



# Patent intelligence in the age of AI: Unlocking strategic insights through granular classification

Guido Giuntelli<sup>a</sup>, Christian Spreafico<sup>a</sup>, Andrea Precorvi<sup>b</sup>, Davide Russo<sup>a,\*</sup>

<sup>a</sup> University of Bergamo, Viale Marconi 5, 24044, Dalmine, BG, Italy

<sup>b</sup> Tinexta Innovation Hub, Corso Mazzini 11, 42015, Correggio, RE, Italy

## ARTICLE INFO

### Keywords:

Artificial intelligence  
Patent analytics  
LLM  
Competitive intelligence  
RAG

## ABSTRACT

Patent intelligence is increasingly constrained by the volume, technical density, and linguistic variability of patent documents, which limit the feasibility of producing high-granularity, analyst-facing outputs, such as matrices, trajectories, and comparative maps, using conventional keyword- or bibliometric-based methods. This study addresses this feasibility gap by proposing an Artificial Intelligence (AI)-based framework designed to stabilize patent content for scalable strategic analysis. Methodologically, the framework transforms patent documents into structured semantic summaries optimized for large language model processing and uses these representations to support AI-driven clustering and multidimensional classification. A retrieval-augmented generation (RAG) architecture ensures that summaries and classifications remain grounded in the underlying patent corpus, enhancing reliability and transparency. The framework is demonstrated through a large-scale robotics case study focusing on Siemens, Toshiba, and Mitsubishi. Results show improved recall and precision compared with keyword-based searches for semantically ambiguous categories, illustrated through a legged-robot benchmark. The approach enables the systematic construction of technology–application matrices that reveal competitive positioning, white spaces, and cross-domain innovation patterns. It further supports cell-level deep dives via grounded clustering and evolutionary timelines, as well as technology-acceleration analysis and normalized portfolio benchmarking. An additional application to European funded-project data highlights misalignments between corporate patenting trajectories and public R&D investment priorities. Overall, the study demonstrates that AI-based semantic stabilization provides a robust and scalable foundation for advanced patent intelligence in complex technological domains.

## 1. Introduction

Contemporary patent analysis necessitates increasingly sophisticated methodologies to navigate the vast volume of information, demanding a balance between efficiency and accuracy. Beyond issues of accuracy, a central and often overlooked challenge concerns analytical feasibility. Many forms of patent intelligence routinely invoked in strategic contexts, such as high-granularity technology–application matrices, evolutionary trajectories, or multi-dimensional competitive maps, are theoretically desirable but practically unattainable at scale using conventional methods.

Although traditional bibliometric approaches, such as the study of trends or litigation counts, offer a broad overview, the identification of specific technological innovations or the assessment of market niches requires an in-depth examination of patent content. However, standard

tools such as the Cooperative Patent Classification (CPC) and International Patent Classification (IPC) systems also exhibit inherent limitations. When used in isolation, these systems can result in missed identifications or erroneous classifications of technological content (Fung et al., 2023; [1,2]).

Recent contributions demonstrate that semantic search, transformer-based models, and large language models substantially improve patent retrieval and document-level classification compared with keyword- or classification-based approaches [3–5]. However, these methods still operate primarily on raw patent texts or vectorized representations, and their analytical outputs remain largely confined to retrieved document sets or flat labels. As a result, limited attention is devoted not only to how patent content is retrieved or labeled, but to whether artificial intelligence (AI)-based approaches demonstrably enable the systematic production of complex, analyst-facing representations, such as matrices,

\* Corresponding author.

E-mail address: [davide.russo@unibg.it](mailto:davide.russo@unibg.it) (D. Russo).

<https://doi.org/10.1016/j.wpi.2026.102454>

Received 29 January 2026; Received in revised form 18 March 2026; Accepted 23 March 2026

Available online 31 March 2026

0172-2190/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

comparative maps, or technological trajectories, across large patent corpora.

Concurrently, manual analysis, while capable of high accuracy, remains economically and temporally unsustainable for large-scale patent intelligence, given the need for highly specialized expertise combining patent-specific language and domain knowledge [6]. While similar analytical outcomes may be conceivable through exhaustive manual review, such approaches do not scale across multiple dimensions and large patent corpora, effectively limiting their use to narrow or illustrative cases.

Large Language Models (LLMs) enable new forms of automated patent analysis, but their direct application to full patent texts is constrained by document length, linguistic variability, and legal-technical complexity. Prior work shows that these limitations arise not from model capability, but from the lack of intermediate representations capable of stabilizing patent content for scalable and reliable downstream analysis [7].

The core novelty lies not merely in the use of LLMs, but in the design of structured semantic summaries as compact, inspectable, and representation-agnostic artifacts that make it feasible to generate, at scale and with reliability, a wide spectrum of high-granularity analytical outputs. These summaries constitute an explicit intermediate representation layer between raw patent texts and downstream strategic analyses.

It is important to clarify the nature of the contribution proposed in this study in relation to existing AI-based patent analysis approaches. While a growing body of literature applies semantic search, embeddings, or large language models to improve patent retrieval or document-level classification, the present work does not primarily address performance gains in these isolated tasks. Instead, its contribution lies in demonstrating how structured semantic representations of patent content enable the systematic and scalable production of complex, analyst-facing analytical outputs, such as technology–application matrices, evolutionary trajectories, and comparative positioning maps, that are often discussed in strategic contexts but rarely achievable in practice at scale. In this sense, the study advances a methodological framework rather than an incremental optimization of existing retrieval or classification techniques.

Accordingly, the contributions of this study are threefold: (i) the introduction of structured semantic patent summaries as a reusable intermediate representation optimized for LLM-based analysis; (ii) the demonstration of how this representation enables scalable, high-granularity clustering and multidimensional classification; and (iii) the illustration of its analytical value through technology–application matrices, evolutionary trajectories, and comparative corporate and public R&D analyses. Rather than benchmarking isolated tasks or error rates, the study demonstrates the breadth and depth of analytical representations that become practically achievable when patent data are semantically stabilized through AI.

The limitations observed in the direct analysis of patent documents motivated the exploration of alternative approaches to maximize the efficacy of LLMs in patent intelligence. Accordingly, this study proposes an AI-based method centered on the generation of structured semantic summaries, designed as compact and semantically rich representations of patent content. By providing a more stable and manageable input for LLMs, this approach mitigates issues related to document length and linguistic complexity, enabling more scalable and reliable patent analysis.

Rather than focusing on benchmark-oriented validation or isolated performance metrics, this study adopts a workflow-level, demonstrative perspective. The objective is not to show that alternative approaches fail, but to make explicit what becomes practically achievable today through AI-based patent analysis in terms of scale, granularity, and reliability of analytical outputs.

To illustrate and stress-test the proposed framework, we present a case study on legged robots, a domain characterized by high semantic

ambiguity and significant classification noise. This case study is not intended as a narrow validation task, but as a means to demonstrate the breadth and depth of analytical representations that can be systematically generated when patent content is semantically stabilized through AI.

## 2. Literature background

This section reviews academic literature on AI-based patent analysis with the objective of identifying methodological limitations that constrain the feasibility of large-scale, high-granularity analytical outputs for strategic patent intelligence. Patent documents and commercial systems are discussed only insofar as they illustrate prevailing application trends, rather than serving as methodological references.

Recent survey work has provided a comprehensive overview of how machine learning and deep learning techniques have been applied to patent analysis tasks. In particular, Krestel et al. [8] organize the literature around datasets, text representations, and neural architectures, identifying classification and retrieval as the most mature and extensively evaluated applications. While this task-centric perspective has been instrumental in consolidating methodological practices and highlighting technical challenges in patent NLP, it primarily frames progress in terms of model performance on well-defined tasks, leaving open the question of how such models can be systematically integrated into end-to-end analytical workflows that generate analyst-facing representations for strategic patent intelligence.

Within the evolution of patent information retrieval, three distinct levels of sophistication in the application of AI can be identified.

### 2.1. Keyword-based search

The traditional paradigm of patent searching has relied primarily on keyword-based methods, often supplemented by Boolean and proximity operators. Despite their widespread adoption, such approaches exhibit well-known limitations when applied to the linguistic complexity and scale of patent literature.

A fundamental issue arises from the lexical gap and polysemy: identical technological concepts may be expressed through diverse formulations, while identical terms may carry different meanings depending on context. As a result, keyword-based searches systematically generate distorted result sets, affected by both overestimation (false positives) and underestimation (false negatives).

Similar distortions are introduced by standard classification systems such as IPC and CPC, whose coarse granularity often produces inflated and weakly specific patent sets [9,10]. Attempts to mitigate noise through restrictive queries or narrow classification codes frequently exacerbate underestimation, excluding relevant documents whose innovative content is expressed implicitly or outside predefined terminological boundaries.

More fundamentally, keyword-based approaches presuppose prior knowledge of the relevant terminology and lack the capacity to infer implicit technical meaning or integrate domain-specific contextual knowledge, limiting their effectiveness in complex patent analysis tasks [11]. While useful for preliminary filtering, these methods fail to produce stable semantic representations of patent content, thereby preventing the systematic construction of high-level analytical artifacts required for strategic patent intelligence.

### 2.2. Semantic-based search

Several recent studies focus on improving patent search and portfolio construction through semantic search and AI-assisted filtering. Sequential workflows combining semantic search with automated analysis have been proposed to support the identification of relevant state-of-the-art documents, particularly in contexts characterized by heterogeneous terminology [3]. Similarly, transformer-based models

trained on patent corpora have been shown to outperform Boolean search strategies in recall- and precision-critical retrieval tasks involving long and complex patent texts [4,5].

These advances build on semantic-based search approaches that move beyond lexical matching to capture the underlying meaning of patent texts (Bergeaud et al., 2017), enabled by Natural Language Processing and Deep Learning techniques [12]. Recent methodologies leverage AI and big data techniques for prior art searching, including vector databases, semantic similarity analysis, and neural network models, with NLP and LLMs supporting synonym extraction, interpretation of user search intent, and advanced document classification (KR10-2025-0079348; CN118170872; WO2023/195769; IN202521040236; Zhang et al., 2023).

A parallel research stream applies AI to the assessment of patentability and the identification of innovative features, either by comparing applications against deep learning models trained on specific patent classes (TR202502698) or by detecting points of innovation through large models (CN119886105). Academic studies further investigate the determinants of AI innovation and the use of machine learning to identify AI-related patents at scale [13,14], as well as applications in specific technological domains such as climate-related inventions [15] and energy systems [16].

AI-based techniques are also employed in adjacent patent-related activities, including proactive monitoring of patent infringement through the analysis of litigation data and product or service content (e.g., WO2025156057 A1), as well as the automation of patent prosecution and management processes using LLM-based frameworks trained on procedural patent knowledge ([17]; WO2020209498).

Collectively, these semantic methodologies enable increasingly granular technological classification. By processing full patent texts into embeddings and leveraging deep learning models, particularly Transformer architectures trained on large technical-scientific corpora, it becomes possible to assign highly specific labels that capture technological micro-niches and application contexts (e.g., distinguishing between Solid-State Batteries with Ceramic Electrolytes and Redox Flow Batteries).

However, optimizing these approaches to achieve stable, scalable, and reusable representations for strategic patent intelligence remains an open challenge. While semantic-based methods substantially improve precision and scalability by overcoming synonymy and polysemy, their automated application for high-granularity patent analysis remains limited, particularly with respect to the definition of granular category vocabularies and the reuse of results across multiple analytical dimensions.

Recent work explores the use of generative AI to automate specific analytical artifacts derived from patent data [18,19]. While these studies demonstrate the feasibility of producing individual outputs, they are typically output-specific. In contrast, the present study introduces a representation-agnostic intermediate layer designed to support multiple analytical views within a unified workflow.

Recent work emphasizes that reproducibility and transparency can be more limiting than raw model performance in AI-assisted patent analytics, particularly when workflows rely on partially disclosed procedures or proprietary tools [3,20]. Accordingly, this study frames reliability in operational terms, as the ability to generate inspectable and repeatable analytical outputs grounded in the patent corpus.

Despite improvements in retrieval accuracy and classification precision, semantic and embedding-based approaches largely remain document-centric. Patent content is rarely transformed into compact, reusable representations that directly support matrices, maps, or evolutionary analyses, leaving higher-level synthesis dependent on manual interpretation.

### 2.3. Reasoning-based search

Reasoning-based retrieval represents an important future direction

for query-time inference in information retrieval [21]. Nevertheless, it primarily addresses how relevance is inferred during search, rather than how patent content is transformed into stable representations for downstream strategic analysis, which remains the focus of the present work.

The central challenge at this level lies in enabling systems to reason about a query by connecting concepts through logical or common-sense chains. For instance, in response to a query such as “What solutions prevent damage to plant roots caused by reusing water from the plant saucer?”, a reasoning-based system would infer that saucer water contains dissolved salts which, when accumulated over time, can harm plant roots. Consequently, documents addressing topics such as soil salt accumulation, root health, or irrigation techniques involving leaching would be considered relevant, even in the absence of an explicit reference to saucer water reuse. In this context, relevance is not directly retrieved but inferred through reasoning.

Systems operating at this level require the integration of structured knowledge bases (e.g., knowledge graphs), common-sense reasoning capabilities, and further advances in LLM, as discussed in work on retrieval-augmented reasoning and neural reasoning architectures [22].

Taken together, prior work improves retrieval and classification performance but leaves largely unaddressed the problem of representation stability: how to transform patent texts into compact, inspectable, and reusable semantic artifacts that directly support analyst-facing analytical outputs.

