional adaption of this particular type of technology.

7.2 Network analysis

In this section I will pivot into the network analysis in order to identify the most central patents through a backwards citation network, the eigenvector centrality measure and subsequent community detection to identify the most central patents for further analysis..

7.2.1 Network centralities

As mentioned in the methodology, the objective of examining the networks eigenvector centralities is primarily to identify the most central patents within technological clusters of AI and ML. Given that the network is so large and contains so many nodes however, the majority of eigenvector centrality measures take on incredibly low values. In order to visualize and analyze these, they must be preprocessed for a more intuitive interpretation. In this case I utilized normalization and log transformations to make the data more interpretable, as the values will be transformed into numbers between 0 and 1. After this, the following boxplot of the distribution could be made

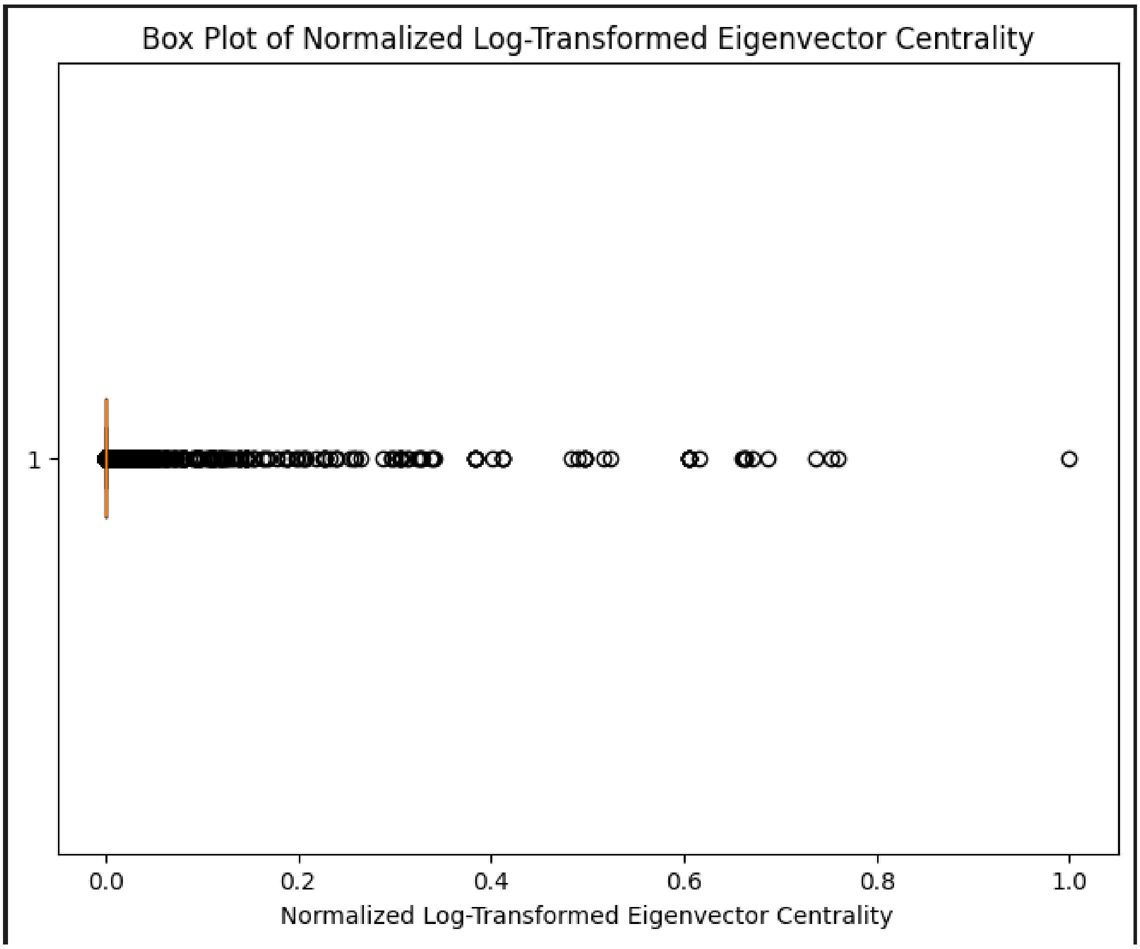

Figure 4: Boxplot of eigenvector centrality distribution

As already mentioned, the figure shows a significant number of low score centralities, which indicates that there a large number of patents that don't connect to the more well-connected ones. This in turn also seems to indicate that a select few of these patents are particularly well-connected to the rest of the network, and thus may function some sort of driving force for the field as a whole.

If we instead try to categorize these values into bins, the picture becomes even clearer. The following figure below shows the distribution across a few categories based on centrality scores. The low category is patents with centralities equal to 0.2 or less, medium is between 0.2 and 0.5, high is between 0.5 and 0.8 and finally very high is the remainder above 0.8. The graph clearly shows a very high concentration of these values lower than 0.2, which is consistent with the figure from before.

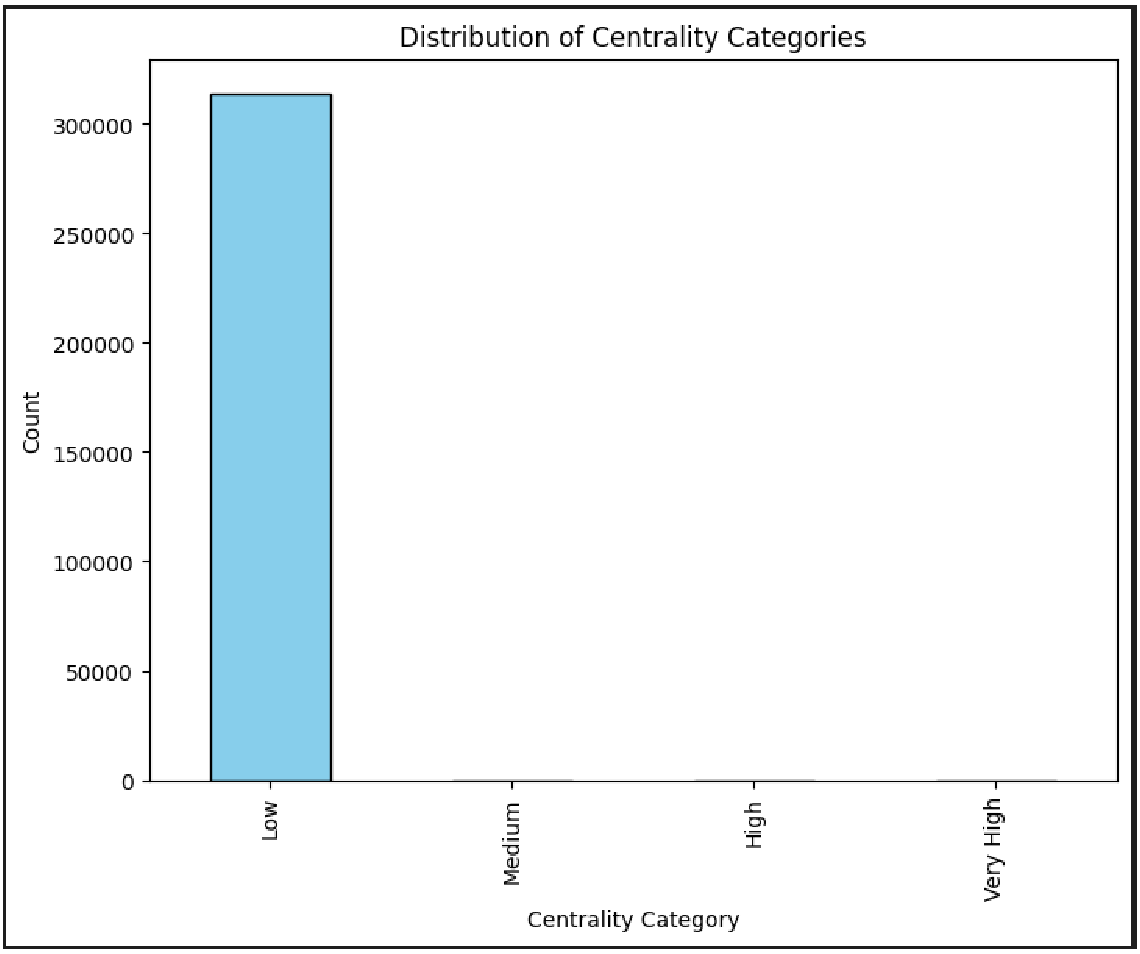

Figure 5: Centrality measures categorized by size

However, due to the extremely large amount of low score patents, it is not feasible to see that there are in fact several patent within the remaining categories. The number of patents within each category are instead listed as follows:

- Low: 314041
- Medium: 48
- High: 18
- Very High: 1

This indicates that we have identified central patents within the network based on the eigenvector centrality, that could prove valuable for testing the proposed framework. The next step will be to extract these particular patent abstracts, and analyze them with the LLM.

