# An AI-Driven Methodology for Patent Evaluation in the IoT Sector: Assessing Relevance and Future Impact

Lelio Campanile[2][a], Renato Zona[1][b], Antonio Perfetti[3] and Franco Rosatelli[3]
[1]*Department of Engineering, Università degli Studi della Campania, Via Roma 29, Aversa (CE), Italy*
[2]*Department of Mathematics and Physics, Università degli Studi della Campania, Caserta, Italy*
[3]*Fondazione Ricerca e Imprenditorialità, Napoli, Italy*
{*lelio.campanile, renato.zona*}*@unicampania.it, perfetti@fondazioneri.it, rosatelli@fondazioneri.it*

Keywords: Patent Evaluation, Patent Classification, Machine Learning, LLM.

Abstract: The rapid expansion of the Internet of Things has led to a surge in patent filings, creating challenges in evaluating their relevance and potential impact. Traditional patent assessment methods, relying on manual review and keyword-based searches, are increasingly inadequate for analyzing the complexity of emerging IoT technologies. In this paper, we propose an AI-driven methodology for patent evaluation that leverages Large Language Models and machine learning techniques to assess patent relevance and estimate future impact. Our framework integrates advanced Natural Language Processing techniques with structured patent metadata to establish a systematic approach to patent analysis. The methodology consists of three key components: (1) feature extraction from patent text using LLM embeddings and conventional NLP methods, (2) relevance classification and clustering to identify emerging technological trends, and (3) an initial formulation of impact estimation based on semantic similarity and citation patterns. While this study focuses primarily on defining the methodology, we include a minimal validation on a sample dataset to illustrate its feasibility and potential. The proposed approach lays the groundwork for a scalable, automated patent evaluation system, with future research directions aimed at refining impact prediction models and expanding empirical validation.



## 1 INTRODUCTION

The proliferation of the Internet of Things (IoT) has led to a marked increase in patent filings, underscoring the rapid growth of innovation win this domain. As IoT technologies continue to expand across various sectors, such as smart cities, healthcare, industrial automation and transportation systems, the global number of patents granted worldwide has seen a significant increase, that poses significant challenges for researchers, policymaker, and industry stakeholders who seek to assess their relevance, technological impact, and potential future influence. The ability to efficiently assess and evaluate patents is becoming crucial, as it supports innovation, assures fair intellectually property practices, and informs investment strategies. However, traditional workflows for patents evaluation and management shows theirs limitations in handle large volumes of documents in special way while capturing the semantic inherent within patents.

The evaluation of patents has traditionally relied on manual review by domain experts and keyword-based retrieval systems. This approach limits the scalability, rendering it ineffective for large-scale analysis. Manual evaluations and classifications, are prone to inefficiencies and inconsistencies due to their reliance on human judgment. Furthermore, keyword-based retrieval systems frequently fail to capture the subtleties and complexities of patent text, particularly within interdisciplinary domains such as IoT, where the interconnections between concepts can be highly context-dependent. Traditional impact assessments rely heavily on citation-based metrics, which may not provide a comprehensive understanding of a patent's potential influence or innovation value. This limitation renders it challenging to accurately evaluate the pertinence, applicability, and overall impact of patents within rapidly evolving fields.

The integration of Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques into patent analysis has activated advancements in automated and scalable evaluation methodologies. Recent breakthroughs in Large Language Model (LLM), in-

[a] https://orcid.org/0000-0003-4021-4137
[b] https://orcid.org/0000-0001-6718-9387



Campanile, L., Zona, R., Perfetti, A. and Rosatelli, F.
An AI-Driven Methodology for Patent Evaluation in the IoT Sector: Assessing Relevance and Future Impact.
DOI: 10.5220/0013519700003944

In *Proceedings of the 10th International Conference on Internet of Things, Big Data and Security (IoTBDS 2025)*, pages 501-508
ISBN: 978-989-758-750-4; ISSN: 2184-4976

cluding GPT, and domain-specific embeddings, have demonstrated superior performance in text comprehension, semantic similarity analysis, and automated classification tasks. These LLMs facilitate the extraction of high-dimensional, context-aware representations of patent texts. The application of LLMs to patent analysis is a novel research are with unresolved challenges. Some of these challenges include data sparsity, which hinders the development of robust models; domain adaptation, which necessitates the customization of models for specific domains; and explainability of AI-driven decisions, which raises concerns regarding the interpretability and trustworthiness of results.

