Agent Ideate: A Framework for Product Idea Generation from Patents Using Agentic AI

Gopichand Kanumolu[1] **Ashok Urlana**[1,2] **Charaka Vinayak Kumar**[1]
Bala Mallikarjunarao Garlapati[1]
TCS Research, Hyderabad, India[1] IIIT Hyderabad[2]
{gopichand.kanumolu, ashok.urlana, charaka.v, balamallikarjuna.g}@tcs.com
ashok.u@research.iiit.ac.in

Abstract

Patents contain rich technical knowledge that can inspire innovative product ideas, yet accessing and interpreting this information remains a challenge. This work explores the use of Large Language Models (LLMs) and autonomous agents to mine and generate product concepts from a given patent. In this work, we design ***Agent Ideate***, a framework for automatically generating product-based business ideas from patents. We experimented with open-source LLMs and agent-based architectures across three domains: Computer Science, Natural Language Processing, and Material Chemistry. Evaluation results show that the agentic approach consistently outperformed standalone LLMs in terms of idea quality, relevance, and novelty. These findings suggest that combining LLMs with agentic workflows can significantly enhance the innovation pipeline by unlocking the untapped potential of business idea generation from patent data.

1 Introduction

With the rapid advancement of large language models (LLMs), there is growing interest in leveraging these models for tasks such as scientific discovery and innovation support. However, generating viable and actionable product ideas from patents requires not only comprehension of complex technical content but also creativity, domain knowledge, and market awareness (Urlana et al., 2024). Patents are legal documents that protect inventions and promote technological innovation (Mossoff, 2000), but their complex and technical language poses unique challenges. Despite the wealth of technical insights contained within patent documents, generating product business ideas from patents remains an underexplored area (Jiang and Goetz, 2024). To achieve this, AgentScen 2025 shared task[1] on Product Business Idea Generation from Patents (PBIG) was introduced as part of the 2^{nd} Workshop on Agent AI for Scenario Planning at IJCAI-25.

[1]https://sites.google.com/view/agentscen/shared-task

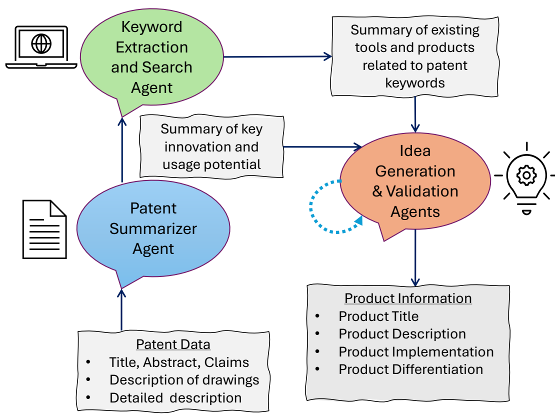

Figure 1: Illustration of the Agent Ideate Pipeline.

1.1 Task formulation

The goal of this task is to evaluate systems that can read a patent and generate a realistic product idea that could be implemented and launched within three years. Each submission is expected to produce four concise outputs for a given patent:

1. ***Product title***: A concise name for the product.

2. ***Product description***: A brief explanation of the product outlining its essential features, target users, their needs, and the benefits provided by the product.

3. ***Implementation***: An explanation describing the implementation of patents technology into the product.

4. ***Differentiation***: An explanation highlighting what makes the product unique.

To support this task, the organizers released a curated dataset consisting of 150 U.S. patents across three categories: Computer Science (CS), Natural

Language Processing (NLP), and Material Chemistry. Participants were allowed to use external resources to enhance idea generation. System outputs were evaluated by both human experts and LLM-based evaluators based on multiple criteria, including technical feasibility, innovation, specificity, market need, and competitive advantage.

In this study, we built the **Agent Ideate** framework, which is a Multi-Agent architecture leveraging an external search tool for generating product ideas from patent text. The pipeline diagram is illustrated in Figure 1. We leverage an LLM-based judging approach to evaluate the ideas generated by the different methods and to select the most effective one. We also analyze the effectiveness of agent-based and LLM-driven architectures for transforming patent knowledge into innovative product concepts.

2 Related Work

The task of generating business ideas from patent documents (Yoshiyasu, 2025; Xu et al., 2025; Terao and Tachioka, 2025; Hoshino et al., 2025; Shimanuki et al., 2025) intersects with multiple research domains, including patent analysis (Sheremetyeva, 2003), patent summarization (Sharma et al., 2019), knowledge extraction (Tonguz et al., 2021), and large language model (LLM)-driven ideation. Prior work has explored the use of NLP and information retrieval techniques to extract technical concepts (Suzuki and Takatsuka, 2016; Tonguz et al., 2021) and commercial potential applications from patent texts (Souili et al., 2015; Jiang et al., 2025). More recently, LLMs have been applied for creative tasks such as product ideation, innovation support, showing promise in structured content generation (Girotra et al., 2023; Radensky et al., 2024; Li et al., 2024; Wen et al., 2006).