7.2.2 Community Detection

This section will go over the analysis and results of the community detection approach, and try to determine whether this provided any meaningful insight in terms of the technological field.

The Louvain community detection algorithm was applied to our patent network to identify clusters of related technologies. Initially, the algorithm detected over 3,000 communities, many of which contained fewer than 10 patents. This unexpected result suggested that the algorithm was identifying too many small and weakly connected communities. Whilst the size of the total network meant that it was not possible, nor really meaningful to visually interpret all of these communities, it was decided to utilize primarily descriptive statistics in an attempt to interpret these clusters.

The distribution of these initial communities can be viewed in the figure below:

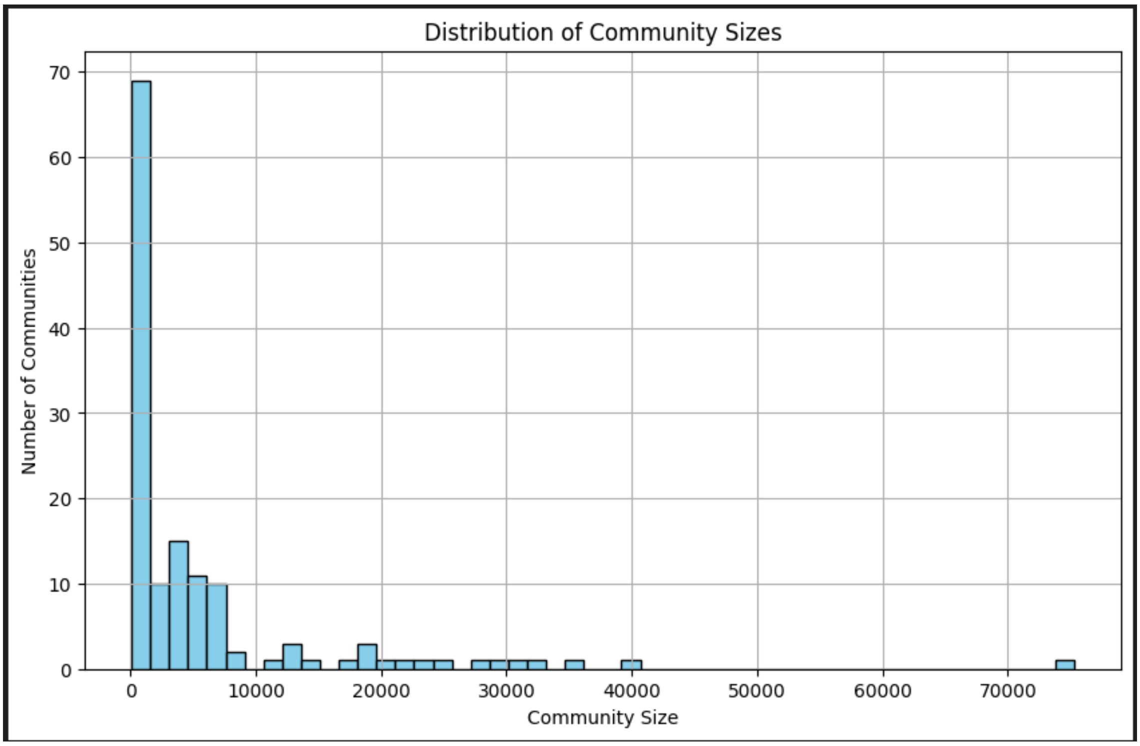

Figure 6: Distribution of the detected community Sizes

To address this, the resolution parameter of the Louvain algorithm was adjusted to 1.5. The resolution parameter influences the granularity of the detected communities; increasing it leads to fewer, larger communities. By experimenting with different resolution values, it was found that a resolution of 1.5 provided a more reasonable partitioning of the network, resulting in a more balanced distribution of community sizes.

This however still resulted in communities that appeared to be significantly fragmented, thus not lending itself to further analysis. Therefore an additional post-processing step was utilized to merge small communities into larger ones. Specifically, any community with fewer than 50 thousand nodes was merged into its neighboring communities based on the most common community among its neighbors. This however still left several communities that only contained a single node, which once again suggests a highly fragmented networked. Therefore, for the sake of analysis, these communities have been excluded from the analysis, as they would not provide any meaningful insight. This final approach helped to ensure that each community had a sufficient number of patents to be meaningful for analysis.

After these adjustments, the community sizes became more balanced, with fewer very small communities. The adjusted partitioning allowed us to identify key patents within each community based on eigenvector centrality, analyze in- and outgoing community edges, and calculate the modularity of the network graph partition.

When analyzing the modularity of the network, which is a measure of the strength of division of the network into individual communities, we derived a modularity score of 0.8285. This score indicates that there is a strong community structure within the network graph, which suggest that there are well-defined themes through the network communities. This partially explains the fragmented nature of the network, as strong thematic clusters are not likely to well-connected to each other as less modular ones.

To support this observation, it was chosen to also calculate the occurence of inter- and intra community edges. Inter-community edges represent connections between nodes across the different communities, and intra-community edges represent the connections between nodes within the same community. The measurements of these community edges, as listed below, indicated the same as the modularity score.

- Inter-community edges: 262,773
- Intra-community edges: 1,674,106

The fact that there are significantly more intra-community edges than inter-community edges is what makes these measurements consistent with the high modularity score. The much higher number of intra-community edges indicate that most connection within the network happen between nodes that exist within the same community.

Finally an attempt was made to identify the main themes within these many different clusters, but due to the otherwise fragmented structure of the network graph, as well as the immense size of this network, the attempt ultimately failed due to processing requirements for such a large network.

Instead it was chosen to focus on the largest community clusters and attempt to identify the 5 most central patents within each. These patents, along with those detected through network centrality, which resulted in a rough total of around 300 patents, are what will be analyzed using the LLM implementation later on in the project.

7.3 Total summary of analysis results

In summary, the network analysis showed that a comprehensive network, whilst possible to analyze, needs significant resources and consideration to properly interpret. Whilst the analysis could have produced significantly better results under a more careful data selection process, it was still possible to derive a few key insights that allows the subsequent analysis to take place.

Firstly, the chosen variables have indicated that AI and ML can indeed be considered a General Purpose Technology, due to their adoptation across industries as explored in section 7.1.2. These indicated that AI is used in a large variety of ways, spanning from healthcare and financial technology, into the realm of gaming and the automotive industry, which aligns with one of the most significant characterizations presented in the introduction.

It was also determined that there are clear indicators for a continuous rate of innovation as explored in section 7.1.1, which also cements AI and ML as a General Purpose Technology.

Through a qualitative assessment of the most common countries within the patent landsdcape in section 7.1.3, we also briefly found some of the key influences on technological innovation within these countries, which can help fuel the discussion of the results further on in section 9.

Finally, the analysis of the network proved not to be as fruitful as expected. Whilst it was possible to determine some indication of key patents based on eigenvectors and community clusters, the results may have been diluted or flawed, due to a fragmented network. This is seemingly due to a technical flaw in the data handling strategy, which resulted in repetitive patents occurring in the dataset. This is far less than ideal, but still provides the ability to test the remaining technical implementation. The implication of this will be explored further in a later section, where I will discuss the results in a more holistic manner.

8 Implementing the Large Language Model

The following section will focus on the implementation of the specific steps required for processing the patent data through a Large Language Model. In order to process the patent abstracts with the chosen LLM, I will need to extract these and inject and use this text with RAG. As mentioned earlier, I can inject new contextual information into the text prompt, in this case being the patent abstracts and combine them with additional instructions to perform the analysis. In order to reduce the computational resources required for the LLM, 300 patents was randomly picked from the central patents identified through community detection and eigenvector analysis.

The type of analysis we wish to perform can be determined by the provided instructions via prompt engineering, which in this case will be to determine which specific AI or ML-based technology that lies at the center of a given patented invention. I will use this method to essentially label each patent with the type of AI or ML technology it believes it to be, and then subsequently add these labels to a dataframe for a more static data analysis. In addition, I will test a embeddings-based RAG implementation of a question-answering system on a few general questions that relate to the technological field, and see if the model can provide comprehensive answer to these.