In this paper, we present a novel AI-driven methodology for patent evaluation that harnesses the power of state-of-the-art LLMs and machine learning techniques to assess the relevance and impact of IoT patents. Our approach is structured around four key components.

We employ LLM-based embeddings and traditional NLP methods for feature extraction, encoding the semantic meaning of patents. This stage produces unstructured data, which is then transformed into a structured one using a structured data output framework for LLM. The next step involves applying relevance classification and clustering to identify the underlying technological patterns within the IoT patent landscape. Finally, our framework estimates the preliminary impact of each patent through a combination of citation analysis, semantic similarity, and time-series forecasting techniques. Unlike previous approaches that rely solely on keyword retrieval or citation counts, our methodology aims to provide a comprehensive, data-driven framework for understanding the significance of patents and saving in a structured way.

The primary contribution of this research is in its methodological framework, which joins conventional patent evaluation workflows with AI-driven techniques. By incorporating LLMs and deep learning models, we created a novel pipeline that not only enhances the accuracy of patent analysis but also addresses the limitations of existing approaches. Furthermore, our work presents an initial validation on a sample dataset, demonstrating the feasibility and potential of the proposed approach. Although this validation is limited in scope, it serves as a foundation for future research aimed at refining impact prediction models and expanding empirical validation to larger datasets.

The remainder of this paper is organized as follow: Section 2 provides an overview of context and background on previous work on patent analysis. Section 3 describes the proposed methodology. Section 4 presents a minimal validation study, demonstrating the applicability of our method to a selected IoT patent set. Section 5 discusses the implications of our findings, comparing our approach with existing techniques and discussing limitations. Finally, Section 6 concludes the paper and highlights future research directions.

## 2 RELATED WORKS

Historically, the patent system was established to encourage and regulate technical progress and innovation (Frumkin, 1947). However, increasing global patent applications and rapid technological development pose significant challenges to patent offices and practitioners (Krestel et al., 2021).

Automating these processes not only improves the efficiency of IP management, but also allows the NLP community to explore a promising and still under-researched field. Previous research has examined the early stages of the use of intelligent, automated methods for patent analysis (Abbas et al., 2014), the introduction of Deep Learning (DL) techniques, which have simplified some patent-related tasks (Krestel et al., 2021), and specific applications, such as patent retrieval (Shalaby and Zadrozny, 2019).

Here, we focus on two main areas: patent analysis and automatic patent text generation. While analysis is aimed at understanding and extracting relevant information from this vast knowledge base (Abbas et al., 2014), generative activities aim at the automated creation of patent content. Several studies have explored the use of Machine Learning (ML) and NLP in the patent domain, which is characterized by highly technical and legal texts (Krestel et al., 2021). In addition, recent LLMs have demonstrated extraordinary capabilities across a wide range of tasks in general domains (Min et al., 2023), suggesting their potential use in the management and editing of patent literature, an essential resource for documenting technological evolution. However, their use in patent-related activities still remains under-explored, due to the complexity of the texts and the field itself. The work of (Abbas et al., 2014) represents some of the first research devoted to patent analysis, adopting text mining and visualization methods, which opened up new perspectives for study. DL has recently gained attention for knowledge management and patents. (Krestel et al., 2021) identified eight main task categories for DL-based methods, including decision support, patent classification and retrieval, patent evaluation, prediction of emerging technologies, automatic patent text generation, litigation analysis and the use of computer vision techniques. Some studies have focused on specific aspects of patent analysis. For example,(Shalaby and Zadrozny, 2019) examined the problem of patent retrieval, i.e. the search for relevant documents, an activity that, although it may seem similar to traditional searching, has well-defined constraints and purposes. Patent literature is an important source of knowledge on technological advances and is known for its potential to inspire innovations, especially when technologies from different fields are combined.