### 2.4. Workflow-level and representation-oriented AI approaches in patent analytics

Beyond improvements in retrieval and document-level classification, recent research has increasingly applied contextual transformer architectures to patent corpora. Fine-tuning large pre-trained language models such as BERT for patent classification tasks has demonstrated improved semantic capture compared with traditional vector-space representations [23], building on foundational transformer architectures for contextual language understanding [24]. Comparative analyses further highlight the advantages of embedding-based feature representations over bag-of-words models in patent classification settings [25].

These advances have significantly enhanced the ability of AI systems to classify and retrieve patents based on semantic similarity. However, most embedding-based approaches produce latent vector encodings optimized for prediction accuracy rather than inspectable and reusable semantic artifacts. As a result, they remain largely document-centric and task-specific, focusing on improving classification or retrieval performance without explicitly addressing how patent content can be stabilized for systematic downstream analytical recombination.

Given the considerable length and technical density of patent documents, specialized long-document transformer architectures have also been proposed to address context limitations [26]. While extended context windows improve the processing of lengthy texts, they do not inherently resolve the challenge of generating structured and persistent intermediate representations suitable for multidimensional strategic analysis.

More broadly, the phenomenon of hallucination in large language models has been extensively documented [27], reinforcing the importance of grounding mechanisms in high-stakes domains such as patent intelligence. Retrieval-augmented approaches mitigate this risk by constraining generative outputs to retrieved source documents, thereby enhancing traceability and reliability. Nevertheless, grounding alone does not define how patent content should be structurally transformed to enable reusable analytical views.

In parallel, the technology intelligence and tech mining literature has long emphasized the importance of structured workflows for converting large-scale textual data into strategic insights [28]. Yet, the integration of large language models into end-to-end analytical pipelines remains an emerging area of research, particularly with respect to the design of

stable intermediate representation layers.

Overall, limited attention has been devoted to transforming patent texts into structured semantic representations that function as persistent and reusable intermediate layers between raw documents and downstream strategic analyses. The present study addresses this gap by introducing structured semantic summaries as compact, inspectable, and representation-agnostic analytical units, enabling the scalable construction of technology–application matrices, evolutionary trajectories, and multidimensional competitive maps within a unified AI-driven workflow.

### 3. Material and methods

This study introduces an AI-based workflow for patent intelligence designed to address the feasibility limitations identified in prior sections, namely the difficulty of systematically producing high-granularity, analyst-facing analytical outputs from large patent corpora. The proposed method is founded on the automated transformation of patent documents into structured semantic representations, conceived as an explicit intermediate layer between raw patent texts and downstream strategic analyses.

These representations provide a stable and reusable foundation for multiple analytical outputs, including multi-dimensional classifications, technology–application matrices, evolutionary trajectories, and comparative visualizations. Rather than focusing on isolated tasks such as retrieval or document-level labeling, the workflow enables the consistent generation of diverse analytical views within a unified process.

Operationally, the workflow comprises a sequence of interconnected, AI-orchestrated steps. Patent documents relevant to a given analytical objective are collected at scale and processed to generate synthetic, semantically enriched representations that condense the most technically and strategically relevant content, with particular emphasis on the inventive core expressed in the claims. These structured summaries are subsequently used as input for AI-based classification and clustering, supporting both predefined and emergent taxonomies.

This study adopts a demonstrative and positioning-oriented evaluation strategy rather than an exhaustive benchmark-based validation. The objective is not to optimize or compare individual AI components, but to make explicit which types of high-granularity patent intelligence analyses become practically feasible when patent content is stabilized through structured semantic representations.

The resulting labeled corpus supports single-label and multi-label classifications and can be reconfigured to generate diverse analytical artifacts and visualizations. The workflow operates end-to-end in a fully automated manner, leveraging patent database APIs and AI models to ensure reproducibility and analytical consistency across large patent datasets.

To enhance clarity and provide a synthetic overview of the proposed methodology, Fig. 1 illustrates the end-to-end AI-based patent intelligence workflow. The diagram highlights the sequential stages of the framework and their logical interconnections, from patent collection to

the generation of multidimensional analytical outputs.

#### 3.1. AI-based patent collection and selection

In the initial collection phase, AI models support the formulation of semantically enriched search strategies. Language models are used to expand the query space beyond fixed terminology by identifying alternative expressions and related technical concepts, thereby reducing the risk of lexical underestimation. In parallel, the analytical objective is decomposed into its core conceptual dimensions, enabling the formulation of queries at different levels of abstraction, from broad recall-oriented searches to more focused queries targeting specific aspects of the technology.

Once the initial patent pool has been collected, a screening phase is applied to isolate the documents that are genuinely relevant to the analytical scope. Screening is performed in a structured and sequential manner by applying a set of independent eligibility criteria, each corresponding to a specific conceptual requirement.

In this phase, AI systems perform a targeted reading of patent content to identify the information required to evaluate each criterion. For every document and for each criterion, a binary decision (YES/NO) is produced, allowing patents to be progressively filtered as the criteria are applied.

For instance, when the objective is to identify patents related to Quadrupedal Robots with Autonomous Navigation, three checks are applied in sequence: verification that the patent claims a robot, assessment of the presence of an autonomous navigation mode, and identification of a quadrupedal structural configuration. Only patents satisfying all criteria are retained, while others are excluded at earlier stages of the screening process.

Patent data were retrieved using Orbit Intelligence (Questel), which provides automated patent family consolidation during data extraction. The initial dataset was exported at the family level, retaining one representative publication per patent family as defined within the platform. This procedure prevents multiple counting of equivalent filings across jurisdictions and avoids duplication of related documents belonging to the same family.

By consolidating data at the family level prior to AI-based processing, the analysis ensures consistency in portfolio-level metrics and reduces potential distortions arising from jurisdictional replication of the same invention.

#### 3.2. AI-based patent summary

To address the limitations of direct analysis of full patent texts, the proposed workflow introduces an intermediate representation layer based on structured semantic summaries. Each patent document is transformed into a compact, semantically stabilized representation designed to preserve the inventive core while reducing linguistic and contextual noise.

The generation of these summaries is based on a targeted reading of patent content guided by the analytical objective. Rather than processing the entire document indiscriminately, AI models selectively extract information from sections that are most relevant for downstream analysis. In particular, patent claims are treated as a primary source, as they formally define the scope and novelty of the invention through a standardized legal structure.

Following the approach introduced by Giuntelli et al. [7], claims are analyzed by decomposing their internal structure into conceptual components, enabling the isolation of inventive elements from ancillary or contextual information. The same principle is applied, when required, to other parts of the patent text, such as sections describing materials, parameters, or application domains, to ensure that only analytically relevant content is retained.

The resulting output is a synthetic text organized into structured paragraphs corresponding to predefined informational dimensions.



Fig. 1. Overview of the AI-driven patent intelligence workflow, showing the transformation of raw patent data into structured semantic representations and strategic analytical outputs.

These summaries provide a stable, semantically rich input for clustering, classification, and cross-sectional analyses.

The specific structure adopted to organize these summaries and capture the inventive core is detailed in Table 1.

To mitigate generation errors, summaries are subject to automated consistency checks, including verification of structural completeness and alignment with claim-level content. Outputs that do not meet predefined structural or semantic criteria are flagged for regeneration or excluded from subsequent analysis.

As with any generative process, some variability in wording may arise across repeated summary generation runs. However, because downstream analyses operate on constrained informational dimensions rather than surface-level phrasing, moderate linguistic variation does not materially affect clustering outcomes or high-level analytical representations.

Patent documents in the dataset originate from multiple jurisdictions and are therefore available in different original languages. In the present implementation, all structured semantic summaries are generated from the English versions of the patent texts obtained through the machine translation system integrated into the patent database platform used for retrieval. Summarization prompts, clustering operations, and classification tasks are consistently performed in English to ensure linguistic uniformity across the corpus. No third-party abstracts are used; summaries are generated directly from the translated full text, with particular emphasis on the claims section. Because all documents are processed in a single target language, no additional cross-lingual terminology normalization is applied prior to clustering or classification. The extension of the framework to fully multilingual processing pipelines constitutes a potential direction for future methodological investigation.

### 3.3. AI-based patent clustering

Clustering is used in the proposed workflow as an inductive mechanism to identify latent thematic structures and emerging technological groupings within large patent corpora. However, the direct application of clustering techniques to full patent texts is constrained by the length, linguistic complexity, and computational cost associated with processing complete documents.

To address these constraints, clustering is performed on the structured semantic summaries described in Section 3.2, rather than on raw patent texts. These summaries provide compact yet semantically rich representations that concentrate analytically relevant information while substantially reducing input size.

Using structured summaries as input enables the simultaneous analysis of larger document sets and allows clustering to be explicitly grounded in predefined informational dimensions (e.g., inventive solution, innovated component, system composition). This shifts clustering from surface-level term co-occurrence toward the identification of technological and functional affinities across patents.

The resulting clusters constitute an initial, data-driven taxonomy that can be reviewed and refined by a domain expert. Expert intervention may include merging closely related clusters, subdividing overly broad groupings, or adjusting cluster labels to improve interpretability and alignment with the analytical objectives.

**Table 1**  
Definition and content of the AI generated structured paragraphs.

Summary Structure	Definition
<b>Innovation:</b>	Describe the innovative solution proposed.
<b>Innovated component:</b>	Specifies the component, element or method that incorporates or is acted upon by the innovation.
<b>System:</b>	List all components of the system described in the claims.

### 3.4. AI-based patent classification

Whereas clustering supports the inductive identification of latent thematic structures within the patent corpus, classification represents the subsequent step aimed at formalizing these structures into consistent and reusable analytical frameworks. In the proposed workflow, classification translates emergent clusters into structured labels that can be systematically applied across the corpus and recombined along multiple analytical dimensions. Accordingly, a multidimensional classification approach is adopted, encompassing several strategic axes:

- **Technological domain and industry sector:** patents are organized based on their core technology (e.g., AI, biotechnology, renewable energy, semiconductors, telecommunications, industrial automation). This perspective helps identify areas of concentration and pinpoints where an enterprise or research center holds a competitive advantage.
- **Functionality and practical application:** patents are categorized according to their function or the problem addressed (e.g., energy efficiency improvement, diagnostic methods, advanced user interfaces, innovative payment systems). This classification mode clarifies how the technology is intended for practical use and what market opportunities it opens.
- **Development stage or technological maturity:** the advancement of an invention is assessed across different levels, from conceptual or theoretical stages to prototyping, commercialization, and market consolidation. Positioning a patent along this trajectory provides insights into growth potential, investment return horizons, and technological risk.
- **Nature of innovation:** a distinction is made between incremental innovations, which refine existing technologies, and radical or disruptive innovations, which can transform markets or create new ones. Identifying this distinction is relevant for assessing the originality and long-term competitive sustainability of a portfolio.
- **Strategic and legal perspective:** the motivations behind patent filing are analyzed (e.g., legal defense, licensing, blocking competitors, revenue generation, role in technology standards), and the validity and strength of the intellectual property right are evaluated. This includes tracking disputes, litigation, licensing agreements, and identifying standard-essential patents.
- **Competitive and external context:** patents filed by key competitors are monitored for benchmarking and identifying strategic trajectories. Analyzing patents that reflect emerging market trends helps capture early signals of transformation. The geographical dimension, based on filing jurisdictions, reveals international expansion strategies.
- **Internal portfolio management:** patents are classified as active or expired, and those to be maintained, abandoned, or licensed are identified, assessing the alignment of the intellectual property strategy with organizational goals.

The proposed methodology supports the systematic assignment of granular classification labels across large patent corpora. Classification is implemented as an iterative and multi-level process, closely integrated with clustering, allowing identified groupings to be formalized, refined, or further subdivided as the analytical scope evolves.

Through this interaction between clustering and classification, the patent corpus is progressively structured into a hierarchically labeled representation that can be reconfigured across multiple analytical dimensions. Such a representation provides the methodological basis for constructing complex analytical views from patent data, as illustrated in Table 2.

## 4. Case study, results and discussion

To validate the proposed methodology in a realistic and

**Table 2**

Overview of the strategic analysis modules and visualizations enabled by the AI-driven methodology.