8.1 Categorizing patent technologies using LLMs

First I will attempt to have the LLM categorize each patent into a specific type of AI technique, in order to try and map which techniques are the most occurring within the selected sample. The reason I utilize a smaller sample as opposed to all the identified central patents, is simply due to API restrictions, as I rely on OpenAIs API to run the LLM calls. Instead this will serve as an experiment and illustration of an LLMs ability to process through and generalize concepts within technical documentation, which is essential for understanding the patent contents in detail.

The defined technique categories will be as follows:

- Machine Learning
- Deep Learning
- Natural Language Processing
- Large Language Model
- Reinforcement Learning
- Computer Vision
- Quantum Computing
- Others

These categories cover some of the most central techniques of AI and ML based systems either directly or indirectly [Hul20], and thus should provide the model with a reasonable selection of categorical labels to choose from. It is however worth noting that many areas of AI research contain significant overlaps between subjects, which may result in broken classifications. For example, Deep Learning and Large Language Models (LLMs) could be considered to have a large overlap, as both Deep Learning models and LLMs consists of neural network structures. Some applications however, only utilize one of the two, and for the sake of testing the reasoning power of the chosen LLM, gpt-3.5-turbo-0125, these categories will suffice for initial evaluation. In addition, Quantum computing which by some is considered the next step in terms of computer processing, has been undergoing heavy research within the field of AI by companies such as Google with their Quantum AI platform [FND$^+$20]. Therefore this example has been added The results of this attempted classification can be seen in the figure below:

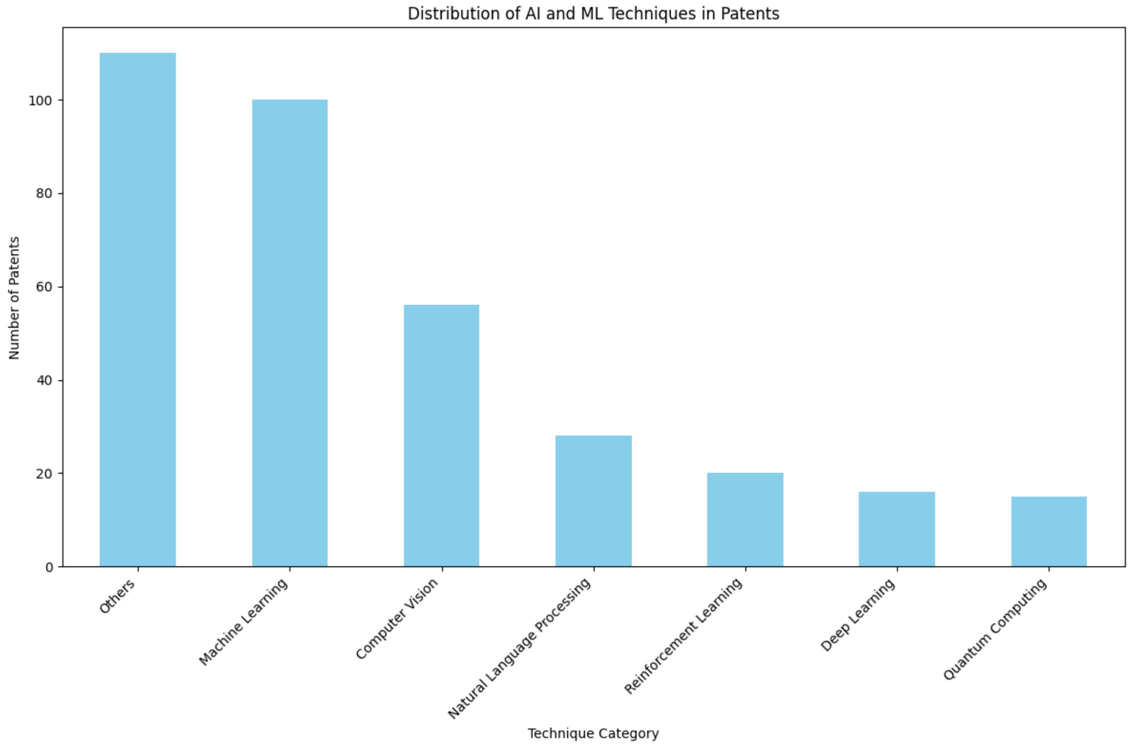

Figure 7: Distribution of AI techniques in the sample

The figure above illustrates, that whilst the model can indeed identify certain key themes, particularly within the general field of Machine Learning, it also fails to identify more than a third of the total patent sample amount, and instead labels these as "Others". This could be due to one of two reasons. First, if the categories are too restrictive, which is likely the case here, and doesn't encompass a wide enough variety of techniques, the model is more likely to group patents that would have otherwise fit into these missing categories as "Others". The second reason could be that the model in these cases cannot determine the technology core techniques through the abstracts directly, which would likewise result in labeling a patent abstract as "Others". However, given that the model could correctly label most of the patent abstracts, it is more likely that the categories are simply too restrictive.

Next I will attempt to categorize the patents into their respective industrial fields that we have determined from the previous distribution analysis. This should indicate whether or not the model can determine the main use cases of different techniques, and categorize these into a specific type of industries. The following categories was used for the second attempt:

- Healthcare AI
- Cybersecurity AI
- Robotics
- Consumer Electronics
- Autonomous Vehicles/Transportation AI
- Financial Services/Fintech AI
- Pharmaceuticals and Biotech
- Aerospace and Defense
- Industrial and Manufacturing AI
- Energy and Utilities
- Retail and E-commerce

- Agriculture and Food
- Telecommunications
- Others

These categories encompass some of the main applications used by companies that have observed and analyzed in the patent data distributions in section 7.1. This should indicate whether or not the model through its inherent reasoning power can identify the application area of the patents based on the abstracts alone, which should prove particularly useful for keyword generation later on when I will attempt to enrich the abstracts further for the RAG implementation.

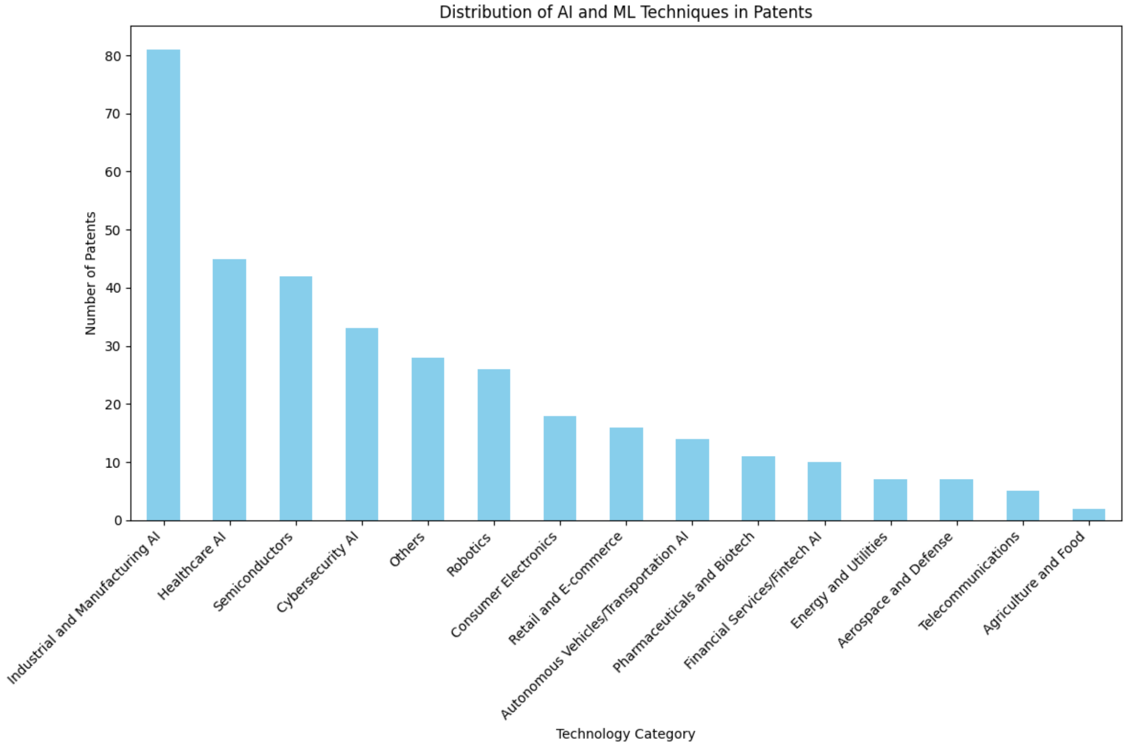

Figure 8: Patents labeled by Industry

The figure above indicates that the model can indeed distinquish between different industries within the patent abstracts. The main application theme going on within the sample appears to be related to industrial and manufacturing which is to be expected, as many companies utilize AI and ML techniques to up the productivity and efficiency of their facilities through predictive analytics [Loh24].