Data science techniques must address these difficulties in order to effectively extract useful information for design and innovation (Jiang et al., 2022). The application of data science to patents not only supports theoretical and methodological design analysis, but also provides tools for innovation strategies. In particular, in the early stages of technology development, NLP can support idea generation, prediction of industry trends and the matching of problems and solutions (Just, 2024). While much research focuses on analyzing existing patent literature, there is a growing interest in the generative use of language models. NLP techniques can be exploited to make patent language more accessible, generate abstracts or even create new patent texts from structured input (Casola and Lavelli, 2022). The recent development in the fields of LLM are reported in (Chang et al., 2024). The field of patents also connects to other areas of intellectual property, such as copyright, trademarks and industrial designs, as well as interdisciplinary areas such as knowledge management and the assessment of the economic value of innovations (Aristodemou and Tietze, 2018). Moreover, the legal language of patents shares many features with other legal texts, so much so that most practitioners in the field have a strong legal background. For this reason, NLP techniques developed for law in general could be applied to patent analysis in the future (Katz et al., 2023).

Although LLMs offer significant potential for knowledge extraction and textual analysis, their application in patenting is still limited. Previous studies have employed word embeddings such as Word2Vec (Mikolov et al., 2013) and DL models such as LSTM, but transformer-based models have opened up new possibilities (Vaswani et al., 2017). Early attempts to apply LLMs to patents were based on relatively small models, such as GPT-2 (Radford, 2018), while more advanced models have not yet been explored in depth in this area. A significant obstacle is the lack of reference datasets and established metrics to assess the performance of LLMs in patent analysis. Although patent offices release raw documents, publicly available datasets for specific patent analysis tasks are still limited. Current searches focus mainly on summary sections of patents, but detailed descriptions and claims are much more relevant for automated patent analysis and generation.

# 3 PROPOSED METHODOLOGY

Patent evaluation and classification is an activity for both businesses and regulators, as it allows them to determine the technological, economic and legal value of an invention. Patent analysis can be conducted through traditional methodologies entrusted to experts in the field, as well as through the use of automated tools based on machine learning and artificial intelligence techniques. The analysis conducted by human experts is based on a set of established criteria that include novelty, inventive step and industrial applicability of the patent. These parameters are examined within the existing patent and scientific literature to assess whether the invention meets the legal requirements for patentability. In addition, the analysis considers the clarity of the claims, scope of protection and potential market coverage. Patent analysts, often with a technical and legal background, conduct extensive searches of specialized databases and compare new applications with prior patents to identify any overlaps or contrasts with the state of the art. In parallel, recent years have seen an increasing development of IT tools for patent classification and evaluation. These tools are based on NLP and ML techniques, which make it possible to extract relevant information from patent documents. For example, the use of text mining algorithms makes it possible to identify semantic relationships between patents and automatically classify inventions into technology categories defined by international standards, such as the International Patent Classification (IPC) and the Cooperative Patent Classification (CPC). Another innovative approach involves DL models and text embedding techniques to improve the ability to identify similarities between existing patents and new applications, facilitating the evaluation process. Furthermore, LLMs are showing significant potential in generating patent abstracts and translating technical language into a more accessible form. Despite these advances, the application of AI techniques to the patent field still presents some challenges. One of the main difficulties concerns the length and complexity of patent texts, which often exceed the processing capacity of traditional language models. Furthermore, patent language is characterized by highly specific terminology and complex syntactic structures, making automated interpretation difficult without a training phase on specific corpora. To address these challenges, scholars are exploring several strategies, including increasing the context window in language models and adopting Retrieval Augmented Generation (RAG) techniques, which combine information retrieval from structured sources with text generation. These developments could improve the ability to classify and evaluate patents more accurately and efficiently. In conclusion, patent evaluation and classification today relies on both the expertise of human analysts and the potential offered by automated tools based on artificial intelligence. While the former provide in-depth and contextualized interpretation, the latter offer a large-scale analysis capability, reducing the time and cost of the process. The combination of these two methodologies appears to be the most promising way to further improve the management and analysis of patent literature.