Analysis type	Description	Critical challenges	AI contribution	Suggested visualization
Technology-application matrix	Cross-references classification data to understand the end-uses of developed technologies and their integration into application systems, highlighting the role of foundational technologies.	A poor assignment of categories (technology or application) can distort the view of linkages, compromising the understanding of technical content or the estimation of served markets.	AI models enable consistent and scalable dual classification, automating the categorization process for both technologies and applications, thus improving analytical precision.	<b>Fractal matrix (heatmap)</b>
Technology-application timeline	Reconstructs the technological development over time for a specific technology/application combination, identifying key players and trajectories.	Selecting irrelevant or out-of-scope patents can distort historical analysis and provide a skewed view of technological development.	AI assists in the accurate selection of relevant patents through retrieval-augmented systems and supports in-depth analysis of the technologies employed, ensuring more accurate and strategically useful results.	<b>Cell drill-down: timeline</b>
Technology positioning matrix	Identifies emerging, established, or promising technologies, allowing for observation of the investment focus of a company (or comparison across companies).	Incorrect technological classification can lead to misplacement on the chart, altering the perception of potential or maturity and compromising strategic decisions.	AI correctly classifies technologies and enables their accurate positioning on the chart. The use of advanced language models and semantic analysis reinforces the consistency and precision of the classification process.	<b>XY scatter plot</b>
Competitive technology benchmarking	Shows which technology categories each company is investing in for comparative analysis and evaluation of IP policy.	Errors in patent classification can distort comparative analyses and strategic decisions, leading to incorrect conclusions about competitor strengths.	AI enables scalable and consistent comparison of competitors by automatically normalizing patent classifications, clustering technologies across heterogeneous portfolios, and reducing bias introduced by manual categorization.	<b>Comparative technology analysis: AI for competitor intelligence</b>
Future trends mind map	Offers a strategic vision of future technological evolutions in a given field, useful for technological forecasting and R&D planning.	Overestimating the weight of a trend or misplacing it chronologically can lead to flawed strategies, excessive investment in uncertain directions, or unfounded expectations.	AI supports this process through web crawling, automated content extraction, and semantic analysis, allowing for the identification and filtering of the most relevant and reliable sources for a more robust and objective overview.	<b>Future trends mind map</b>
EU project investment clustering	Understands how companies allocate funds in EU projects, identifying the technologies on which efforts and financial commitments are concentrated.	A discrepancy between the stated project goal and the technology actually developed can distort the interpretation of investments, producing a misleading representation.	AI analyzes project texts and classifies projects based on the actual technologies involved, not just the stated objectives. The use of NLP and semantic classification ensures a more precise association between projects and technologies.	<b>European project investment clustering</b>

heterogeneous industrial setting, the patent analysis focused on companies that primarily employ robotics as an operational and enabling technology within broader industrial processes, rather than on specialized robot manufacturers. On this basis, three multinational corporations were selected: Siemens, Toshiba, and Mitsubishi.

The robotics domain was selected as the empirical case study for methodological and contextual reasons. Robotics represents a technologically heterogeneous and semantically complex field, integrating hardware, software, sensing systems, control architectures, and application-specific adaptations across multiple industrial sectors. Such heterogeneity generates significant classification ambiguity and terminological variability, making it particularly suitable for stress-testing structured semantic representations and multidimensional analytical frameworks.

In addition, robotics is currently characterized by accelerated innovation dynamics and growing strategic relevance across industrial ecosystems, driven by advances in artificial intelligence, autonomous systems, and smart manufacturing. Its rapid technological convergence and cross-sector deployment make it an especially appropriate domain for evaluating the robustness, scalability, and discriminative capacity of the proposed intermediate representation layer under realistic and industrially relevant analytical conditions.

A preliminary exploration of more than 800,000 robotics-related patent documents published since 2015 and filtered at the claims level was conducted. This analysis showed that the largest patent volumes are concentrated in two main groups: academic institutions, particularly Chinese universities such as Beijing University of Technology, Harbin Institute of Technology, and Zhejiang University; and companies with a highly specialized technological focus, such as State Grid Corporation of

China in the drone sector or DJI in aerial platforms. While these actors exhibit substantial patenting activity, their portfolios are typically concentrated on a single robot type or a narrow application domain, limiting their suitability for validating a methodology intended to support multi-dimensional and cross-domain analysis.

In contrast, Siemens, Toshiba, and Mitsubishi present more balanced and representative patent portfolios. Although they are not the largest patent holders in absolute terms (with 1,020, 475, and 1238 patents, respectively), they operate as advanced users, and in part developers, of robotic technologies across a wide range of industrial sectors. Their portfolios span multiple robot categories, including mobile robots, drones, manipulators, autonomous vehicles, and underwater robots, resulting in a high degree of technological heterogeneity. This diversity makes them particularly suitable for assessing the ability of the proposed framework to capture and structure complex innovation patterns in real-world industrial contexts. To support the analysis, this heterogeneity was formalized through an internal classification map of robot types, which served as the reference structure for defining analytical categories throughout the case study.

To implement the proposed methodology at scale and support the empirical positioning analysis, the study relied on a combination of AI-based tools. A general-purpose large language model was used to support prompt formulation and the initial expansion of search strategies during the patent collection phase.

The core methodological steps, namely structured semantic summarization, multidimensional classification, and the recombination of analytical dimensions into positioning matrices and evolutionary trajectories, are independent of any specific software environment. In the present study, these steps were operationalized using Omnia, a software

platform developed by Tinexta Innovation Hub in collaboration with the University of Bergamo, which provides an integrated implementation of a retrieval-augmented generation (RAG) architecture [29].

In this implementation, RAG is used to ensure that all classifications and summaries remain grounded in the underlying patent corpus. Following the RAG paradigm introduced by Lewis et al. [30], each analytical query is supported by the retrieval of relevant patent documents, which act as an explicit contextual constraint for the generative component.

The role of Omnia in this study is therefore illustrative and operational rather than methodological: it enables the scalable execution of the proposed framework but does not define its analytical logic or determine its validity. Equivalent workflows could be instantiated using alternative retrieval-augmented pipelines, combining open-source language models, vector databases, and standard clustering algorithms, provided that document grounding and structured prompting are preserved.

A key operational characteristic of this architecture is its strict dependence on retrieved patent content for query-specific responses. When information relevant to a given query is absent from the patent corpus, the system is designed to explicitly indicate this absence rather than generating speculative or unsupported content. This behavior substantially reduces the risk of hallucinations, a known limitation of general-purpose LLMs, and enhances the reliability of the analytical outputs in a patent intelligence context.

Fig. 2 provides an operational view of the Omnia interface, illustrating the interaction between user prompting, document retrieval, and grounded response generation.

The integration of these components is implemented as a fully automated, end-to-end workflow. Patents are retrieved through patent database APIs and processed using large language models, including Gemini, to generate the structured semantic summaries described in Section 3. These summaries are subsequently provided as input to a proprietary AI-based classification engine, which assigns labels to patents based on structured classification prompts.

Classification is performed with respect to predefined conceptual classes derived through knowledge engineering. In parallel, an additional AI module based on unsupervised and generative techniques is applied to large sets of summaries to perform clustering. This module supports the identification of latent thematic groupings and the emergence of new conceptual classes directly from the data, enabling the taxonomy to be iteratively refined.

The resulting corpus is consistently labeled and supports both single-label and multi-label classification schemes. In the latter case, multiple

classes may be associated with the same patent, optionally accompanied by relevance scores reflecting their relative importance. The level of granularity achieved through this process exceeds that obtainable through conventional keyword-based searches or CPC-based classification alone, as illustrated in the following analyses.

#### 4.1. AI-based patent search and selection

Regarding data acquisition, the absence of direct AI integration within patent database query interfaces required the initial collection phase to rely on keyword-based search. This stage was explicitly designed to maximize recall rather than precision. Accordingly, the query was intentionally broad, yielding a large and heterogeneous initial document set intended to function as a pre-filter for subsequent AI-based content analysis.

Using this strategy, an initial pool of 2726 patent documents was retrieved with the following query:

```
(SIEMENS OR TOSHIBA OR MITSUBISHI)/PA/FPAN AND EPRD ≥ 2015 AND (ROBOTIC + OR ROBOT? OR ((AUTONOMOUS OR UN_MANN + OR OPERATOR_LESS) 2D (ROVER OR ROVERS OR VEHICL + OR AERIAL + OR FLY + OR AUTOMOBIL+)) OR (REMOT+_CONTROL+ 2D (VEHICL + OR AERIAL + OR FLY+)) OR DRONE? OR AUTO+_GUID+_VEHICL+)/ETI/EAB/OBJ/ECLM.
```

The query did not include CPC classifications because, in the specific case under analysis, their inclusion would not have provided a significant advantage. An analysis of the most frequent CPC classes in the retrieved patent pool showed that the only highly specific class (B25J-009/16, related to program-controlled manipulators) was already covered by the keywords used in the query. In contrast, other frequent CPC classes (e.g., G05D-001/02 and G05D-001/00) refer to general control systems applicable to many technologies beyond robotics, while subclasses related to vision-controlled systems also span multiple non-robotic applications. Therefore, including these CPC codes would likely have increased noise without significantly improving recall.

Keyword-based and classification-based searches remain essential, auditable components of professional patent analysis. The present framework is designed to augment, not replace, these methods by applying AI-driven semantic stabilization downstream of conventional Boolean and classification-based retrieval, thereby enhancing analytical depth while maintaining traceability.

Once the initial corpus was established, a cleaning phase was applied to isolate patents genuinely relevant to the analytical scope. Keyword-based filtering, while suitable for preliminary delimitation, proved insufficient due to well-known semantic limitations related to

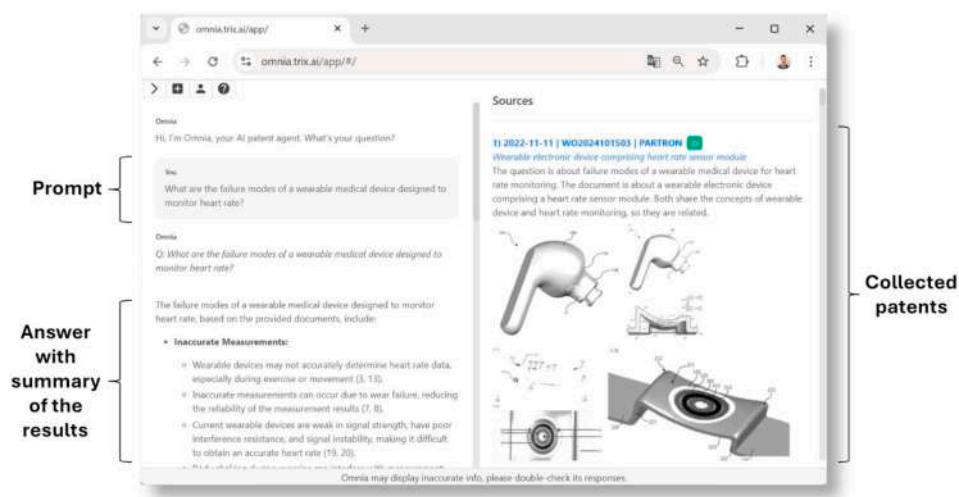


Fig. 2. Screenshot of the Omnia RAG-based platform, illustrating the user prompt, the generated summary, and the retrieved source patents used for grounding the response.

contextual ambiguity, synonymy, and the presence of misleading terminology within patent texts.

To address these limitations, an AI-driven, concept-based classification system was applied to the entire initial document set. The system performed a content-level analysis of each patent, enabling relevance assessment beyond surface terminology.

To validate the robustness of the AI-based classification across the dataset, a comparative benchmark was conducted using the Legged Robot category as an illustrative test case. This category was selected to highlight the performance gap between the proposed framework and traditional keyword-based strategies. A precise conceptual definition of Legged Robot was provided within the classification prompt to guide the AI model.

From the initial pool, the AI identified 66 patents as relevant to this category. These documents were subsequently reviewed manually to establish a ground truth. Table 3 compares the results obtained through AI-based classification with those produced by different keyword-based approaches, including variations in query breadth, search fields (full text versus title and claims), and the use of CPC codes, reporting the corresponding recall and precision metrics.

As shown in Table 3, the most accurate keyword-based strategy corresponds to search 5 conducted on the full text. Nevertheless, this approach achieves a recall only slightly above 50% and remains affected by relatively low precision. To clarify the causes of this suboptimal performance and to illustrate the sources of the performance gap relative to the AI-based approach, representative examples of false negatives and false positives are discussed below.

Several relevant documents are missed by keyword-based searches and therefore constitute false negatives. For example, patent EP4495423 (Siemens Gamesa, 2023) is not retrieved by queries 1, 2, or 3 reported in Table 3, as the concept of a legged robot is expressed through indirect formulations such as “one or more legs” or “legged driven,” rather than through explicit terminology. The absence of a standardized label

**Table 3**  
Comparison between keyword-based search queries and AI concept search.

Query	Where was it searched	Patents found	Cross-reference with the 66 identified by AI and manually verified		
			Found equal	Recall	Precision
1 (legged 1d robot+)	Full text	10	7	7/66	7/10
	Title + claims	3	2	2/66	2/3
2 B62D-057/032	CPC	7	7	7/66	7/7
3 ((LEGG + 1D ROBOT+) or quadrupedal + or (dog 2w robot+) or (four_leg? 2d robot+) or 4_leg?_robot + or hexapod+)	Full text	10	7	7/66	7/10
	Title + claims	3	2	2/66	2/3
4 (ROBOT? \$ (MULTI_LEGG + OR FOUR_LEG + OR 4_LEG + OR QUADRUPEDAL + OR LEGS)) *The \$ operator identifies a search for terms in the same sentence	Full text	45	25	25/66	25/45
	Title + claims	13	8	8/66	8/13
5 (ROBOT? and (MULTI_LEGG + OR FOUR_LEG + OR 4_LEG + OR QUADRUPEDAL + OR LEGS))/ETI/ECLM/DESC	Full text	97	36	36/66	36/97
	Title + claims	16	9	9/66	9/16

prevents keyword queries from identifying the document as relevant. A second type of false negative is illustrated by WO2018/074101 (Mitsubishi Electric, 2017), which describes a humanoid robot using exclusively anatomical terminology without explicitly referring to legged locomotion. In this case, the implicit nature of the locomotion system renders the document invisible to conventional keyword-based searches unless additional terms such as humanoid or bipedal robot are included a priori.