Additionally, healthcare seems to be also be a large field, which is likely due to an increased attention on the possible applications for diagnosis and recommendation for treatments through the use of AI systems [DK19].

As this is just a very small sample of the total patent landscape, this picture is obviously not encompassing reality in its entirety. The same method would need to be applied on a much larger scale across all patents, which would require access to GPU power to run the labeling for throughout the patent landscape. It does however show the potential of analyzing industry applications of AI technology through LLMs, which can be utilized to form more meaningful insights into the landscape as a whole.

8.2 RAG implementation analysis

In this section I will go over the testing and analysis of the question-answering implementation of the LLM. As mentioned in the section of technical implementation, this involved utilizing a FAISS index for storing the patent abstracts, and subsequently fetching relevant ones based on embedding vectors that are most closely related to the user query. To increase the likeliness that relevant patents are fetched, I have enriched the abstracts further with keywords generated by the LLM beforehand. For this purpose it is necessary to both evaluate the systems ability to fetch relevant data, as well as evaluate the responses generated by the LLM based on this data.

8.2.1 Evaluating the RAG system

For the purpose of testing, I will test a variety of general technical questions, as well as some questions that are more specific to a field or technique within a specific field, based on the LLM-based classified labels from earlier sections. By asking about these specific classified labels, it increases the likelihood that the model will be able to find relevant patents. However, for a full scale implementation, the system should ideally be able to answer any given query for AI technology in general. The would only require embeddings and inserting all the patents into an index for fetching, as well as utilizing a model with a larger context window, so that more patents can be evaluated.

The following table shows the test queries used, the respective average similarity scores across the five collected patents for each query, as well as a total average similarity across all queries:

Query	**Average Similarity Score**
What are the key trends in AI and machine learning patents?	0.294
What innovations in deep learning are highlighted in the patents?	0.334
What are the main applications of computer vision in these patents?	0.415
What are the common challenges addressed in AI patents?	0.281
How is reinforcement learning being applied in technology?	0.461
What are the advancements in computer vision technology?	0.365
Describe the advancements in healthcare AI technologies	0.511
Describe how autonomous vehicles are being developed using AI	0.498
How is quantum computing being integrated with AI technologies?	0.342
What innovations in robotics are highlighted in the patents?	0.293
Describe the use of neural networks in technology	0.536
What are the latest innovations in natural language processing according to the patents?	0.401
How are transformer models being utilized in technology?	0.225
What are the primary applications of reinforcement learning in patented technology?	0.399
How is object detection being improved according to recent patents?	0.451
How is AI being used to improve healthcare diagnostics according to the patents?	0.528
How is AI being used in healthcare applications?	0.514
What are the primary applications of AI in finance?	0.392
How is the method of machine learning being used in financial technologies?	0.476
Overall Average Similarity Score	0.407

Table 2: Average Similarity Scores for Each Query

The average similarity score of 0.407 indicates that the retrieved patent abstract embedding vectors are only around 40.7 percent similar to the query embedding vectors across all the different queries. More specifically this means that the fetched patents on average are moderately relevant to the posed question, which in turns indicate that there is room for improvement within the system. I will discuss some possible options for optimization further in section 9 and 10.

Looking at the queries with a relatively high similarity score, such as "Describe the advancements in healthcare AI technologies" and "How is AI being used to improve healthcare diagnostics according to the patents?" indicate that the sample likely contains high numbers of healthcare related patents, which aligns with figure 8 in the previous section 8.1, as the Healthcare AI category was the second largest. Similarly, we see that questions related to autonomous vehicles and object detection also have a relatively high score. This is surprising, as the transportation category only held about 20 out of the 300 patents, however it's likely due the test query aligning well with the injected keywords. Object detection is a more general technique that is used in a wide variety of applications, which is likely the reason this query scored so high.

On the opposite, we see that the query "How are transformer models being utilized in technology?" and "What are the common challenges addressed in AI patents?" have scored relatively low, suggesting that the system cannot answer questions that are more general in nature. One possible reason for this is due to the abstract contents being either more diverse or less specific than expected, meaning that the resulting query embeddings don't align particularly well with the abstracts and injected keywords that were generated by the LLM beforehand. A more likely reason to this is the limited number of patents within the sample, as this would result in a smaller coverage of topics, which does not align well with the relatively broad and varied nature of AI systems.

Next I will evaluate the calculated precision, recall and mean reciprocal scores (MRR) in the table below:

Metric	Value
Precision	0.832
Recall	0.832
MRR	0.947

Table 3: Precision, Recall, and MRR for the Retrieval System

As mentioned earlier in the paper, the above scores function as an additional indicator as to whether or not the system is correctly retrieving relevant documents.

The calculated precision score of 0.832 shows that roughly 83.2 percent of the retrieved documents were relevant to their respective queries. This relatively high score indicates that the retrieval system based on cosine similarity, overall is proving efficient in returning relevant documents to user queries.

It can also be observed that the recall score is the same, 83.2 percent, which indicates that the system is also efficient in fetching the most relevant patent abstracts within the index.

The MRR score of 0.947 shows that on average, the system not only finds the most relevant abstracts, but in 94.7 percent of the queries also rank these correctly by their respective relevance to the users query. This means that the documents that are injected into the prompt for context, are always the most relevant documents within the index. This score is important for further optimization of the system, as injecting the most relevant documents into the prompt first, can help reduce context window issues by always making sure that they get included. In summary, these scores collectively mean that based on the small sample utilized for the proof of concept, the system is capturing a significant portion of the relevant information within the index, and is therefore able to efficiently answer questions relating to the technology.

Next we will attempt manually evaluate a few of the generated responses in relation to their respective queries, and determine whether the response seem relevant to the query. I will test queries that have scored both relatively high and low in terms of the average similarity scores provided earlier. The table below shows some of the earlier queries, and the system-generated responses, as well as their average similarity score

Query and Response
Query: What are the common challenges addressed in AI patents? **Response:** Common challenges in AI patents include data privacy and security, improving the accuracy and efficiency of AI models, and addressing ethical concerns related to AI deployment. (Average Similarity score: 0.281)
Query: What innovations in robotics are highlighted in the patents? **Response:** Innovations in robotics patents include advancements in robotic control systems, sensor integration, and AI-driven decision-making processes, enhancing the functionality and autonomy of robots. (Average Similarity score: 0.293)
Query: Describe the advancements in healthcare AI technologies **Response:** Healthcare AI technologies have seen advancements in predictive analytics, personalized medicine, and diagnostic tools, leveraging AI to improve patient outcomes and streamline healthcare processes. (Average Similarity score: 0.511)
Query: What are the main applications of computer vision in these patents? **Response:** Computer vision patents mainly apply to areas like autonomous vehicles, healthcare diagnostics, facial recognition, and surveillance systems, highlighting the versatility of computer vision technologies. (Average Similarity score: 0.415)

Table 4: Selected queries and corresponding Responses from the RAG-system

The model appears to be responding with relevant and complete answers. Despite a low similarity score between the query and fetched abstracts, it has correctly identified some of the most central and common challenges that exists within AI technology. Data privacy, security and accuracy of models remains a hot topic of research and policy making within the field of AI to this day. The application of AI have also raised a series of ethical questions as to how data should be handled and how these models can be interpreted [AMR19].

It has also correctly identified innovations within the field of robotic AI, such as AI-driven decision making, which likewise is extensively researched for autonomous agent systems in academic literature [RAAA22].

The system has also managed to identify some of the key trends within healthcare related AI, which focuses on utilizing predictive methods for diagnosis, and increase efficiency of processes within the field. In this case, it does however not mention the usage of computer vision, which means that it has failed to capture some aspects of AI use cases within healthcare that are highly relevant When asked specifically about computer vision however, it does identify healthcare as a field in which this is applied [BM20].

It has also identified several other key areas in which computer vision have been utilized, such as autonomous vechicles where computer vision is used for object and lane detection [GSS23], as well as facial recognition and surveillance systems [SSN23].