### 3.1 The Data Model

The proposal is an advanced patent analysis tool based on information extraction technology and semantic, functional and graphical analysis of patent texts and related search reports. It leverages AI algorithms, trained on a large patent domain, to provide an in-depth analysis of the individual patent, its relative positioning and reference context. The reference patent domain consists of the worldwide dataset of existing patents. The advantage is the free availability of the worldwide database of existing patents as it is in the public domain. The reference patent domain databases used as a first approximation are WIPO, ESPO, ESPACENET, GOOGLE PATENT.

Thanks to the advanced analysis of the proposal, it is possible to:

- Make strategic decisions based on concrete data;
- Identify emerging technology trends;
- Define patent policies and technology improvement drivers;
- Discover new customers, suppliers, partners or competitors;
- Identify new markets.

The proposal differs from traditional tools in that it operates on unstructured information databases, making consultation and analysis more efficient, reducing time and associated costs. The proposal for patent documents is a comparative test prior to patent document submission or parallel to the patent process for patents already filed but still reserved. This tool acts as a complement to the prior art analysis, supporting the investment decision-making process and the start of the formal patenting process. The output of the proposal is a radar diagram, which assigns a score on a scale of 1 to 5 for each of the seven patent indexes, already trained on the patent reference database:

- Feasibility, literally represents the feasibility of the patent from a realization point of view, taking into account complexity, availability of materials, the need for ad hoc structures for relitigation, and the maturity of the technological environment ;
- Multipurpose intended as the ability of the invention to have applications outside the originally chosen field ;
- Obsolescence intended as how far it is from the end of its useful life. This parameter does not coincide with the time span of the legal concession;
- Geo Context indicates how attractive a market is in the world market;
- Protection represents how protected the invention is;
- Tech Interest, represents how the subject of invention is of global interest;
- Tech Resonance, represents how much redundancy there is in the world panorama of inventions with respect to the subject matter.

These indexes assess both the intrinsic value of the patent and how it compares with the external context:

- Technological state of the art ;
- Territorial extension level ;
- Geographical distribution of comparable patent families ;
- Population of similar inventions.

A Global Score summarizes the overall patent value based on the above indexes. The proposal allows:

- Attribution of ‘weighted’ index scales (arithmetic or logarithmic);
- Categorization and ranking of patents contained in a portfolio;
- Comparison with predefined minimum thresholds or derived from repetitive analysis;

The proposal allows to compare patents within your portfolio or with previously analyzed patents. In addition, it allows you to generate a customized Global Score, based on specific weightings, to assess the overall patent value.

The proposal represents an innovative and repeatable system for patent evaluation, in pre-grant stage and in the whole patent life cycle, providing strategic support for investment decisions, technological improvement and competitive positioning. Thanks to its advanced technology, it is able to offer an in-depth and predictive view of the value and impact of patents in an increasingly dynamic and competitive market environment.

### 3.2 Feature Extraction and NLP Techniques

Our methodology is founded on conventional NLP techniques but extends them by utilizing LLMs, particularly LLaMA 70B, to obtain and examine patent data. Although traditional methods are effective for preliminary text analysis, LLMs provide the capacity to understand deeper semantic connections and contextual subtleties frequently ignored by simpler models.