Keyword-based approaches also generate numerous false positives. For instance, patent US20250162147 (Mitsubishi, 2023) is incorrectly retrieved as a legged robot because references to wheeled robots appear in the background section, even though the claimed invention exclusively concerns a legged robotic system. Another false positive is WO2022/097256 (Mitsubishi, 2020), in which the term “legs” refers to mechanical anchoring or connecting elements rather than to locomotion. Similar semantic ambiguity is observed in WO2023/084280 (Siemens, 2021), where the proximity of expressions such as “robot arm” and “table leg” leads even refined or proximity-based queries to infer a spurious conceptual relationship.

Taken together, these examples show that query refinement alone is insufficient to resolve semantic ambiguity in patent texts. This confirms that keyword-based approaches cannot reliably discriminate relevant documents in complex technological domains without deeper, content-level semantic analysis.

#### 4.2. AI-based patent summary

Once the relevant patent set was established, structured semantic summaries were generated to support subsequent analyses. Depending on the analytical objective, different summary formats were produced via AI, each tailored to extract specific types of information.

Table 4 provides an example of this summary format, illustrating how the novelty of patent JP2024097565 is synthesized into structured components.

#### 4.3. AI-based patent clustering

Based on the structured summaries, a clustering analysis was performed, resulting in the identification of four primary criteria, defined as follows:

- **Robot type:** describes the physical form and architecture of the robot, including the locomotion system (e.g., wheels, tracks, legs), manipulation capabilities (e.g., articulated arms, grippers, end-effectors), and overall configuration (e.g., humanoid, vehicular, mobile platform). Hybrid systems combining multiple modalities are also identified.

**Table 4**  
Example of a structured synthetic summary.

Summary Structure	Example (Summary of Patent JP2024097565)
<b>Innovation</b>	The innovation lies in an operation management system designed to handle a mixed fleet of unmanned guided vehicles (AGVs) that use different methods for position acquisition. [...]
<b>Innovated component</b>	The innovated component is the operation management system itself, specifically the setting unit within the management device that assigns the right of passage. This setting unit considers factors such as the order of arrival at a pre-set approach standby position, the route of the AGV, its positional relationship with other AGVs, the type of AGV, and the type of article it carries.
<b>System</b>	The system comprises the following components: <ul style="list-style-type: none"> <li>• A management device.</li> <li>• First unmanned guided vehicles (AGVs) that estimate their position indirectly. [...]</li> <li>• A notification unit.</li> </ul>

- **Function:** identifies the task performed by the robot, including the activity type, the target of the activity (e.g., inspection of wind turbine blades), and the operational mode (autonomous, collaborative, or remote-controlled).
- **Operational field:** defines the intended environment and industrial sector of use, specifying both the application domain (e.g., automotive, logistics, aerospace, healthcare) and the physical characteristics of the operating context (e.g., structured or unstructured, hazardous, remote, public).
- **Sensors:** identifies the sensing technologies integrated in the system, specifying the sensor types (e.g., cameras, LiDAR, IMU, GPS, force or tactile sensors) and their functional role in perception or state estimation.

A key feature of the proposed framework is its recursive application. Once a macro-cluster is identified, the corresponding subset of patents can be re-processed to explore finer-grained technological distinctions. To illustrate this capability, a second-level clustering was applied to the Legged Robot category.

This analysis enabled the identification of specific robot sub-architectures beyond generic classification labels, distinguishing between humanoids, bipeds, quadrupeds, hexapods, and octopods, as well as more specialized configurations such as one-legged robots, adaptive suspension vehicles, and hybrid systems combining legs with wheels or tracks. The resulting hierarchical organization, structured along robot type, function, and sensor dimensions, is shown in Fig. 3.

#### 4.4. AI-based patent classification

Once the taxonomy was defined, the entire patent corpus was systematically classified. In this stage, the AI system operated as a classifier rather than as a discovery tool, assigning to each structured summary the most appropriate labels based on the established criteria.

This automated labeling process transformed the patent corpus into a structured, high-dimensional dataset indexed according to the dynamic taxonomy. The resulting representation provides the basis for the cross-referencing and quantitative analyses presented in the following sections.

##### 4.4.1. Technological Intersection Analysis

Technological intersection analysis examines how technologies are distributed and combined across application domains by jointly considering the technological content of patents and their intended application context. In this study, each patent is classified along two primary dimensions: robot type and functional application.

Systematic cross-referencing of these dimensions produces a technology–application matrix that highlights patterns of technological convergence and diffusion across end-use domains. This representation supports the identification of cross-cutting enabling technologies as well as application areas characterized by higher concentrations of patenting activity.

Constructing a matrix of this granularity through manual analysis would require extensive document-by-document review and substantial domain expertise, making it impractical for large patent corpora. The use of AI-generated structured summaries enables the consistent construction of such matrices even for numerically significant datasets.

Fig. 4 presents the resulting Technological Intersection Matrix for Siemens, Toshiba, and Mitsubishi. The matrix is visualized as a high-density heatmap obtained by cross-referencing Robot Type (columns) and Functional Application (rows), providing a comparative view of the companies' technological positioning across these dimensions.

As shown in Fig. 3, the Technological Intersection Matrix reveals distinct competitive profiles that are not readily captured by aggregate bibliometric indicators. The matrix highlights a convergence in the Handling and Manipulation domain, while simultaneously exposing divergent technological strategies. Siemens shows a strong

concentration in Robot Arms, consistent with a focus on industrial automation, whereas Toshiba and Mitsubishi display a more distributed presence involving Mobile Robots and Drones.

In addition, Fig. 3 enables the identification of white spaces and niche overlaps across technology–application intersections. One example is the use of Legged Robots for Inspection tasks, a specific intersection that suggests an emerging area of R&D activity shared by the analyzed companies.

To illustrate the scalability and multidimensional flexibility of the proposed framework, the intersection analysis was subsequently extended beyond functional tasks to two additional strategic axes: Industrial Application Fields and Sensor Technologies. This reconfiguration was performed without manual re-indexing, relying solely on the recombination of AI-generated classification labels, and provides complementary views of the competitive landscape.

Fig. 5 maps Robot Type against Industrial Sector, revealing distinct market footprints across the analyzed companies.

The heatmap shown in Fig. 4 highlights a clear strategic differentiation: Siemens shows a dual concentration in Industrial Manufacturing and Hospital/Medical sectors, primarily driven by extensive use of robotic arms. In contrast, Mitsubishi exhibits a strong presence in Warehouse/Logistics and the Automotive sector, supported by a broad deployment of mobile, wheeled, and aerial robotic platforms. This analysis illustrates how the intersection framework differentiates between robotic functionality and operational deployment contexts, enabling a more nuanced interpretation of companies' technological positioning.

Fig. 6 correlates Robot Type with Sensor Type, highlighting the hardware configurations that enable different forms of autonomy.

The analysis shown in Fig. 5 reveals consistent technological patterns across companies. Siemens' emphasis on manipulation is associated with a high concentration of Force/Torque Sensors, which are critical for precise physical interaction. In contrast, Mitsubishi's portfolio shows a strong integration of LiDAR and GPS technologies, reflecting its strategic focus on autonomous mobility and outdoor navigation.

This functional and structural decomposition provides a granular view of how sensing technologies support different robotic capabilities. By examining intersections at this level of detail, it becomes possible to identify enabling technologies and structural dependencies that are difficult to capture through aggregate or document-level analysis. At the level of individual matrix cells, smaller and semantically coherent subsets of patents emerge, characterized by high internal similarity. These focused document pools can be further analyzed using RAG-based tools to automatically extract and synthesize their technical content, enabling targeted deep-dive analyses.

For Fig. 7, a specific matrix cell corresponding to Siemens patents related to wheeled robot handling applications was selected for further analysis. This subset of patents was queried in Omnia using the question: "How many patents does Siemens have related to wheeled robot handling applications?". The system returned a set of automatically clustered documents, which were subsequently visualized using an evolutionary timeline. This representation supports expert-level patent analysis by enabling a detailed reconstruction of technological developments within a narrowly defined technology–application intersection.

The timeline shown in Fig. 6 reveals a coherent innovation trajectory within the Siemens patent portfolio, showing how an initially broad functional focus progressively evolves into more specialized applications. The sequence highlights a strategic oscillation in R&D activities, moving from early Medical Procedure Assistance systems to Industrial Mobile Platforms and Robotic Packing solutions, before reapplying these matured mobility technologies to advanced medical intervention robots in 2024.

This longitudinal reconstruction makes it possible to trace specific threads of technological development within a narrowly defined technology–application intersection, which would be difficult to identify

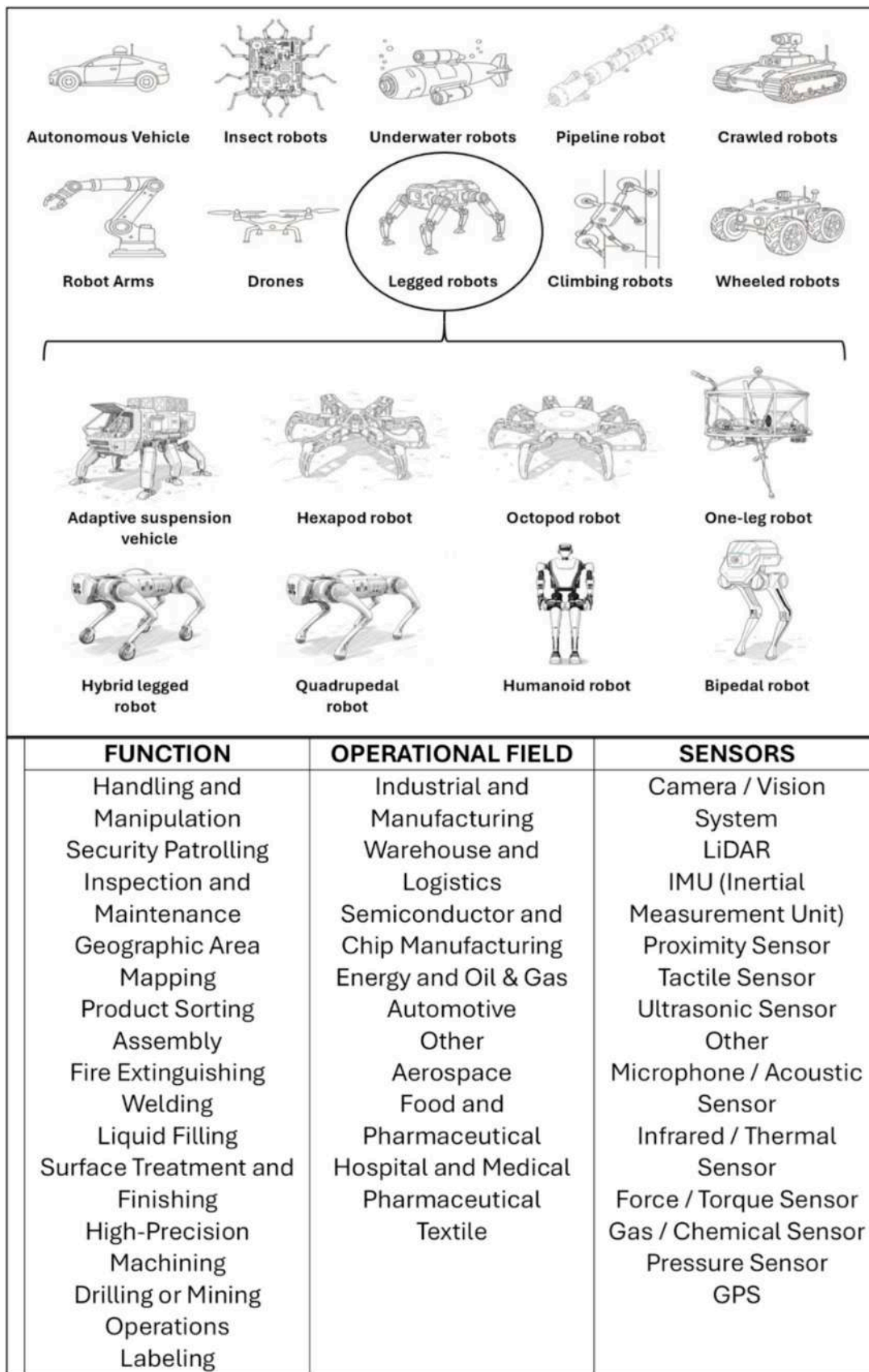


Fig. 3. Visual taxonomy of the robotics categories and functional criteria identified through the AI clustering process.