In summary, the responses have shown that the model can apply reasoning power to queries and utilize the contextual information provided through the enriched abstracts in a meaningful way. This means that further improving the system should only enable even more contextually relevant answers, and as mentioned earlier in the paper, I will discuss some possible avenues of improvement in the following section.

9 Discussion

In order to conclude the project, I will discuss each of the different questions posed in the introduction and attempt to answer and discuss these in light of what may have could have been improved upon with the implementation. Finally, I will summarize the discussion of the results in this paper, before making a final conclusion in the last and following section.

The field of patent analysis in technology mapping has grown significantly over the years, incorporating a wide range of methodologies and tools to extract meaningful insights from patent data such as patent counts, as well as more advanced techniques such as citation and network analysis. When applied successfully, these methods are essential for understanding technological trends, innovation pathways, and the impact of intellectual property on economic and technological development. The strengths and insights gathered of these methods was shown in the review to vary significantly, depending on the objective of the project at hand.

One of the main methods and points of data that have been used both in literature and in this project, is patent counts to determine innovation activity within AI across geographical hot spots. By counting the number of patents filed or granted over a period, researchers can gauge the level of innovative activity within a specific field or region. Patent counts offered a straightforward measure of inventive output and are useful for identifying trends and patterns in innovation. However, this method has notable limitations. It does not account for the quality or economic impact of the patents, leading to potential misinterpretations of innovation activity. This has also helped identify some of the main companies that act on this particular field. By extending with a qualitative examination of the results from this analysis, it has shown great potential, as it has indicated several key industries as well as underlying factors for innovation both in countries and companies. Many of these factors, such as collaborative efforts between different actors that were found, align with the literature review, that indicated that innovation was to be seen as a collective effort that involve continuous interactions between these actors. Additionally, by combining the insights of patents counts across industries, countries and time with a qualitative evaluation of these, it was possible to determine that AI does function in alignment with the characteristics of a General Purpose Technology. It exhibited a wide variety of use-cases across many different technological fields, such as robotics, quantum computing, finance, healthcare and so forth as presented in section 7.

Network and citation analysis was shown to be a method that maps the relationships among patents, inventors, organizations, and other entities within technology, as well as measure their impact on technology. By examining the references made by patents to prior patents, it becomes possible to understand the impact of either a specific entity or technology in a broader perspective. It can also help determine the flow of knowledge, predict emerging trends through recent patent activity, identify central actors, and the overall structure of innovation ecosystems. The use of centrality measures helps pinpoint patents that are pivotal within the network, acting as foundational technologies or crucial bridges between different technological domains. By employing this type of advanced analytical techniques, researchers and organizations can extract actionable insights from patent data, guiding their efforts in fostering innovation, developing new technologies, and maintaining a competitive edge within AI technology.

Network analysis is also used to identify clusters or communities within the network, revealing groups of related technologies and innovations. In this project, I attempted to utilize citation network analysis to identify these central clusters by the help of Louvain Community detection. This is a specific technique within network analysis that focuses on identifying clusters or groups of nodes within a network that are more densely connected to each other than to the rest of the network. However, as the network construction in this project was based on backwards citations between patents, the analysis inherently relies on these citations between patents to construct a valuable network. That is likely the reason that the community analysis showed the network as heavily fragmented, and therefore did not provide much insight into the network structure and interactions as expected. The unusually high number of clusters indicated that there appears to be little to no interaction between the patented technology.

There was however a fewer larger clusters that could be identified, each containing several thousand patents. These individual clusters was shown to have a high internal cohesiveness, which does suggest that some thematic trends was detected, and despite the remaining network not being particularly insightful, these could still provide some key patents. This has however likely had some negative effects on the subsequent evaluations, as this would have become less inclusive in terms of the gathered information, and not encompassing a larger picture of the field. Additionally, limiting the analysis to only account for patent data as opposed to widening the scope to include other technological fields as well as academic literature, has likely also been a contributing factor, as innovation does not just happen in industry, but also within academia as well as the open source community of AI.

An additional method within patent analysis literature was found to be text mining, which leverages data science techniques to analyze the textual content of patents. This method has involved extracting specific keywords from patent documents to identify trends and patterns in technological development. This technique is also the main field in which Large Language Models

(LLMs) was shown to be utilized. Literature showed that the integration of LLMs into patent analysis represented a significant advancement in how intellectual property can be managed and utilized. LLMs are adept at categorizing text into predefined categories, which is essential in the patent domain where accurate classification facilitate efficient patent processing. Using LLMs, patents can be automatically classified according to the Cooperative Patent Classification (CPC) system or the International Patent Classification (IPC) system based on their textual content. This automation not only speeds up the classification process but also enhances consistency in how patents are categorized across different patent offices.

A similar method was attempted in this project in section 8, where an attempt was made to categorize a selected sample size of roughly 300 patents that were derived from the the eigenvector analysis in section 7.2.1 as well as the largest clusters identified in the community detection analysis in section 7.2.2. Categorizing these patents had a dual function, as it contributed both to determining the characteristics of AI technology within a large variety of possible applications and industries, but also to evaluate the reasoning power of the chosen model in the context of technical patent text. The purpose of evaluating the reasoning power of the LLM, was due to the objective of implementing a proof-of-concept version of a RAG-based system, the strengths of which was covered in section 4.3.6.

The already large and growing volume of patent documentation makes manual search and retrieval of relevant patents challenging. LLMs can transform patent search by understanding complex queries in natural language and retrieving relevant documents with high precision. Unlike traditional keyword-based search engines, LLMs can interpret the intent of a query, allowing them to return results that are contextually related. In addition, LLMs can be trained to recognize and generate legally precise text, aiding legal professionals and patent offices in drafting claims and descriptions that have strict legal requirements. Therefore, the integration of advanced AI tools like LLMs represents a significant advantage if applied correctly, providing more nuanced and comprehensive analysis of patent data. As these models continue to evolve, their impact on the efficiency, accuracy, and strategic insight of patent-related activities can only be expected to grow, transforming the landscape of innovation management.

As discussed, the field of patent analysis is critical for understanding the dynamics of technological innovation. The methodologies discussed, patent counts, text mining, citation analysis, network analysis, and community detection, each offer unique insights and advantages. However, the integration of advanced AI tools like LLMs represents a significant advantage if applied correctly, providing more nuanced and comprehensive analysis of patent data. The integration of LLMs into patent analysis not only compliments traditional methodologies but also paves the way for innovative approaches to handling an ever-growing corpus of patent literature. As these models continue to evolve, their impact on the efficiency, accuracy, and strategic insight of patent-related activities is expected to grow, transforming the landscape of innovation management. Combing the methodologies should therefore not only be feasible, but also valuable in terms of gaining key insights into this innovation system.

The implemented proof of concept RAG-system did also show capable of performing relatively well, given a small sample size. Whilst the calculated average cosine similarity scores initially did not appear to show a particularly high semantic relevance, the collective precision, recall and MRR scores remained relatively high, indicating that the major point of improvement in the system would be to introduce even more patent data, and perhaps inject even further contextual information into the abstracts, before embedding these and injecting them into the FAISS index. It was observed, that it did manage to provide reasonable responses to the test queries, and that these were in factually validated as being correct applications and perspectives through the provided sources in section 8.2.1. The attempt to implement this system however may have been fundamentally flawed due to the initial issues in relation to the constructing a fragmented network graph. Due to its fragmented construction, the indicated patents may or may not actually be the most central, thus not providing a fully indicative picture of the central technologies within the field. Subsequent analysis of these patents therefore can not be deemed successful as a result.

9.1 Summary

In summary, this discussion has highlighted the significant growth in the field of patent analysis and technology mapping, as patent data has emerged as a valuable resource for understanding technological trends and the impact of intellectual property on economic and technological development. Through this project, we have examined various methodologies used in patent analysis. Furthermore the implementations of some of these have been discussed, such as patent counts for pinpointing key companies and industries in AI, citation network analysis for identifying key patented technology and text mining by using LLMs to determine impact of these key patents. By combining these quantitative methods with qualitative examinations, the project identified underlying factors for innovation in different countries and companies, aligning with existing literature that views innovation as a collective effort involving continuous interactions between various actors. The findings also support the characterization of AI as a General Purpose Technology, with applications spanning multiple fields such as robotics, quantum computing, finance, and healthcare.

The proof-of-concept RAG system implemented in the project demonstrated reasonable performance, with high precision, recall, and MRR scores. Despite initial challenges with semantic relevance, the system managed to provide reasonable responses to test queries. This performance suggests that the major point of improvement would be to introduce more patent data and inject further contextual information into the abstracts before embedding and indexing them. However, the fragmented network graph likely affected the overall success of the project, as it might not fully represent the central technologies within the field.