A significant advancement in our methodology is the generation of structured data from LLMs generated content. Patent documents are intrinsically unstructured, rendering methodical processing challenging. We employ advanced structured output generation algorithms to transform free-text patent descriptions into a standardized format that corresponds with the main criteria outlined in previous section. Techniques such as prompt engineering, instruction tweaking, and RAG are utilized producing structured representations of patents. Recent scholarly research has proven the efficacy of these strategies in retrieving domain-specific knowledge.

To guarantee high-quality structured data, we utilize schema-based extraction, directing LLM to produce data in specified types, such as JSON or XML. This facilitates easy integration with subsequent machine learning models. Furthermore, few-shot learning enhances the dependability of structured extraction by supplying instances of accurately formatted data during inference, improving consistency and diminishing ambiguity.

Another significant feature is constraint-based extraction, wherein the LLM-generated content is verified against established rules or knowledge bases. Post-processing approaches, including regex validation, ontology-based filtering, and entity linking, are utilized to enhance structured data and rectify errors. These procedures guarantee that extracted fields, including assignees, citations, patent classification, and important claims, comply with domain-specific specifications.

Following the extraction process, the structured data is consolidated into a centralized dataset designed for ML applications. This dataset supports patent comparison and review, facilitating more rigorous study. The organized dataset has fields like technological categorization, novelty indicators, existing relationships, impact scores, and key entities such as assignees, inventors, citations, and jurisdictional coverage.

We primarily utilize zero-shot and few-shot learning paradigms, while also investigating fine-tuning as a supplementary upgrade. Fine-tuning LLaMA 70B on specialized corpora, such historical patent records and annotated datasets, may enhance the model's precision and recall. Nevertheless, considering computational limitations and ethical concerns related to proprietary data, fine-tuning is not the central emphasis of this study. We utilize adaptive prompting tactics and context-aware embeddings to enhance model efficiency without necessitating significant retraining.

We assess various approaches for the practical implementation of structured data extraction from LLMs. Among the most promising solutions are LangChain, Hugging Face's Transformers, and LLamaIndex, each providing comprehensive APIs for dealing with various LLMs and producing structured outputs.

Our methodology provides a scalable and automated framework for obtaining significant insights from patents by integrating traditional NLP pipelines with sophisticated NLP capabilities.

### 3.3 Relevance Classification and Impact Estimation

Patent relevance classification is essential for assessing the importance of a patent within the IoT domain. We plan to utilize a combination of supervised machine learning models and unsupervised clustering techniques to evaluate and classify patents according to their technological and economic worth. We will apply random forest classifiers, Support Vector Machines (SVM) , and DL models for relevance classification to assess the significance of a patent based on extracted data. These models are trained on historical patent datasets with annotated relevance ratings, ensuring strong performance in patent categorization. Furthermore, we employ hierarchical clustering to discern clusters of patents that display analogous technological attributes, thereby offering insights into nascent trends in IoT technologies.

In addition to classification, estimating the impact of a patent is crucial for forecasting its long-term effects. We combine several predictive indicators, such as citation trajectory analysis, semantic similarity to high-impact patents, and time-series forecasting of technological uptake. Citation-based methodologies analyze the historical citation trends of patents to