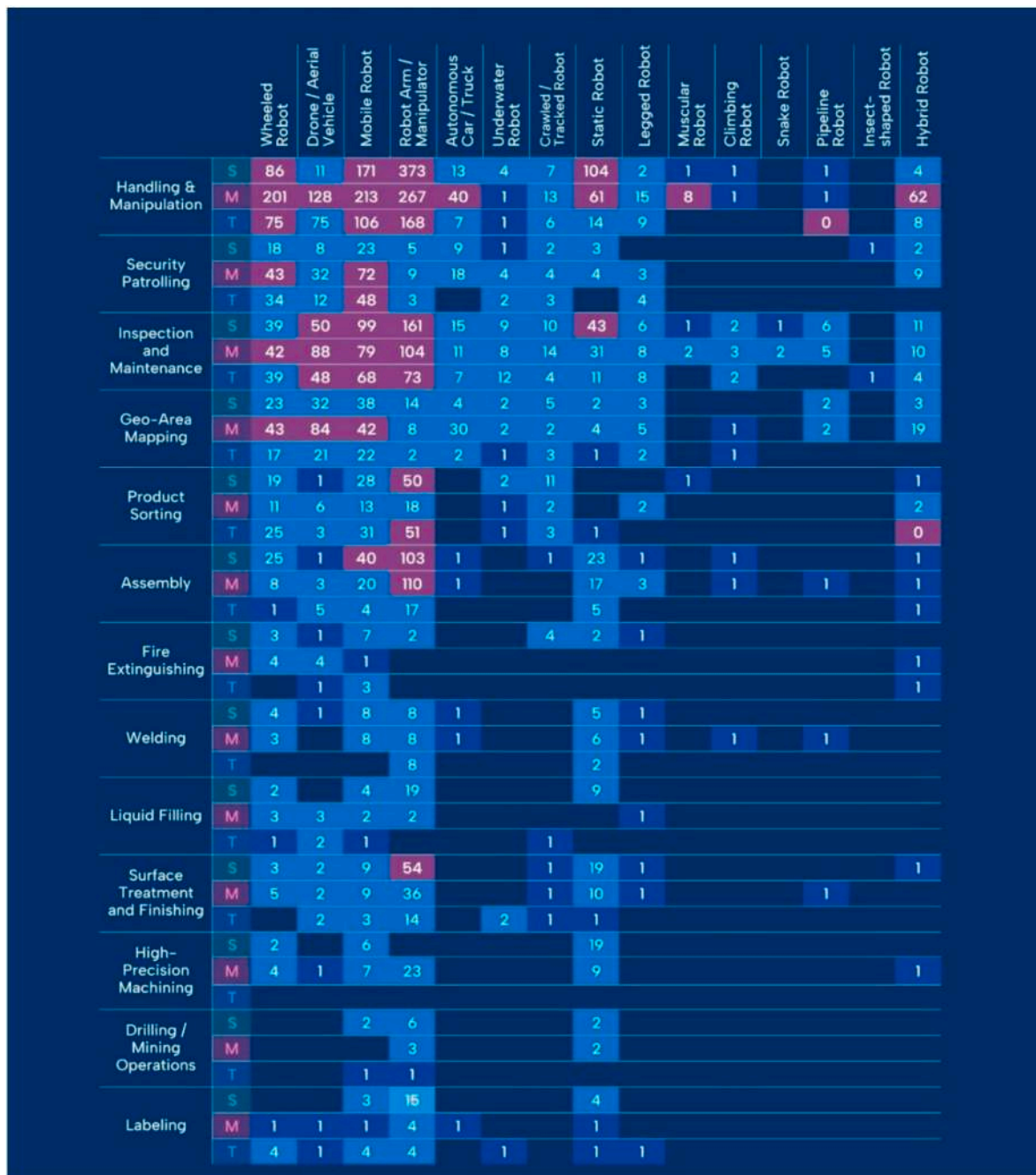


Fig. 4. Technological intersection analysis between the robot type (columns) and the functional application (rows).

through aggregate portfolio-level analyses.

This level of analysis makes it possible to isolate patenting activity within a narrowly defined technology–application intersection and to track the emergence of early solutions as well as subsequent phases of technological refinement. Such reconstructions are difficult to obtain through conventional keyword-based searches, which are often affected by the inclusion of contextually irrelevant documents or by imprecise temporal attribution. In the present framework, these limitations are mitigated through the use of a retrieval-augmented architecture that grounds clustering and interpretation in the underlying patent corpus. By constraining analysis to semantically coherent document sets and preserving their chronological context, the approach supports a more accurate reconstruction of technological development trajectories.

Analyses of this kind, both at the level of the overall matrix and at the level of individual cells, are difficult to perform at scale without

automated support. Manual approaches typically rely on keyword searches, which are effective for direct and unambiguous functions (e.g., welding or fire extinguishing) but are less suitable for abstract or cross-cutting categories such as inspection and maintenance, which can be expressed through heterogeneous terminology and applied in diverse contexts.

In other cases, such as distinguishing between a wheeled robot and an autonomous vehicle, the technological function can only be determined through detailed content-level analysis of the patent. Combined with the volume of documents involved, these requirements make exhaustive manual classification impractical for constructing complete and statistically robust matrices.

By leveraging semantic representations and clustering mechanisms, the proposed framework supports the identification of functional relationships that are not captured by simple term co-occurrence. This

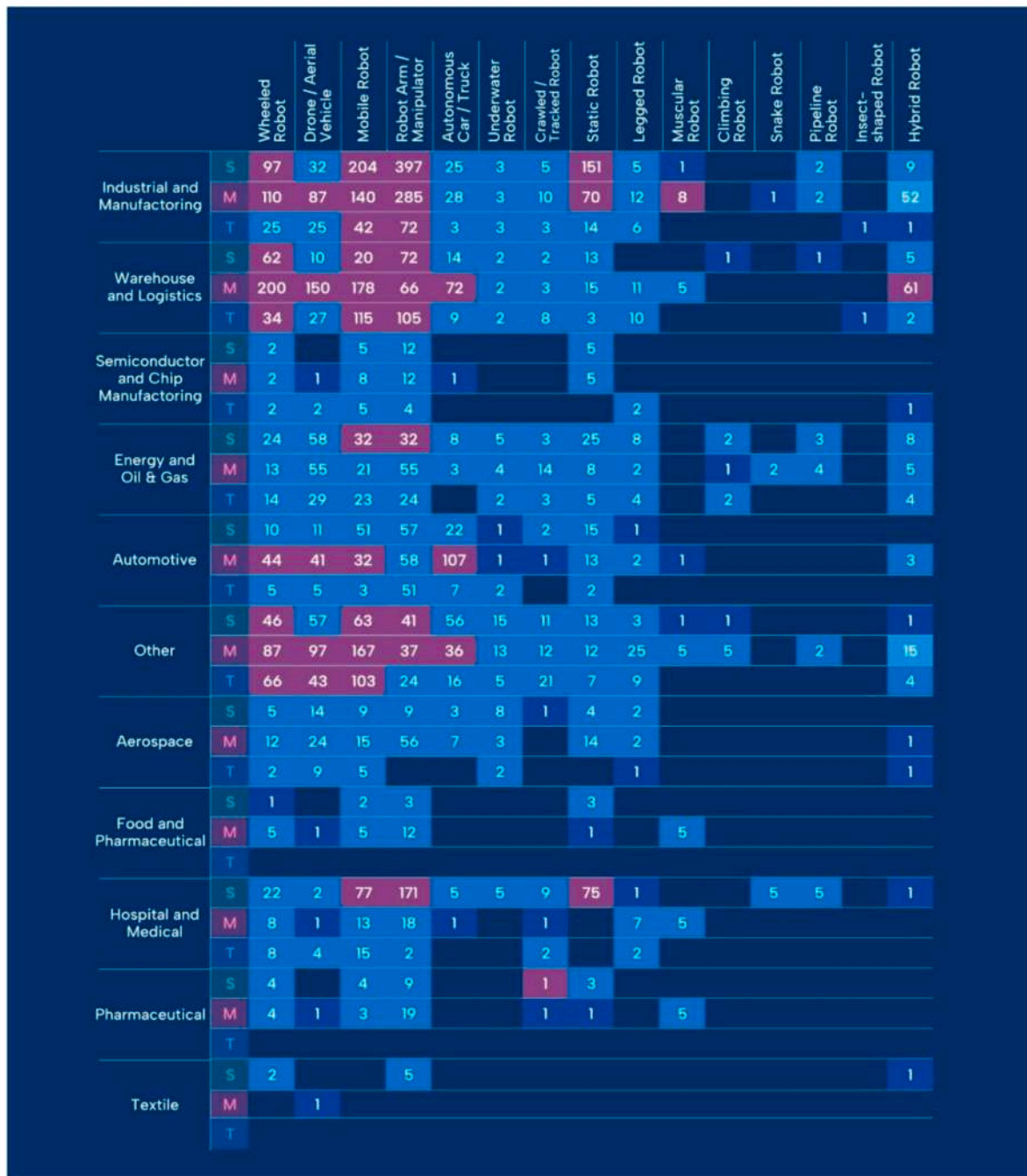


Fig. 5. Technological intersection analysis by the Robot Type (columns) and the Industrial Sector (lines).

enables a more systematic exploration of technological intersections, facilitating the analysis of convergence patterns and the identification of emerging application areas.

4.4.2. Technology acceleration analysis

The dynamic analysis of technological trajectories combines information on patent volumes with temporal patterns of patenting activity, enabling the joint assessment of technology diffusion and innovation momentum. Rather than relying on static patent counts alone, this perspective incorporates the acceleration of filings over time, providing a more informative view of technological and strategic interest. Based on this approach, technologies can be distinguished into consolidated technologies (high patent volumes with stable growth), emerging technologies (limited patent bases with high acceleration), mature-in-

growth technologies (large portfolios with sustained innovation activity), and marginal or declining technologies (low patent volumes and weak growth).

This classification supports the interpretation of corporate strategies in terms of investment focus and technological positioning. Approaches based solely on keyword frequency or term co-occurrence are prone to distortion, as they do not reliably capture firms' engagement with specific technologies. By contrast, semantic classification enables a more consistent attribution of patents to technological categories, improving the robustness of trajectory-based analyses.

Fig. 8 presents the technology acceleration matrix for Mitsubishi. The analysis reveals a clear differentiation across robot categories. Consolidated technologies such as manipulator robots, mobile robots, and drones exhibit stable or slightly declining patenting activity,

		Wheeled Robot	Drone / Aerial Vehicle	Mobile Robot	Robot Arm / Manipulator	Autonomous Car / Truck	Underwater Robot	Crawled / Tracked Robot	Static Robot	Legged Robot	Muscular Robot	Climbing Robot	Snake Robot	Pipeline Robot	Insect-shaped Robot	Hybrid Robot
Camera/ Vision System	S	104	37	179	215	58	10	19	91	9		5		7		9
	M	222	203	247	123	137	9	22	54	25	5	4	2	4		91
	T	113	71	139	131	12	7	12	15	19		2			1	6
LiDAR	S	45	50	61	15	25	7	10	11	1				2		2
	M	124	93	55	45	78	5	5	14	6	1	2	1	1		54
	T	33	22	66	52	3	5	4	2	6						2
IMU (Inertial Measurement Unit)	S	5	11	17	22	5	2	4	4	5			5	2		
	M	42	31	44	15	41	7	1	6	7	2			1		15
	T	18	12	24	5	4	6	5		3						1
Proximity Sensor	S	55	7	45	70	5	1	2	21							2
	M	28	21	60	47	18	1	7	12	3	2					7
	T	31	15	60	48	2		5	5	2						
Tactile Sensor	S	8	1	15	25			1	14	1						
	M	8	4	11	58	4		1	4	6	7			1		1
	T	10	3	12	20		1	5	2	2						1
Ultrasonic Sensor	S	15	15	23	52	10	2	4	12	2		1		4		1
	M	48	41	24	15	41	1	1	21	5		2				22
	T	54	7	42	14	1	1	6	1	2						2
Other	S	65	40	141	202	55	2	11	73	7	1	2	5	6		6
	M	202	162	223	154	140	3	19	47	21	7	5		5		31
	T	512	65	150	103	13	11	15	11	17		2			1	6
Microphone/ Acoustic Sensor	S	14	2	26	22	3	2		7	5				2		1
	M	17	24	23	22	10	5	5	3	9				1		9
	T	21	12	29	3	1	1	1	1	7		2			1	2
Infrared/ Thermal Sensor	S	23	21	53	44	6	5	7	15	5		1		4		5
	M	56	44	40	26	14	2	2	4	3		1	1	1		17
	T	51	21	45	12	5	2	7	1	3					1	5
Force/Torque Sensor	S	4	1	17	215		5		15	2			5			1
	M	55	5	55	512	22		1	14	7	4		2	2		5
	T	11	7	12	45	1		2	4	5						1
Gas/ Chemical Sensor	S	1	4	4	4	2			1							
	M	4	3	3	2			1	1	1	1					
	T	3	4	3	3			2								
Pressure Sensor	S					1										
	M															
	T															
GPS	S	12	23	17	2	17	2	5	4	2				2		2
	M	102	150	23	21	107	6	4	5	7						55
	T	24	22	56	3	7	3	1		4						5

Fig. 6. Technological intersection analysis by the Robot Type (columns) and Sensor Type (lines).

indicating signs of maturation. In contrast, categories including legged, muscular, underwater, and climbing robots display limited patent volumes but high acceleration rates, positioning them as emerging technologies with increasing strategic relevance. To ensure a clear and consistent allocation of each class across the four areas, a relative comparison criterion was adopted. Specifically, for each company and for each individual class, a ratio was computed between the average observed over the last three years and the overall average of the entire pool. This metric enabled a coherent positioning of each class within its respective area, thereby providing a more structured and comparable representation of the data. The choice of a three-year window reflects two methodological considerations: first, it allows sufficient time for patent applications to overcome the typical confidentiality period and become publicly available; second, it provides a balanced temporal

horizon for capturing acceleration dynamics, avoiding both excessive short-term volatility and the dilution of recent trends associated with longer time spans.

4.4.3. Competitive intelligence analysis

Comparative analysis of patenting strategies examines how companies distribute their patenting activity across different technological domains. Absolute patent counts, however, are not directly comparable across firms with different sizes and R&D capacities. To address this limitation, patent counts are normalized by expressing each category as a percentage of the company's total patent portfolio. This normalization provides a relative benchmark that highlights not only the volume of patenting activity, but more importantly the technological areas in which each company concentrates its innovation efforts.