10 Conclusion

In order to conclude the project, I will answer the central question posed at the beginning of the project, based on the discussion and results. The main question posed at the begging was the following:

'How can technology mapping be utilized in conjunction with current AI and network graphing methods to extract meaningful insights into the field of technological innovation and General Purpose Technologies?'

This project has been an attempt to not only explore a framework that consist of various selected methodologies for analyzing patent data but to utilize these to map technological trends within AI. By employing patent counts, citation analysis, network analysis, and the integration of Large Language Models (LLMs), the project aimed to provide a comprehensive understanding of the AI patent landscape.

The use of patent counts highlighted the level of innovative activity in different regions and industries. This method, despite its limitations, effectively identified key companies and technological fields, aligning with existing literature that views innovation as a collective effort involving continuous interactions between various actors. The project's findings support the characterization of AI as a General Purpose Technology, demonstrating its wide range of applications across multiple fields. In addition, it appears that the driving forces of innovation within the technological field of AI, which enable it to function as a General Purpose Technology, can also be attributed to the efficiency gains of these types of systems. Since AI makes processes so much more efficient, it not only undergoes continuous development itself but likewise spurs innovation by being integrated into other systems that are further innovated upon. This gives it a particularly strong position within the entire innovation ecosystem as a whole, being further evidence of its position as a General Purpose Technology.

However, due to unforeseen issues, the knowledge derived from the analysis that supports these claims may not be as insightful as they may seem. It appears that some error may have occurred in terms of the patent data queries, which have resulted in repeating patents due to the way in which companies file across different jurisdictions. As PATSTAT contains patents from all over the world, failing to properly filter based on additional criteria means that the patent counts may have been overestimated due to this issue. Therefore, the insights generated from the initial analysis cannot be concluded as valuable in their current state due to this issue of repeating patents. Among other observations, this could explain the large concentration of AI and ML-related patents within the US, as observed in Figure 3 in Section 7.1.3, as well as the large concentration of patents related to IBM in Figure 2 in Section 7.1.2.

Citation analysis provided insights into the impact of specific inventions and the flow of technological knowledge, while network and citation analysis mapped the relationships among patents, inventors, organizations, and other entities. This approach helped identify central actors and the structure of innovation ecosystems, with centrality measures pinpointing pivotal patents within the network. However, challenges with the network construction, based on backward citations, led to a fragmented network that limited the insights gained from community detection. The discovery of this issue was primarily due to the rather high number of community clusters detected in Section 7.2.2

Text mining and the integration of LLMs represented significant advancements in patent analysis. These methodologies enabled the automatic classification of patents and improved the consistency and speed of the classification process. The proof-of-concept RAG system demonstrated reasonable performance, with high precision, recall, and MRR scores, indicating its potential for providing nuanced and comprehensive analysis of patent data. However, further improvements could be made by increasing the volume of patent data and injecting more contextual information into the abstracts before embedding and indexing them

The limitations encountered, such as the fragmented network graph and reliance on patent data alone, suggest areas for future research. Expanding the scope to include other technological fields and academic literature could provide a more comprehensive understanding of innovation within AI. Additionally, exploring more sophisticated methods such as GraphRAG, which combines graph-based neural networks with retrieval-augmented generation, could leverage the rich relational information inherent in patent data and offer enhanced retrieval and generation capabilities.

Overall, this project has shown that combining various methodologies has the potential to provide significant and valuable insights into the dynamics of technological innovation. The integration of advanced AI tools like LLMs complements traditional methodologies, offering a more nuanced analysis of patent data. As these models continue to evolve, their impact on the efficiency, accuracy, and strategic insight of patent-related activities is expected to grow, transforming the landscape of innovation management. This project underscores the importance of a multifaceted approach to patent analysis, leveraging both traditional and advanced techniques to understand the complexities of technological innovation. By addressing the identified limitations and incorporating more advanced methodologies, future research can build on these findings to further enhance the understanding and management of technological innovation in the field of AI.

11 Future research

This project's attempt at a complex combination of methodologies with the the integration of Large Language Models (LLMs) needs further exploration. The landscape of AI innovation is a system of continuous interaction among various actors, and this project has highlighted potential areas for future research that could help cover gaps the current understanding and expand the limits of traditional methodologies.

One direction I aim to dive deeper, into is the system of patent data that I have only just begun to explore. By addressing the issue of repeating patents across jurisdictions, it should be possible to enhance the accuracy of patent counts as well as correct the network analysis and ensure that it is reflecting the true nature of the technological innovation of AI. This adjustment will likewise provide a clearer picture of the innovative activity within different regions and industries. Additionally, this should also ensure that the final RAG implementation will be significantly more functional, as it will be receiving even more patents with further contextual information. Furthermore, including other data sources, such as scientific publications, inventor and legal data, this further can enrich the analysis, creating a much better perspective of technological trends, as well as significantly improve the tested RAG system.

Finally, a particularly promising approach for future research is the implementation of a GraphRAG method. Given the network structure of the data used in this project, GraphRAG could offer a more appropriate and effective way to handle the relationships between patents. GraphRAG combines graph-based neural networks with retrieval-augmented generation, allowing the model to leverage the rich relational information inherent in the patent data. This method could enhance the retrieval and generation capabilities of the system, providing more accurate and contextually relevant insights. The integration of graph-based techniques aligns well with the interconnected nature of patent data and could lead to significant improvements in the analysis and visualization of technological trends.

In conclusion, my future research should aim to integrate more diverse data sources, employ additional advanced network analysis techniques, and explore innovative methodologies like GraphRAG. By doing so, it should be possible to gain a more comprehensive understanding of the AI patent landscape and further enhance the tools available for innovation management and technological forecasting.

References

[AA] Dean Alderucci and Kevin Ashley. Using ai to analyze patent claim indefiniteness using ai to analyze patent claim indefiniteness.

[AAV02] Zoltan J. Acs, Luc Anselin, and Attila Varga. Patents and innovation counts as measures of regional production of new knowledge. *Research policy*, 31(7):1069–1085, 2002. ObjectType-Article-1; ID: cdi$_{p}roquest_{j}ournals_{2}$23247162.

[Ada01] Stephen Adams. Comparing the ipc and the us classification systems for the patent searcher. *World Patent Information*, 23(1):15–23, 2001. ID: 271774.

[AH92] Philippe Aghion and Peter Howitt. A model of growth through creative destruction. *Econometrica*, 60(2):323–351, 1992. ObjectType-Article-2; ID: cdi$_{p}roquest_{j}ournals_{2}$03879986.

[AMR19] Ari Alamäki, Marko Mäki, and RM Ratnayake. Privacy concern, data quality and trustworthiness of ai-analytics. *Proceedings of Fake Intelligence Online Summit 2019*, 2019.

[Ara19] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models, 2019.

[Bai24] Baidu. Baidu ai open platform-the world's leading artificial intelligence service platform, 2024.

[BCB16] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate, 2016.

[BCZ$^+$16] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29, 2016.

[BDPd$^+$92] Peter F. Brown, Vincent J. Della Pietra, Peter V. deSouza, Jenifer C. Lai, and Robert L. Mercer. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–480, 1992.

[BGLL13] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks, -07-31 2013.

[BGMMS21] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.

[BJT23] Francis Bloch, Matthew O. Jackson, and Pietro Tebaldi. Centrality measures in networks, -04-24 2023.

[BM14] Erik Brynjolfsson and Andrew McAfee. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* WW Norton & Company, 2014.

[BM20] Adam Bohr and Kaveh Memarzadeh. The rise of artificial intelligence in healthcare applications. *Artificial Intelligence in Healthcare*, pages 25–60, 2020.

[BMR$^+$20] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.

[Bon72] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, 2(1):113–120, 1972. doi: 10.1080/0022250X.1972.9989806.

[BT95] Timothy F. Bresnahan and M. Trajtenberg. General purpose technologies 'engines of growth'? *Journal of Econometrics*, 65(1):83–108, 1995. ID: 271689.

[CAP$^+$17] Kwok Tai Chui, Wadee Alhalabi, Sally Shuk Han Pang, Patricia Ordóñez de Pablos, Ryan Wen Liu, and Mingbo Zhao. Disease diagnosis in smart healthcare: Innovation, technologies and applications. *Sustainability*, 9(12):2309, 2017.

[CH14] Jinho Choi and Yong-Sik Hwang. Patent keyword network analysis for improving technology development efficiency. *Technological Forecasting and Social Change*, 83:170–182, 2014. ID: 271733.