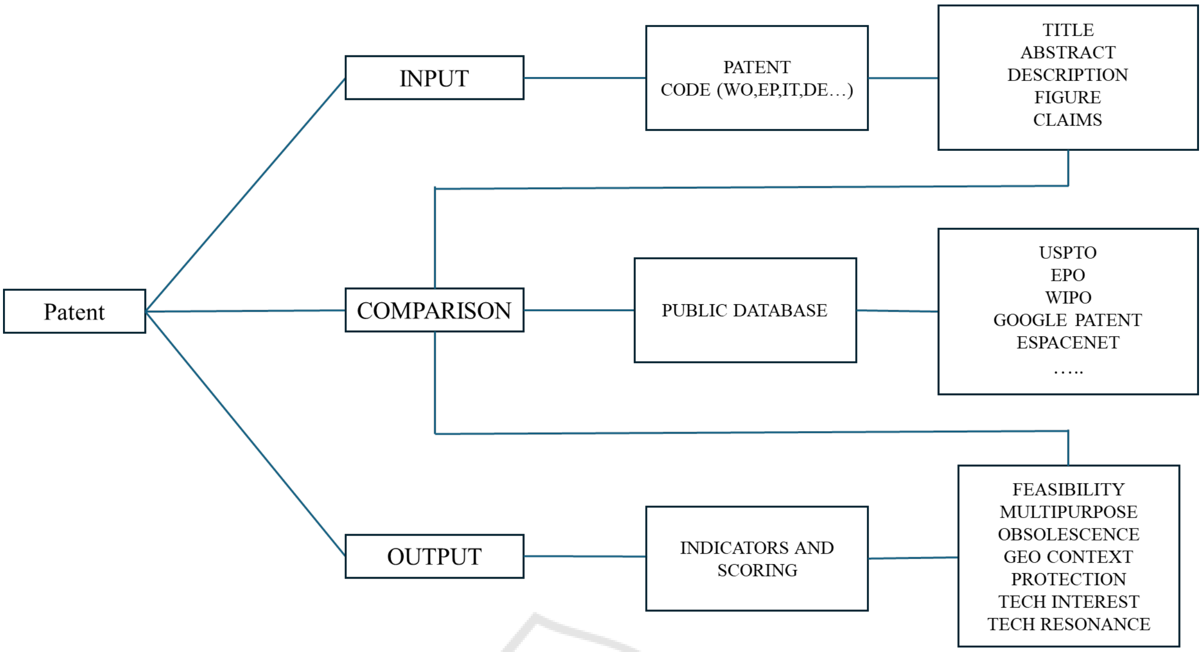



Figure 1: Flowchart of the proposal.

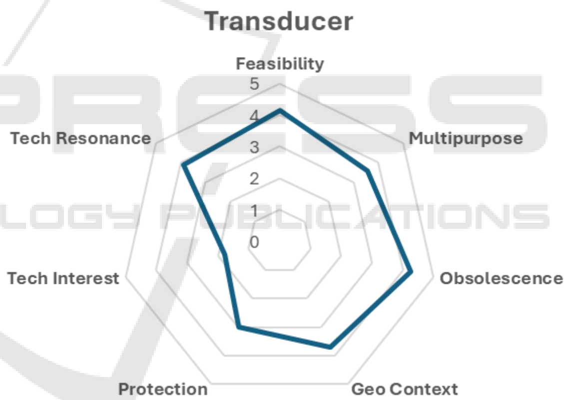
evaluate their anticipated future impact. Furthermore, embedding-based semantic similarity metrics enable the assessment of the alignment of a specific patent with historically significant patents.

Our methodology integrates classification and effect estimation, offering a thorough evaluation framework that connects traditional and AI-driven patent examination. Utilizing ML and insights from LLM, we develop a system that efficiently categorizes patents and forecasts their significance in influencing future technical progress.

## 4 PRELIMINARY VALIDATION

We evaluate the feasibility of our proposed methodology through a preliminary validation utilizing a representative case study based on patent WO2020193804A1, widely described in (Di Gennaro et al., 2024; Minutolo et al., 2020; Gennaro et al., 2022), which relates to an IoT-related invention. This validation demonstrates the efficacy of our structured extraction pipeline and classification models in assessing patent relevance and estimating potential impact.

WO2020193804A1, Transducer, is selected as a case study due to its considerable relevance in the IoT domain. The patent outlines a new method for distributed data management in IoT networks, positioning it as a suitable candidate to assess the efficacy




Global Score 2.42

Figure 2: Index scoring of the evaluated patent.

of our approach in extracting structured insights and forecasting its technological implications.

The results in terms of patent indexes are reported, see Figure 2.