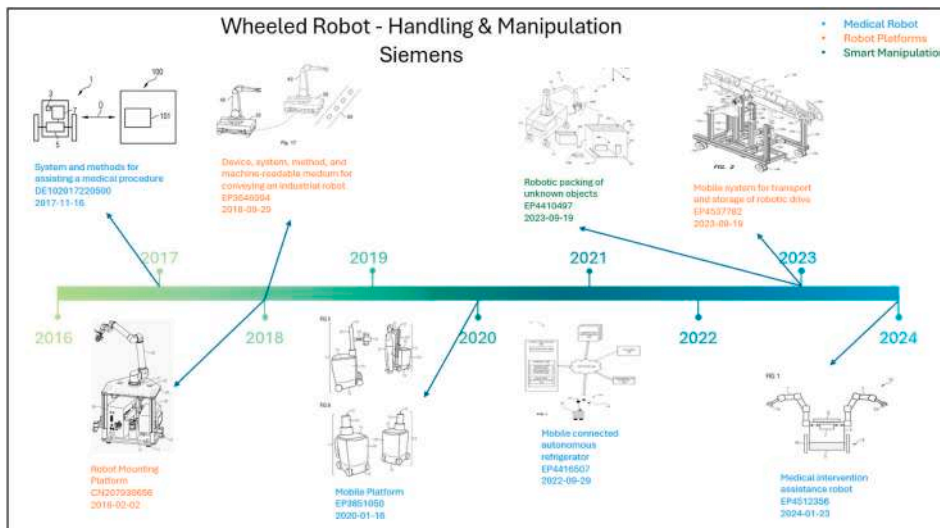


Fig. 7. Deep-dive timeline for the Siemens intersection, Wheeled Robot and Handling and Manipulation.

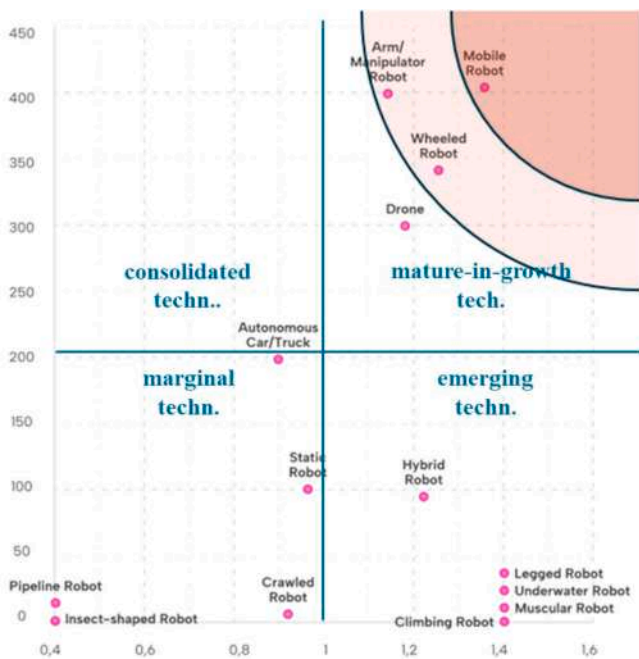


Fig. 8. Technology acceleration matrix for Mitsubishi in Robot Types.

This perspective enables the identification of priority technological domains for each company, highlights areas of specialization, and supports the assessment of the internal coherence of their strategic positioning. Firms with substantially different patent volumes may exhibit similar relative investment patterns, while others may follow clearly divergent technological trajectories.

The drone category illustrates this effect. Although Mitsubishi holds nearly three times as many drone-related patents in absolute terms (296) as Toshiba (96) and Siemens (87), the relative weight of this technology within its overall portfolio is comparable to that of Toshiba. A similar pattern is observed for wheeled robots: Mitsubishi owns more than twice the number of patents (333) held by Toshiba (152) and Siemens (172), yet these account for a smaller share of its portfolio (16%) than Toshiba (19%) and only slightly more than Siemens (11%). These results indicate that absolute patent counts alone may obscure meaningful similarities or differences in companies' strategic allocation of innovation efforts.

This type of analysis depends on the accurate classification of patents

into predefined technological categories and on the subsequent normalization of patent counts. Classification errors may distort relative benchmarks and affect the interpretation of results, particularly when comparing heterogeneous patent portfolios. The use of AI-based semantic classification supports a more consistent and scalable attribution of patents to technological categories, reducing such distortions when applied to large document sets.

Fig. 9 compares the normalized distribution of patent portfolios across the identified categories.

The analysis of Fig. 8 shows that all three companies concentrate a substantial share of their patenting activity in a common set of robot categories, including wheeled robots, drones, mobile robots, and robot arms. Within this shared focus, distinct strategic differences emerge.

Siemens exhibits a stronger relative concentration in robotic arms and static robots compared with the other companies, reflecting its emphasis on industrial automation. Mitsubishi displays a generally balanced distribution across categories, with localized peaks in autonomous vehicles and hybrid robots. Toshiba, by contrast, shows greater versatility, allocating its patenting activity more evenly between autonomous vehicles and mobile robots. Overall, the analysis indicates that, despite these differences, all three companies primarily concentrate their innovation efforts on robotic technologies traditionally associated with logistics and manufacturing applications.

#### 4.4.4. Future trends analysis

The early identification of emerging technological trends requires a forward-looking analysis of industrial innovation dynamics. Patents constitute a particularly suitable data source for this purpose, as they provide legally grounded and systematically structured representations of technological development that often anticipate future innovation directions.

In this study, future trend analysis was conducted using a RAG framework to ensure that semantic interpretation and temporal comparisons remained grounded in patent source data. Omnia was used to retrieve and aggregate a comprehensive corpus of patents relevant to a given technological domain or sector, which served as the explicit knowledge base for subsequent analysis.

The patent corpus was then segmented into discrete time windows (e.g., yearly intervals), enabling longitudinal comparison of patenting activity rather than a static snapshot. Within each time window, AI-based clustering was applied to the retrieved patents to identify technological macro-categories based on semantic similarity. The use of RAG allowed clustering and interpretation to remain anchored to the content of the retrieved patent documents, mitigating distortions arising from

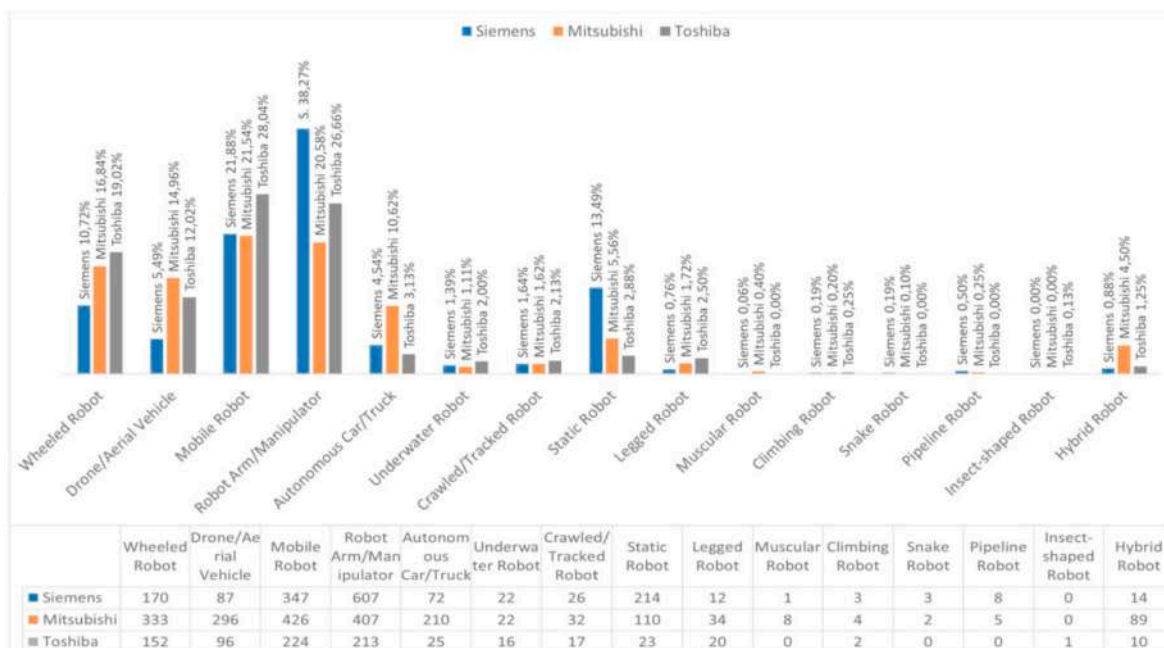


Fig. 9. Comparative analysis of how Siemens, Mitsubishi and Toshiba distribute their own patent portfolios (in %) across different types of robots.

heterogeneous or evolving terminology.

Semantic clusters identified in the most recent time window (e.g., 2024) were subsequently compared with those from earlier periods (e.g., 2023) through an iterative temporal analysis. This process supports the identification of newly emerging technological clusters as well as the evolution and refinement of previously established ones.

The analysis aimed to identify three complementary phenomena: the emergence of new patent clusters corresponding to previously unobserved technological themes; the introduction of novel elements within established clusters, reflecting the evolution of consolidated technologies; and significant changes in the relative relevance of existing clusters over time.

To this end, novelty detection focused on identifying clusters characterized by accelerating patenting activity, which may indicate increasing R&D intensity and potential future impact. The use of patents as the primary data source strengthens the robustness of this analysis, as patent filings are directly linked to strategic and economic investment decisions and are less volatile than alternative information sources such as general-audience reports or online commentary.

By combining retrieval-augmented search, semantic clustering, and temporal comparison, the proposed pipeline supports a systematic and reproducible analysis of technological evolution across different time horizons, while reducing the subjectivity inherent in large-scale manual assessments.

Fig. 10 summarizes the emerging technological clusters identified through the temporal semantic analysis of patent data.

As illustrated in Fig. 10, analysis of the patent landscape over the last three years indicates a shift in strategic focus across major technological areas. Patent activity is increasingly concentrated in Nano Robotics and Smart Materials, which together account for approximately 7000 patent documents, compared with about 4000 patents collectively associated with Autonomous Robots, Human–Robot Interaction, and Swarm Robotics.

This distribution suggests that while the latter fields remain active, they exhibit characteristics of technological maturity and commercialization, with a comparatively lower rate of radical innovation. In contrast, the growing patenting intensity in nanotechnology and advanced materials points to a strategic reorientation toward foundational hardware technologies that may underpin future generations of

robotic systems.

The framework also supports fine-grained exploration within individual clusters. As illustrated by the Smart Materials breakout in Fig. 9, clusters can be expanded to reveal specific technological directions, including self-healing polymers and bio-hybrid materials, providing a detailed view of the solution spaces driving these emerging domains.

More broadly, this analysis highlights the importance of semantic disambiguation in longitudinal technology assessment. In domains where technological evolution is not accompanied by stable or updated terminology, advanced AI-based methods enable consistent classification of patents over time by distinguishing between genuinely novel concepts, incremental refinements of existing solutions, and references to obsolete or contextually unrelated technologies.

The importance of semantic disambiguation can be illustrated by considering the concept of an autonomous robot and its evolution over time. A comprehensive trend analysis requires not only tracking the distribution of solutions across time (temporal trends), but also assessing whether and how the underlying technological concept itself has changed (technological trends).

Searches based on simple keywords or coarse classification schemes can retrieve very large patent sets, for example, more than 43,500 documents using the full-text keyword “autonomous robot”. However, closer inspection reveals that while the term has remained largely unchanged, the notion of autonomy has undergone substantial transformation over the past decade. This persistence of terminology despite conceptual change introduces significant ambiguity and renders traditional classification approaches increasingly inadequate.

Given the scale of patent data involved, manual analysis of such conceptual shifts is impractical. Addressing this challenge requires AI-based methods that move beyond surface-level relevance checks and explicitly model the interaction between the operating environment, enabling technologies, and the target concept, allowing technological meaning to be interpreted consistently over time.