[Chr97] Clayton M. Christensen. *The innovator's dilemma : when new technologies cause great firms to fail.* Harvard Business Review Press, Paris, France, reprint edition, 1997.

[CKG$^+$19] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*, 2019.

[CQX15] George R. G. Clarke, Christine Zhenwei Qiang, and Lixin Colin Xu. The internet as a general-purpose technology: Firm-level evidence from around the world, -07-31 2015.

[CSB00] Pere Escorsa Castells, Marisela Rodriguez Salvador, and Ramon Maspons Bosch. Technology mapping, business strategy, and market opportunities. *Competitive Intelligence Review*, 11(1):46–57, /01/01 2000.

[CvMBB14] Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder–decoder approaches. In *SSST@EMNLP*, 2014.

[CXMR07] P. Chen, H. Xie, S. Maslov, and S. Redner. Finding scientific gems with google's pagerank algorithm, -01 2007.

[DCLT19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[DGD$^+$24] Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. The faiss library, 2024.

[DK98] Emmanuel Duguet and Isabelle Kabla. Appropriation strategy and the motivations to use the patent system: An econometric analysis at the firm level in french manufacturing. *Annales d'économie et de statistique*, (49):289–327, 1998. ObjectType-Article-1; ID: cdi$_{h}al_{p}rimary_{o}ai_{H}AL_{h}al_{0}3592981v1$.

[DK19] Thomas Davenport and Ravi Kalakota. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2):94–98, June 1, 2019.

[Dor88] Ronald Dore. Technology policy and economic performance; lessons from japan : Christopher freeman, (frances printer publishers, london, new york, 1987) pp. 155, £20.00, 1988. ID: cdi$_{r}epec_{p}rimary_{e}eerespol_{v3}a17_{3}ay_{3}a1988_{3}ai_{3}a5_{3}ap_{3}a309_{3}10_{h}tm$.

[DSM$^+$21] Jesse Dodge, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. Documenting large webtext corpora: A case study on the colossal clean crawled corpus. *arXiv preprint arXiv:2104.08758*, 2021.

[EK02] Dieter Ernst and Linsu Kim. Global production networks, knowledge diffusion, and local capability formation. *Research policy*, 31(8-9):1417–1429, 2002.

[Elm90] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990. ID: 272076.

[Ent24] Huawei Enterprise. Infusing intelligence into enterprise 'neurons' - huawei publications, 2024.

[ERC$^+$20] Julian Martin Eisenschlos, Sebastian Ruder, Piotr Czapla, Marcin Kardas, Sylvain Gugger, and Jeremy Howard. Multifit: Efficient multi-lingual language model fine-tuning, 2020.

[FASR15] Patrícia Fortes, António Alvarenga, Júlia Seixas, and Sofia Rodrigues. Long-term energy scenarios: Bridging the gap between socio-economic storylines and energy modeling. *Technological Forecasting and Social Change*, 91:161–178, 2015. ID: 271733.

[FC22] Luciano Floridi and Josh Cowls. A unified framework of five principles for ai in society. *Machine learning and the city: Applications in architecture and urban design*, pages 535–545, 2022.

[FKJ07] Lee Fleming, I. II King, Charles, and Adam I. Juda. Small worlds and regional innovation. *Organization science (Providence, R.I.)*, 18(6):938–954, 2007. ObjectType-Article-2; ID: cdi$_{p}roquest_{m}iscellaneous_{3}6821268$.

[FND$^+$20] Brooks Foxen, Charles Neill, Andrew Dunsworth, Pedram Roushan, Ben Chiaro, Anthony Megrant, Julian Kelly, Zijun Chen, Kevin Satzinger, Rami Barends, et al. Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms. *Physical Review Letters*, 125(12):120504, 2020.

[For10] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010. ID: 271542.

[Fre77] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977. 13.

[FV02] Jan Fagerberg and Bart Verspagen. Technology-gaps, innovation-diffusion and transformation: an evolutionary interpretation. *Research policy*, 31(8-9):1291–1304, 2002.

[FY11] Maryann P. Feldman and Ji Woong Yoon. An empirical test for general purpose technology: an examination of the cohen–boyer rdna technology. *Industrial and Corporate Change*, 21(2):249–275, July 23, 2011.

[GG23] A. Shaji George and A. S. Hovan George. A review of chatgpt ai's impact on several business sectors. *Partners Universal - International Innovation Journal*, 1(1):9–23, 2023.

[GHT$^+$18] S. Ghosh, M. Halappanavar, A. Tumeo, A. Kalyanaraman, H. Lu, D. Chavarrià-Miranda, A. Khan, and A. Gebremedhin. Distributed louvain algorithm for graph community detection. In *2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 885–895, 2018. ID: 1.

[Goo24] A. I. Google. Making ai helpful for everyone - google ai, 2024.

[Gri90] Zvi Griliches. Patent statistics as economic indicators: A survey. *Journal of economic literature*, 28(4):1661–1707, 1990. ObjectType-Article-1; ID: cdi$_{p}roquest_{m}iscellaneous_{5}$9550229.

[Gro24] Ping An Group. Ping an group 2023: A year of strategic development and sustained growth, Feb 8, 2024.

[GSS23] Henil Gajjar, Stavan Sanyal, and Manan Shah. A comprehensive study on lane detecting autonomous car using computer vision. *Expert Systems with Applications*, 233:120929, 2023. ID: 271506.

[GTT23] Avi Goldfarb, Bledi Taska, and Florenta Teodoridis. Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings. *Research Policy*, 52(1):104653, 2023. ID: 271666.

[Gue24] Federico Guerrini. Uk and germany double down on joint ai, clean energy rd efforts, Mar 15, 2024.

[GXG$^+$24] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. Retrieval-augmented generation for large language models: A survey, 2024.

[Hen24] Calvin Hennick. Getting to systemic sustainability, May 29, 2024.

[HKMSZ24] Amin Hajivand, Ali Khosh Manzar, and Saber Sayari Zuhan. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *Fares Law Research*, 6(18):73–90, 2024.

[HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[HTJ01] Bronwyn H. Hall, Manuel Trajtenberg, and Adam B. Jaffe. The nber patent citation data file: Lessons, insights and methodological tools, 2001. ID: cdi$_{p}roquest_{j}ournals_{1}$689555710.

[Hul20] Steve Hulet. Artificial intelligence 101: The key concepts of ai, -09-22T18:12:00+00:00 2020.

[Int24] Intel. Explore intel® artificial intelligence solutions, 2024.

[JHY⁺24] Hongye Jin, Xiaotian Han, Jingfeng Yang, Zhimeng Jiang, Zirui Liu, Chia-Yuan Chang, Huiyuan Chen, and Xia Hu. Llm maybe longlm: Self-extend llm context window without tuning. *arXiv preprint arXiv:2401.01325*, 2024.

[Jon] Karen Spärck Jones. A statistical interpretation of term specificity and its application in retrieval.

[JTH93] Adam B Jaffe, Manuel Trajtenberg, and Rebecca Henderson. Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3):577–598, 1993.

[KL] Mingyu Kim and Juyong Lee. What are the future trends in natural gas technology to address climate change? patent analysis through large language model.

[KLLL20] Dylan Myungchul Kang, Charles Cheolgi Lee, Suan Lee, and Wookey Lee. Patent prior art search using deep learning language model. In *Proceedings of the 24th Symposium on International Database Engineering & Applications*, IDEAS '20, New York, NY, USA, 2020. Association for Computing Machinery.

[KPSS21] Leonid Kogan, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller. *Technology-skill complementarity and labor displacement: Evidence from linking two centuries of patents with occupations.* National Bureau of Economic Research, 2021.

[LCB05] R.G. Lipsey, K.I. Carlaw, and C.T. Bekar. *Economic Transformations: General Purpose Technologies and Long-Term Economic Growth.* OUP Oxford, 2005.

[LH20a] Jieh-Sheng Lee and Jieh Hsiang. Patent claim generation by fine-tuning openai gpt-2. *World Patent Information*, 62:101983, 2020. ID: 271774.

[LH20b] Jieh-Sheng Lee and Jieh Hsiang. Patent classification by fine-tuning bert language model. *World Patent Information*, 61:101965, 2020. ID: 271774.

[Li09] Xibao Li. China's regional innovation capacity in transition: An empirical approach. *Research policy*, 38(2):338–357, 2009.