The structured data extraction module is utilized on Transducer, producing a machine-readable representation that includes essential attributes such as technological classification, key claims, prior art references, and citation trajectories. Our classification model predicts the relevance score of a patent by analyzing its textual content and assessing its semantic similarity to high-impact patents within our dataset.

The evaluation of classification performance is conducted through three primary metrics: best fit similarity to established high-impact patents within an embedding space, semantic overlap with glsiot-specific patent clusters identified via hierarchical clustering, and citation-based influence indicators obtained from historical patent citation trends.

We compare the structured outputs produced by LLaMA 70B, with manually annotated patent data to validate our methodology. Factors considered include the accuracy of extracted entities, such as assignees, inventors, and technological components; the completeness of structured data fields in comparison to human-extracted insights; and the consistency of relevance classification against expert-labeled patent categories.

This case study illustrates our approach, yet it has certain limitations. The validation relies on a singular patent, necessitating expansion to a broader dataset for enhanced generalization. Estimating impact based on citations necessitates longitudinal data, which may not be completely accessible for recent patents. Furthermore, although LLMs offer structured insights, their dependability in patent-specific contexts necessitates additional benchmarking against evaluations by domain experts.

This preliminary validation indicates that our methodology effectively extracts structured information from IoT-related patents, establishing a foundation for evaluating relevance and impact. Further research is necessary to enhance the classification models, broaden the dataset, and strengthen the reliability of impact predictions. Future research will concentrate on expanding the validation scale and incorporating supplementary benchmark datasets to enhance the evaluation comprehensiveness.

# 5 DISCUSSION

The proposed AI-driven methodology for patent evaluation presents notable improvements in the analysis of intellectual property. The integration of LLM with ML techniques signifies a significant shift from traditional methods, providing enhanced accuracy, scalability, and automation. This method improves the assessment of patents through more accurate relevance classification and impact forecasting. LLM embeddings enhance the comprehension of patent texts by effectively capturing semantic relationships that keyword-based methods frequently miss. This advancement facilitates more efficient clustering of patents, thereby assisting in the identification of emerging technological trends. The integration of citation analysis and semantic similarity measures enhances impact prediction, offering a comprehensive perspective on a patent's significance over time.

This AI-driven evaluation framework provides advantages to stakeholders, including patent offices, research institutions, and corporations engaged in IoT innovation. Patent examiners can utilize the proposed method to enhance the efficiency of the patent approval process by more effectively identifying relevant prior art. Research institutions may employ the methodology to monitor technological advancements and evaluate the potential impact of specific patents on future research trajectories. Businesses can employ the system to inform decisions regarding patent investments and strategic collaborations, utilizing predicted impact metrics. AI-based evaluation serves as a transformative instrument in the management of intellectual property.

While this approach offers certain advantages, it is essential to recognize several limitations. The efficacy of LLM is contingent upon the quality and diversity of the training data utilized. LLMs trained on general corpora exhibit robust performance across various NLP tasks; however, fine-tuning on domain-specific patent texts may be necessary for optimal outcomes. Future research should investigate the creation of specialized LLMs that are specifically trained on patent data to improve accuracy and relevance. The preliminary validation study is constrained in scope due to its reliance on a limited dataset of IoT patents. A comprehensive evaluation utilizing larger and more diverse patent datasets is essential to validate the generalization of the methodology. The selection of impact prediction metrics must be refined to ensure robustness across various technological domains.

A significant challenge in AI-driven patent evaluation is the interpretability of the results. LLMs offer robust semantic representations; however, their decision-making processes frequently lack transparency. Enhancing the usability and trustworthiness of relevance classification and impact prediction can be achieved through the development of explainable AI techniques that elucidate the reasoning behind these processes. The comprehensibility and validation of AI-generated evaluations by human experts is a critical focus for future development.