A retrospective analysis highlights substantial conceptual heterogeneity associated with the term “autonomy” over time. Earlier interpretations of reactive autonomy typically referred to systems constrained to pre-programmed paths and rule-based behavior within structured environments. In contrast, contemporary notions of adaptive and intelligent autonomy describe AI-driven systems capable of dynamic

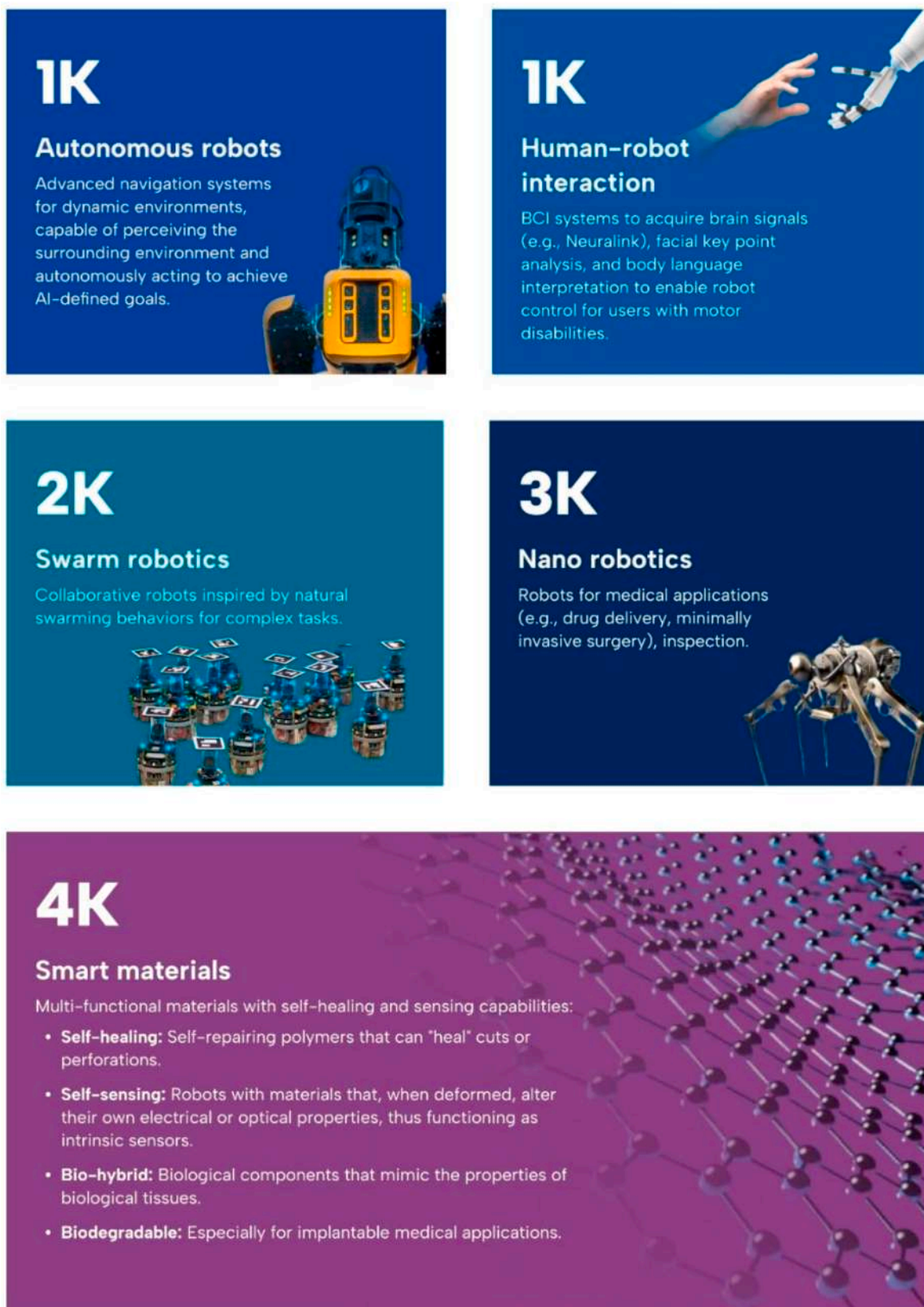


Fig. 10. Snapshot of emerging technological trends in robotics.

planning, multimodal perception, and real-time decision-making in connected and unstructured contexts.

This conceptual evolution is clearly reflected in patent literature. Earlier filings, such as US20150185027, characterize autonomy primarily as rigid waypoint navigation, whereas more recent documents, including JP2020116669, describe advanced capabilities such as multi-robot coordination and autonomous task identification. As a result, identical terminology may mask substantial functional differences,

limiting the effectiveness of keyword-based analysis and underscoring the need for AI-based approaches that interpret patent content at a semantic level to accurately trace technological development.

#### 4.4.5. Analysis of financial data from European project fundings portal

The AI-based methodology previously illustrated for patent analysis can be extended to other relevant documentary sources. One notable application concerns the analysis of European project funding, which

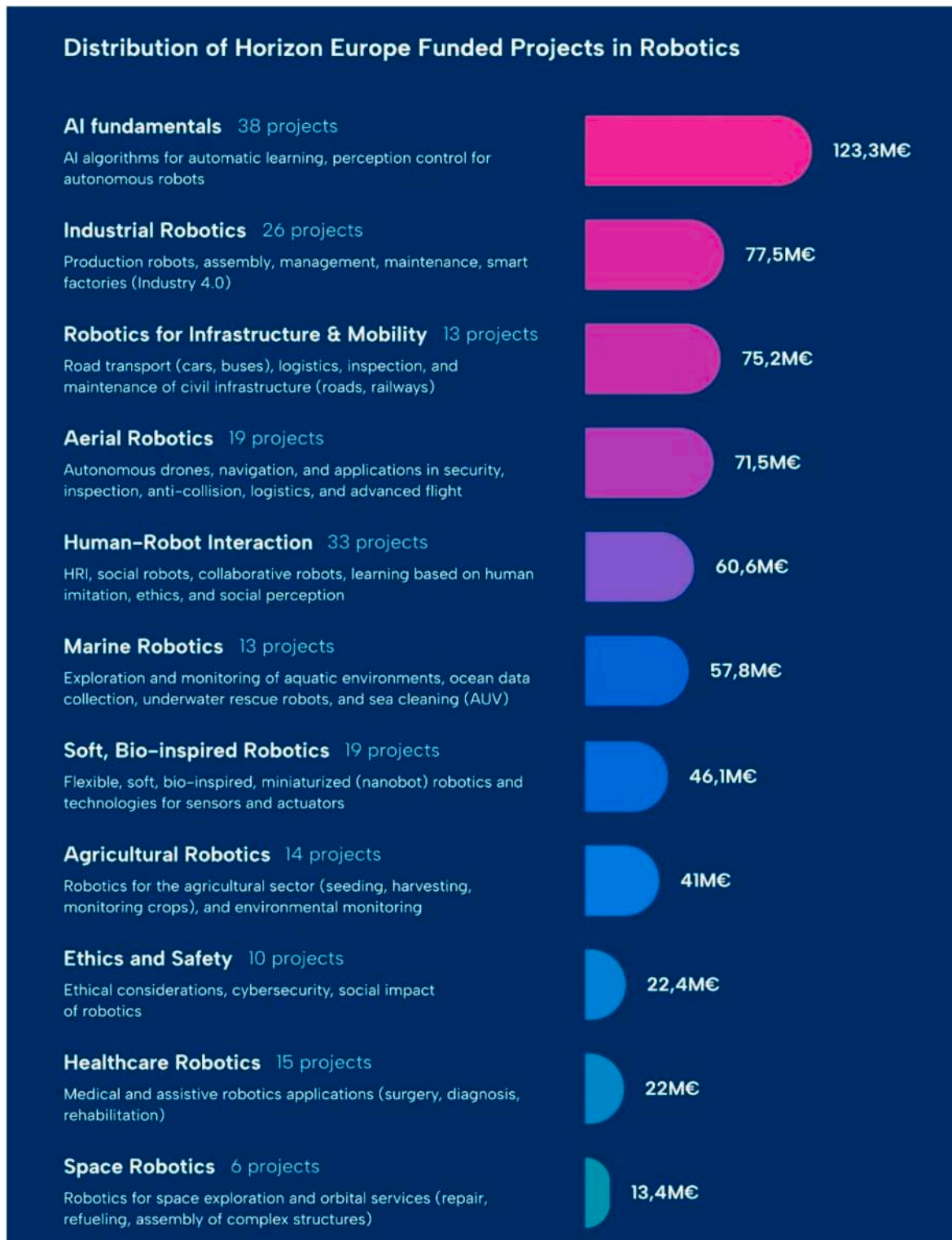


Fig. 11. Distribution of 2024–2025 European project funding (206 projects) across the identified robotics sub-sectors.

provides a complementary perspective to patent-based evaluations by enabling the investigation of investment trends and strategic priorities at the European level. Despite their apparent geographical focus, European projects attract participation from companies and organizations worldwide, reflecting global interest in the funded technologies.

Although distinct from patent data, this information complements the preceding analyses by enabling a comparison between emerging technological macro-categories and areas receiving European funding. Such comparisons support the assessment of alignment or divergence between corporate innovation strategies and public funding priorities. The application of the same NLP and semantic clustering techniques allows structured analysis of funded projects, contributing to a strategic overview of investment flows within the European innovation landscape.

Historically, a key limitation in the analysis of funded research projects has been the potential mismatch between stated project objectives and the technologies actually developed. Classifications based solely on project calls or administrative categories are often too coarse to accurately reflect technological content, which can lead to distorted interpretations of funding priorities. Applying AI-based NLP and semantic classification techniques to full project documentation enables a more precise association between projects and reference technologies, thereby reducing classification errors.

Using this approach, 1735 European projects involving robotic systems were identified to provide insight into public investment patterns in this domain. For analytical clarity, the discussion focuses on a subset of 206 projects funded in the 2024–2025 period, which were classified according to their technological context to identify emerging innovation clusters. Fig. 11 presents a visual overview of the resulting distribution of European project funding across the identified robotics sub-sectors.

As shown in Fig. 10, the distribution of European research funding exhibits a distinct strategic orientation, with the largest share of financial resources allocated to AI fundamentals and industrial robotics. A cross-analysis with the patent trends reported in Fig. 7 reveals a divergence between corporate R&D trajectories and public investment priorities.

While private patenting activity has increasingly shifted toward emerging physical technologies, European funding remains largely concentrated on areas such as autonomous robotics, human-robot interaction, and swarm robotics, which patent-based evidence suggests are relatively mature technological domains. Targeted investments are also observed in soft and bio-inspired robotics, corresponding to developments in smart materials, and in healthcare robotics, which relates to advances in nano robotics; however, these areas account for a comparatively smaller share of total funding.

Overall, this comparison indicates that private-sector R&D is currently oriented toward the development of next-generation enabling technologies, whereas public funding continues to prioritize the application, deployment, and integration of established robotic solutions within industrial and societal contexts.

#### 4.5. Overall discussion and limitations

The high degree of heterogeneity observed across classification dimensions would render large-scale manual patent analysis extremely complex, requiring the involvement of experts from multiple disciplinary backgrounds to adequately cover each analytical axis. By contrast, the AI-based framework enables the use of a unified and specialized prompting strategy to generate diverse classifications in a coherent and scalable manner, while preserving methodological consistency and reducing the coordination burden associated with multi-expert analysis.

A further limitation of this study concerns the scope of quantitative validation. The proposed framework is not evaluated through an exhaustive, benchmark-oriented comparison across multiple technological domains. Such an evaluation would require the definition of

domain-specific ground truths and extensive manual annotation, which would shift the focus of the study toward performance optimization rather than methodological positioning. Instead, the present work adopts a demonstrative perspective, using a limited quantitative benchmark to illustrate typical failure modes of conventional approaches and to support the broader objective of showing which types of patent intelligence analyses become practically feasible through AI-based semantic stabilization.

Despite these advantages, the proposed AI-based approach is subject to several limitations that require careful consideration. A primary concern is the risk of classification errors, particularly those arising from hallucinations in language models. In highly specialized patent contexts, such errors can be subtle and difficult to detect, as they may produce outputs that appear plausible while remaining semantically incorrect.

Unlike classification in stable knowledge domains, where categories are well defined and universally recognized (e.g., distinguishing between animal species), patent analysis presents a higher degree of conceptual complexity. Technological taxonomies, particularly in emerging fields, are dynamic and continuously evolving, and the terminology used to describe innovations is often weakly standardized, qualitative, or metaphorical.

As a result, AI systems are more likely to encounter ambiguity when distinguishing between closely related and informally defined concepts, such as “snake robots” and “crawling robots”, whose descriptions may be discursive and context-dependent, than when classifying mature technologies such as a two-stroke engine, which are characterized by a consolidated and objective technical vocabulary. This semantic fluidity increases the level of contextual reasoning required and, consequently, the risk of generating plausible but inaccurate inferences.

For these reasons, domain-expert supervision constitutes an essential component of the proposed framework. Expert involvement serves a dual role. Upstream, domain experts contribute to the definition of taxonomies and to prompt design, ensuring alignment between automated classifications and the intended analytical and strategic objectives. Downstream, domain experts are responsible for the critical validation of results.

This validation does not require exhaustive review of individual data points. Instead, experts focus on evaluating aggregated patterns, investigating anomalies, and interpreting outcomes in light of domain-specific knowledge, particularly when results appear counter-intuitive. Within this collaborative model, AI systems provide the computational capacity required to analyze patent data at scale, while human experts contribute contextual understanding, critical judgment, and strategic interpretation. The objective of the framework is therefore not to replace human analysts, but to augment their capabilities by transforming large volumes of patent data into actionable strategic intelligence.

The large-scale deployment of AI-based patent analysis workflows is subject to economic and computational constraints. Access to high-quality patent databases involves non-trivial costs: free services such as Espacenet are not designed for high-volume automated retrieval, whereas commercial databases provide reliable APIs at subscription costs that increase with data-intensive usage. These constraints become particularly relevant when analyses require full-text access and iterative processing.

A further limitation concerns the computational cost associated with the use of advanced language models. Open-source systems, e.g., LLaMA and Mistral, may not consistently provide the reliability required to interpret complex patent language, making commercial models preferable despite their higher associated costs. Given the length and technical density of patent documents, token consumption increases rapidly, constraining the economic feasibility of large-scale processing.

In addition, the development of an operational AI-based patent analysis workflow requires substantial software engineering expertise. Access to a conversational interface alone (e.g., GPT Pro) is insufficient, as automated analysis depends on the orchestration of multiple components, including data retrieval, segmentation, prompting,

classification, and quality control. This requirement introduces additional technical complexity and operational costs.