[Loh24] Peter Loh. The dawn of ai in manufacturing: Understanding its wide reaching impact on industry, -04-09T19:32:53+00:00 2024.

[LPP19] Ian Marques Porto Linares, Alex Fabianne De Paulo, and Geciane Silveira Porto. Patent-based network analysis to understand technological innovation pathways and trends, -04-26 2019.

[LT24] Jonathan Larson and Steven Truitt. Graphrag: A new approach for discovery using complex information, -02-13T20:00:00+00:00 2024.

[Mac24] International Business Machines. Ibm watsonx.ai, 2024.

[Mar20] Gary Marcus. The next decade in ai: Four steps towards robust artificial intelligence, 2020.

[MCCD13] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[Mic24] Microsoft. Artificial intelligence solutions | microsoft ai, 2024.

[MR99] David C Mowery and Nathan Rosenberg. *Paths of innovation: Technological change in 20th-century America.* Cambridge University Press, 1999.

[MRS08] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval.* Cambridge university press, 2008.

[NH20] Mark J Nelson and Amy K Hoover. Notes on using google colaboratory in ai education. In *Proceedings of the 2020 ACM conference on innovation and Technology in Computer Science Education*, pages 533–534, 2020.

[NW84] Richard R. Nelson and Sidney G. Winter. An evolutionary theory of economic change // review, 1984. ID: cdi$_{p}roquest_{j}ournals_{1}$99957755.

[oESoC23] Universty of Electronic Science and Technology of China. Master majors research fields-university of electronic science and technology of china, 2023.

[Off] EPO European Patent Office. Patstat online | wipo inspire.

[Pav95] Keith Pavitt. National systems of innovation: Towards a theory of innovation and interactive learning: Bengt-ake lundvall (editor), (pinter, london, 1992) pp. 317, £45 (hardbook) isbn 1-85567-063-1, 1995. ID: cdi$_{r}epec_{p}rimary_{e}eerespol_{v}3a24_{3}ay_{3}a1995_{3}ai_{3}a2_{3}ap_{3}a320_{3}20_{h}tm$.

[Pen95] Edith T. Penrose. *The Theory of the Growth of the Firm.* Oxford University Press, Oxford, 1 edition, 1995. SourceType-Books-1; ID: cdi$_{p}roquest_{m}iscellaneous_{5}6684956$.

[Pet20] Sergio Petralia. Mapping general purpose technologies with patent data. *Research Policy*, 49(7):104013, 2020. ID: 271666.

[RAAA22] Ammar Abdul Ameer Rasheed, Mohammed Najm Abdullah, and Ahmed Sabah Al-Araji. A review of multi-agent mobile robot systems applications. *International Journal of Electrical and Computer Engineering*, 12(4):3517–3529, 2022.

[Ref] Wipo technology trends 2019 artificial intelligence.

[Ref03] Patents, citations innovations: A window on the knowledge economy: By adam b. jaffe and manuel trajtenberg. the mit press, cambridge, ma. (2002). 478 pages. 50. $Computersmathematicswithapplications(1987),45(10):1774,2003. ID: cdi_{e}lsevier_{s}ciencedirect_{d}oi_{1}0_{1}016_{S}0898_{1}221_{0}3_{8}0127_{X}$.

[Res24] LG AI Research. Data intelligence, 2024.

[RG20] Nils Reimers and Iryna Gurevych. Making monolingual sentence embeddings multilingual using knowledge distillation. https://huggingface.co/sentence-transformers/all-mpnet-base-v2, 2020.

[Rik22] Cecilia Rikap. Becoming an intellectual monopoly by relying on the national innovation system: the state grid corporation of china's experience. *Research Policy*, 51(4):104472, 2022. ID: 271666.

[RM21] Laria Reynolds and Kyle McDonell. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2021.

[Rog03] Everett M. Rogers. *Diffusion of innovations.* Simon Schuster, 5 edition, 2003. SourceType-Books-1; ID: cdi$_{p}roquest_{m}iscellaneous_{3}7691479$.

[RSR^{+}23] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer, 2023.

[Rud19] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019.

[RVD20] Vincent Van Roy, Daniel Vertesy, and Giacomo Damioli. Ai and robotics innovation, 2020.

[Sam24] Samsung. Galaxy ai | mobile ai and ai features on devices, 2024.

[Sch27] Joseph Schumpeter. The explanation of the business cycle. *Economica*, (21):286–311, 1927.

[Sch76] Joseph A. Schumpeter. *Capitalism, Socialism and Democracy.* Routledge, 1976. ID: cdi$_{i}nformaworld_{t}aylorfrancisbooks_{9}780203202050$.

[Sch80] Joseph Schumpeter. *The theory of economic development: an inquiry into profits, capital, credit, interest, and the business cycle.* Routledge, 1980. ID: cdi$_{a}skewsholts_{v}lebooks_{9}781351472210$.

[Sco04] Suzanne Scotchmer. *Innovation and incentives.* MIT press, 2004.

[SGdV^{+}18] Priscila Gonçalves Vasconcelos Sampaio, Mario Orestes Aguirre González, Rafael Monteiro de Vasconcelos, Marllen Aylla Teixeira dos Santos, José Carlos de Toledo, and Jonathan Paulo Pinheiro Pereira. Photovoltaic technologies: Mapping from patent analysis. *Renewable and Sustainable Energy Reviews*, 93:215–224, 2018. ID: 271969.

[SGM19] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*, 2019.

[SSN23] Manan Shah, Arpit Singh, and Vishal Nayak. Automation of surveillance systems using deep learning and facial recognition | international journal of system assurance engineering and management. *International Journal of System Assurance Engineering and Management*, 14(1):236–245, Jan 6, 2023.

[SSW14] Andreas Schaefer, Daniel Schiess, and Roger Wehrli. Long-term growth driven by a sequence of general purpose technologies, -02 2014.

[STG^{+}23] Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Le Hou, Kevin Clark, Stephen Pfohl, Heather Cole-Lewis, Darlene Neal, Mike Schaekermann, Amy Wang, Mohamed Amin, Sami Lachgar, Philip Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Aguera y Arcas, Nenad Tomasev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle Barral, Dale Webster, Greg S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Natarajan. Towards expert-level medical question answering with large language models, 2023.

[SVL14] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks, 2014.

[SYAN20] H. Sasaki, S. Yamamoto, A. Agchbayar, and N. Nkhbayasgalan. Extracting problem linkages to improve knowledge exchange between science and technology domains using an attention-based language model. 10(4):5903–5913, 2020.

[Ten24] Tencent. About us - research fields, 2024.

[Tra90] Manuel Trajtenberg. A penny for your quotes: Patent citations and the value of innovations. *The Rand journal of economics*, 21(1):172–187, 1990. ObjectType-Article-1; ID: cdi$_{p}roquest_{m}iscellaneous_{5}$6387780.

[VSP^{+}23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[VT00] Ellen M. Voorhees and Dawn M. Tice. The TREC-8 question answering track. In M. Gavrilidou, G. Carayannis, S. Markantonatou, S. Piperidis, and G. Stainhauer, editors, *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, Athens, Greece, May 2000. European Language Resources Association (ELRA).

[VUB^{+}23] Dave Van Veen, Cara Van Uden, Louis Blankemeier, Jean-Benoit Delbrouck, Asad Aali, Christian Bluethgen, Anuj Pareek, Malgorzata Polacin, Eduardo Pontes Reis, Anna Seehofnerova, Nidhi Rohatgi, Poonam Hosamani, William Collins, Neera Ahuja, Curtis Langlotz, Jason Hom, Sergios Gatidis, John Pauly, and Akshay Chaudhari. Clinical text summarization: Adapting large language models can outperform human experts, -10-30 2023.

[WDG20] Tobias Witt, Marcel Dumeier, and Jutta Geldermann. Combining scenario planning, energy system analysis, and multi-criteria analysis to develop and evaluate energy scenarios. *Journal of Cleaner Production*, 242:118414, 2020. ID: 271750.

[WTB^{+}22] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. Emergent abilities of large language models, 2022.

[WZS^{+}24] Yafei Wang, Zihan Zhao, Ming Shi, Jing Liu, and Zhixiong Tan. Public environmental concern, government environmental regulation and urban carbon emission reduction—analyzing the regulating role of green finance and industrial agglomeration. *Science of The Total Environment*, 924:171549, 2024. ID: 271800.

[ZWZ^{+}] Zhuosheng Zhang, Yuwei Wu, Hai Zhao, Zuchao Li, Shuailiang Zhang, and Xiang Zhou. Semantics-aware bert for language understanding.