# 6 CONCLUSIONS AND FUTURE WORKS

This study introduces an AI-based approach for assessing IoT-related patents, including sophisticated

NLP techniques and ML models. Our methodology offers a scalable and automated solution for patent analysis through the integration of LLM-based structured data extraction, relevance categorization, and impact calculation. The findings from our initial validation demonstrate that the methodology successfully extracts significant insights from patents, presenting a viable alternative to conventional review techniques.

The suggested methodology improves patent analysis by utilizing advanced LLMs to extract structured information, categorizing patents according to their technological relevance, and assessing their potential impact through predictive modeling. Our research illustrates that the integration of NLP approaches with machine learning allows for a more thorough evaluation of patents, aiding in the detection of significant advances in the IoT sector.

Notwithstanding the encouraging outcomes, numerous obstacles persist. The initial validation was performed on a restricted dataset, and subsequent research should aim to broaden the methodology to encompass a wider and more varied collection of patents. Furthermore, enhancing the effect estimation model through advanced time-series forecasting and the integration of external elements like market adoption patterns could improve predictive accuracy. A vital topic for enhancement is the augmentation of the interpretability of AI-driven patent evaluation, guaranteeing that the methodology yields practical insights for academics, enterprises, and policymakers.

The next steps will focus on augmenting the dataset, refining domain-specific LLM models for patent analysis, and assessing the methodology across several technological areas outside IoT. Additionally, incorporating expert feedback into the review process could enhance the trustworthiness of automated assessments. This research addresses these problems, enhancing AI-assisted intellectual property analysis and facilitating more efficient, data-driven decision-making in patent review.

## REFERENCES

Abbas, A., Zhang, L., and Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis.

Aristodemou, L. and Tietze, F. (2018). The state-of-the-art on intellectual property analytics (ipa): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (ip) data.

Casola, S. and Lavelli, A. (2022). Summarization, simplification, and generation: The case of patents.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., and Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15.

Di Gennaro, L., de Cristofaro, M., Loreto, G., Minutolo, V., Olivares, L., Zona, R., and Frunzio, G. (2024). In-situ load testing of an ancient masonry structure using fibre optics. *Structures*, 70:107567.

Frumkin, M. (1947). Early history of patents for innovation. *Transactions of the Newcomen Society*, 26.

Gennaro, L. D., Damiano, E., Cristofaro, M. D., Netti, N., Olivares, L., Zona, R., Iavazzo, L., Coscetta, A., Mirabile, M., Giarrusso, G., D'Ettore, A., and Minutolo, V. (2022). An innovative geotechnical and structural monitoring system based on the use of nsht. *Smart Materials and Structures*, 31.

Jiang, S., Sarica, S., Song, B., Hu, J., and Luo, J. (2022). Patent data for engineering design: A critical review and future directions. *Journal of Computing and Information Science in Engineering*, 22.

Just, J. (2024). Natural language processing for innovation search – reviewing an emerging non-human innovation intermediary. *Technovation*, 129.

Katz, D. M., Hartung, D., Gerlach, L., Jana, A., and Bommarito, M. J. (2023). Natural language processing in the legal domain. *SSRN Electronic Journal*.

Krestel, R., Chikkamath, R., Hewel, C., and Risch, J. (2021). A survey on deep learning for patent analysis. *World Patent Information*, 65.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. In *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*.

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heintz, I., and Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, 56.

Minutolo, V., Cerri, E., Coscetta, A., Damiano, E., Cristofaro, M. D., Gennaro, L. D., Esposito, L., Ferla, P., Mirabile, M., Olivares, L., and Zona, R. (2020). Nsht: New smart hybrid transducer for structural and geotechnical applications. *Applied Sciences (Switzerland)*, 10.

Radford, A. (2018). Improving language understanding by generative pre-training. *Homology, Homotopy and Applications*, 9.

Shalaby, W. and Zadrozny, W. (2019). Patent retrieval: a literature review.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Łukasz Kaiser, and Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 2017-December.