Furthermore, current LLM architectures lack persistent memory and remain susceptible to hallucinations, both of which affect the reliability of automated patent summaries and classifications. The repeated reconstruction of context increases computational overhead, while hallucinated outputs may distort the interpretation of claims or introduce subtle inaccuracies, requiring additional validation steps.

Prompt engineering therefore plays a critical role in achieving adequate precision, particularly in domains characterized by closely related technical terminology, such as LNG and hydrogen technologies. Careful prompt design and, where necessary, domain-specific adaptation can reduce ambiguity and steer model outputs toward evidence-based reasoning grounded in patent claims, albeit at the cost of additional expertise and resources.

Finally, large-scale analytical pipelines require careful optimization to control computational costs. Complex, multi-step workflows may generate redundant iterations or unproductive processing cycles, which scale poorly when applied to thousands of patents. Even minor inefficiencies can therefore result in disproportionate cost increases, highlighting the importance of lean pipeline design, continuous monitoring, and iterative optimization to achieve practical scalability.

## 5. Conclusions

This study has introduced and validated an innovative framework for patent intelligence. Through the application of advanced AI methodologies, this framework transforms patent documents into synthetic, structured representations, thereby enabling granular classifications and strategic analyses of a depth that would otherwise be unachievable.

The findings of this study carry fundamental implications for the practice of patent intelligence in the era of AI, which can be summarized as follows:

- In-depth analysis and classification of a patent corpus require the individual examination of each document. Reliable strategic outcomes cannot be achieved without a granular reading of the single patent.
- The complexity and length of patent documents present a significant hurdle even for advanced AI models. The generation of structured summaries emerges as an indispensable intermediate step to enable scalable and effective analysis.
- The quality of the results is critically dependent on prompt engineering. The ability to formulate structured and conceptually complex instructions is essential to guide AI in extracting specific information and executing accurate classifications.
- Despite efficiency gains, processing costs and times remain a significant consideration. The pursuit of more precise and granular classifications entails increased computational complexity and, consequently, higher costs associated with the use of APIs and advanced models.

Alongside these implications, it is important to acknowledge the inherent limitations of the proposed framework. AI-based patent analysis systems entail significant economic and operational constraints, including the cost of reliable data sources, the computational demands of advanced language models, and the complexity of orchestrating multi-step analytical workflows at scale. Additional limitations arise from current language model architectures, such as the lack of persistent memory and the risk of hallucinated outputs, which require careful system design and validation strategies. These limitations delineate clear directions for future research, aimed at increasing the autonomy and efficiency of AI-driven patent intelligence systems. In particular, addressing the static and memory-less nature of current language models represents a key challenge for the evolution of the proposed framework.

In this context, a promising direction for future development lies in the evolution toward autonomous agent-based systems. Such systems are designed to plan and execute multi-step analytical workflows while maintaining state across different phases, thereby addressing key limitations of current memory-less language models. Agent-based architectures also offer opportunities for improved cost and time efficiency, enabling adaptive strategies that allocate computational resources selectively across different stages of the analysis. Ultimately, this evolution would support a paradigm in which human analysts define high-level strategic objectives, while autonomous agents plan and execute the most efficient sequence of actions to achieve them.

The proposed approach proves particularly effective for the systematic identification of so-called white space opportunities. As demonstrated in the case study, a funneling approach can be adopted: starting from a macroscopic perspective, using acceleration graphs to identify the fastest-evolving robotic systems (e.g., drones), the data are then cross-referenced with emerging technology curves, which highlight the most recent and promising innovations, such as advanced materials or swarm robotics.

The analysis of the “robot type vs. application domain” matrix ultimately enables the identification of specific market niches, such as fire extinguishing, that are highly relevant yet characterized by low patent coverage or a disproportionate concentration of patents in favor of a single player. This indicates a clear competitive vulnerability or an unexplored strategic space.

The preliminary semantic processing of each individual patent (through the extraction of abstracts, claims, and the assignment of specific tags) enables immediate drill-down capabilities for the analyst, which are essential for accurately mapping the state of the art and the boundaries of novelty already claimed by competitors. In the case examined, the analysis revealed that current innovations are limited to drones used as advanced sensing platforms (for image acquisition and oxygen level monitoring) or as mechanical carriers designed to deploy static fire-retardant blankets to contain flames and transport extinguishing agents through dedicated channels.

The final and decisive step consists of leveraging this deep analytical granularity to trigger targeted technology transfer actions and New Concept Generation, with the aim of producing novel patentable ideas. By understanding the precise limitations of the current state of the art (i.e., static and logistically heavy solutions), it becomes possible to implement advanced problem-solving processes, potentially supported by generative AI, to apply emerging technologies (such as swarm systems and advanced materials) to highly complex scenarios, such as fires in the Oil & Gas sector.

For instance: by surrounding the facility, a swarm of drones creates a real-time thermodynamic and chemical Digital Twin. By instantaneously cross-referencing environmental parameters (local wind conditions, O<sub>2</sub> concentration, and fuel type), onboard intelligence can mathematically predict the evolution of the fire in real time. Representing a radical technological leap beyond the use of physical blankets, the drones act fluidodynamically: they generate a downward ‘cyclonic air curtain’ that prevents the lateral dispersion of toxic fumes, reduces oxygen supply at the base of the fire, and confines the blaze within a vertical column, thereby significantly facilitating ground-based extinguishing operations.

## CRedit authorship contribution statement

**Guido Giuntelli:** Investigation, Formal analysis, Data curation. **Christian Spreafico:** Methodology, Formal analysis, Conceptualization. **Andrea Precorvi:** Data curation, Visualization. **Davide Russo:** Methodology, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## References

- [1] K. Meguro, Y. Osabe, Lost in patent classification, *World Pat. Inf.* 57 (2019) 70–76.
- [2] B. Degroote, P. Held, Analysis of the patent documentation coverage of the CPC in comparison with the IPC with a focus on Asian documentation, *World Pat. Inf.* 54 (2018) S78–S84.
- [3] A.M. Villa, M. Wirz, A sequential patent search approach combining semantics and artificial intelligence to identify initial State-of-the-Art documents, *World Pat. Inf.* 68 (2022) 102096.
- [4] P. Soutrenon, A. De Valon, R. Billet, C. Lecante, D. Boulard, J.Y. Legendre, Analyzing green science-related patents in cosmetics by using artificial intelligence, *World Pat. Inf.* 74 (2023) 102223.
- [5] T.Y. Lin, L.C. Chou, A systematic review of artificial intelligence applications and methodological advances in patent analysis, *World Pat. Inf.* 82 (2025) 102383.
- [6] J. Son, H. Moon, J. Lee, S. Lee, C. Park, W. Jung, H. Lim, Ai for patents: a novel yet effective and efficient framework for patent analysis, *IEEE Access* 10 (2022) 59205–59218.
- [7] G. Giuntelli, C. Spreafico, A. Precorvi, D. Russo, AI-Powered patent classification: a hybrid approach merging algorithms and LLMs, in: *International TRIZ and Artificial Intelligence Conference*, Springer Nature Switzerland, Cham, 2025, pp. 169–179.
- [8] R. Krestel, R. Chikkamath, C. Hewel, J. Risch, A survey on deep learning for patent analysis, *World Pat. Inf.* 65 (2021) 102035.
- [9] T. Montecchi, D. Russo, Y. Liu, Searching in cooperative patent classification: comparison between keyword and concept-based search, *Adv. Eng. Inform.* 27 (3) (2013) 335–345.
- [10] P. Carrara, D. Russo, Patent searches opinion: how to minimize the risk when reviewing patent applications, *World Pat. Inf.* 49 (2017) 43–51.
- [11] D. Russo, Knowledge extraction from patent: achievements and open problems. A multidisciplinary approach to find functions, in: *Global Product Development: Proceedings of the 20th CIRP Design Conference*, Ecole Centrale De Nantes, Nantes, France, 19th–21st April 2010, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 567–576.
- [12] B. Balsmeier, M. Assaf, T. Chesebro, G. Fierro, K. Johnson, S. Johnson, L. Fleming, Machine learning and natural language processing on the patent corpus: data, tools, and new measures, *J. Econ. Manag. Strat.* 27 (3) (2018) 535–553.
- [13] I. Igna, F. Venturini, The determinants of AI innovation across European firms, *Res. Pol.* 52 (2) (2023) 104661.
- [14] M. Miric, N. Jia, K.G. Huang, Using supervised machine learning for large-scale classification in management research: the case for identifying artificial intelligence patents, *Strateg. Manag. J.* 44 (2) (2023) 491–519.
- [15] V. Verendel, Tracking artificial intelligence in climate inventions with patent data, *Nat. Clim. Change* 13 (1) (2023) 40–47.
- [16] A. Entezari, A. Aslani, R. Zahedi, Y. Noorollahi, Artificial intelligence and machine learning in energy systems: a bibliographic perspective, *Energy Strategy Rev.* 45 (2023) 101017.
- [17] J.M. Chu, H.C. Lo, J. Hsiang, C.C. Cho, Patent response system optimised for faithfulness: procedural knowledge embodiment with knowledge graph and retrieval augmented generation, in: *Proceedings of the 1st Workshop on Towards Knowledgeable Language Models (Knowllm 2024)*, 2024, pp. 146–155.
- [18] H.Y. Wang, S.H. Chang, C.Y. Chuang, Integrating generative artificial intelligence techniques into technology function matrix analysis, *World Pat. Inf.* 81 (2025) 102352.
- [19] S. Sakaoka, S. Kano, Quantitative insights on artificial intelligence in the pharmaceutical industry: a patent-basis analysis of technological trends and key players, *World Pat. Inf.* 82 (2025) 102381.
- [20] A.L. Kalinichenko, K.W. Willoughby, The effective use of artificial intelligence in patent searches: a case study in using AI-based classifiers to identify AI inventions, *World Pat. Inf.* 82 (2025) 102387.
- [21] Y. Zhu, H. Yuan, S. Wang, J. Liu, W. Liu, C. Deng, J.R. Wen, Large language models for information retrieval: a survey, *ACM Trans. Inf. Syst.* 44 (1) (2025) 1–54.
- [22] H. Tran, Z. Yao, Z. Yang, J. Wang, Y. Zhang, S. Han, H. Yu, Rare: retrieval-augmented reasoning enhancement for large language models, in: *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2025, pp. 18305–18330.
- [23] S. Lee, J. Hsiang, Patent classification by fine-tuning BERT language model, *World Pat. Inf.* 58 (2019) 101841.
- [24] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, pp. 4171–4186.
- [25] J. Risch, R. Krestel, Bag-of-words vs. embeddings: feature representation for patent classification, *World Pat. Inf.* 57 (2019) 100931.
- [26] I. Beltagy, M.E. Peters, A. Cohan, Longformer: the long-document transformer, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 3431–3447.
- [27] Z. Ji, T. Yu, Y. Xu, N. Lee, E. Ishii, P. Fung, Towards mitigating LLM hallucination via self reflection, in: *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, pp. 1827–1843.
- [28] A.L. Porter, S.W. Cunningham, *Tech Mining: Exploiting New Technologies for Competitive Advantage*, second ed., Wiley, 2022.
- [29] S. Avogadri, G. Alzetta, D. Russo, Discovery omnia: dynamic RAG for enhanced patent analysis and systematic innovation, in: *International TRIZ and Artificial Intelligence Conference*, Springer Nature Switzerland, Cham, 2025, pp. 87–99.
- [30] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, D. Kiela, Retrieval-augmented generation for knowledge-intensive NLP tasks, *Adv. Neural Inf. Process. Syst.* 33 (2020) 9459–9474.

**Davide Russo** is a Full Professor at the University of Bergamo. His research expertise lies in the development of systematic methods for product innovation, Triz and “design for patenting” strategies with a current focus on AI-assisted technologies designed to enhance corporate Research and Development (R&D). He is the founder of two academic spin-offs, including **TRIX (now acquired by Tinexta group SPA)**, a company specialized in AI-based search engines for advanced patent analysis. With an extensive track record in intellectual property, he is the inventor of 22 patents and has authored numerous scientific publications including many on patent intelligence and search engine methodologies. For over 20 years, Professor Russo has led high-level patent investigations and patent intelligence projects for major international manufacturing companies, bridging the gap between academic research and industrial application.

**Christian Spreafico** is an Associate Professor at the University of Bergamo. His research focuses on the development of methods and tools for knowledge retrieval to support engineering design, product innovation, and sustainable design. To this end, he integrates large language models, multi-agent architectures, patent analysis, life cycle assessment (LCA), machine learning, and FMEA. On these topics, he has served as coordinator of a European Union-funded research project under the NextGenerationEU program. He is the inventor of six international patents developed in collaboration with companies across multiple industrial sectors and regularly serves as a reviewer for several international scientific journals.

**Guido Giuntelli** holds a degree in Mechanical Engineering and is currently a PhD candidate in the Technology, Innovation and Management (TIM) program at the University of Bergamo (Italy). His doctoral research focuses on the application of Artificial Intelligence (AI) techniques for the analysis of large-scale document corpora, with a specific interest in extracting strategic insights for innovation management. He has already authored some scientific publications in this field and has actively contributed to the development of software solutions for the automated analysis of patents leveraging AI-driven methodologies.

**Andrea Precorvi** is an Innovation Advisor and Software Developer in the Digital Division of Tinexta Group. His expertise lies in developing AI-based search engines for advanced patent analysis and technical documentation. With a multi-year background as a patent documentalist, he provides high-level patent intelligence and research services to multinational corporations across diverse industrial sectors, conducting comprehensive searches and analyses using advanced AI tools.