One closely related line of work is by Si et al. (2024), who investigate the research ideation capabilities of LLMs. They pose a critical question: Are current LLMs capable of generating novel ideas that rival those produced by human experts? To answer this, the authors conducted a large-scale study involving over 100 qualified NLP researchers who generated human baselines and performed blind evaluations of both human and LLM-generated ideas. Their findings reveal that LLM-generated ideas are often judged as more novel than those produced by domain experts.

Section	CS	NLP	Chemistry
Title	10	11	8
Abstract	134	138	130
Background	1058	910	6215
Claims	1499	1708	535
Description of Figures	4636	868	700
Detailed Description	1499	5068	156

Table 1: Average number of words present in each section for different datasets. CS - Computer science, NLP - Natural Language Processing.

In another study, SciMON(Wang et al., 2024) is a framework that enhances language models' ability to generate novel scientific ideas by leveraging literature-based inspirations and iterative novelty optimization. Unlike traditional link-prediction approaches, it takes contextual inputs (e.g., research problems) and produces natural language hypotheses, using retrieval from semantic, knowledge graph, and citation sources. While evaluations show improvements over GPT-4, the generated ideas still lack the depth and novelty of human-authored research. To this end, in contrast to the existing works, this study aims to generate product-based business ideas from patents by building a multi-agentic framework.

3 Dataset

The dataset provided by the shared task organizers comprises a total of 150 U.S. patents, with 50 patents each from three distinct domains: Computer Science(CS), Natural Language Processing (NLP), and Material Chemistry(MC). Each patent entry includes structured metadata such as the title, abstract, claims, description, publication number, and publication date.

Preprocessing: Among these fields, the description section is notably extensive, often exceeding the input length limitations of most large language models (LLMs). To address this challenge, we implemented a preprocessing strategy that segments the description into semantically meaningful subsections. This was achieved through regular expression-based matching, which identifies and extracts parts such as: Background information, Brief description of drawings and claims, and Detailed description of the patent technology.

This segmentation allows for more efficient and focused processing by LLMs and downstream agents. Detailed statistics about the dataset distribution and content lengths across categories are summarized in Table 1.

4 Methodology

As presented in Figure 1, we adopt three distinct methods to generate innovative business ideas from patent documents. These methods are increasingly sophisticated in terms of architecture and capability:

1. Prompt-based LLM Approach: This is the simplest baseline. We use a single-prompt approach with a large language model (LLM), wherein the entire patent (or its reduced components: title, abstract, claims, and summarized description) is passed as input to the model. The prompt is crafted to guide the model in generating business ideas, specifying the required structure in JSON format with fields such as product title, product description, implementation, and differentiation.

2. Multi-Agent LLM Architecture: The second approach builds on modularization via a multi-agent system, where different tasks are handled by different specialized agents. Specifically:

- A Patent Analyst Agent summarizes the core innovation and usage of the patent.
- A Business Idea Generator Agent uses the summarized insight to generate a structured business idea.
- A Business Validator Agent ensures the output adheres to format, character limits, and originality constraints.

Each agent uses the same LLM backend but is provided with a distinct goal and context. Tasks are executed sequentially with inter-agent context passing, allowing for better modularity, reliability, and control compared to single-shot prompting. In the rest of the paper, we refer to this method as the *Agent without Tool* approach.

3. Multi-Agent LLM with External Search Tool: The third and most comprehensive method incorporates a search tool to enrich the reasoning process with external information. It extends the second approach by introducing:

- A Keyword Extractor Agent, which identifies two core keywords from the summarized patent content.
- A Research Agent, which performs a DuckDuckGo tool-based web search using these keywords to gather information about existing tools, libraries, or products in the domain.
- The Business Idea Generator Agent utilizes both the patent summary and external market insights to create a business idea that is clearly differentiated from known solutions.
- Finally, the Business Validator Agent ensures the output is well-formed, concise, and novel.

We provide the role, goal, backstory, tool usage, task description, and expected output instructions for each agent in Appendix Table 4 and Table 5. In the rest of the paper, we refer to this method as the *Agent with Tool* approach.

5 Experiments and Evaluation

We conduct experiments with prompt-based, agent with Tool and agent without Tool based approaches. For all experiments, we used the llama-4-scout-17b-16e-instruct[2] model for response generation in both architectures: the prompt-based LLM model and each agent in the multi-agent setup. Due to resource constraints and the lack of access to proprietary APIs such as OpenAI, we opted to experiment with open-source LLMs hosted via the Groq API[3]. The LLM was configured with a temperature of 0.7 and a maximum token limit of 1000. All experiments were conducted using the free-tier access provided by Groq. For all agentic framework experiments, we used the CrewAI[4] framework to create agents and integrate with external search tools.

To assess the relative quality of business ideas generated by different methods, we employed an LLM-as-a-judge evaluation strategy. Specifically, we designed a structured prompt where the model is provided with a patent description and two product ideas generated using different approaches (e.g., baseline prompting vs. multi-agent with search). The LLM is then instructed to critically evaluate the ideas across six well-defined dimensions: technical validity, innovativeness, specificity, need validity, market size, and competitive advantage.

The evaluation setup and criteria are provided in Appendix Table 6 and Table 7. Explicitly listing the criteria reduces ambiguity and encourages the model to weigh each dimension before issuing a verdict. The output follows a strict JSON format, containing the selected better idea (idea 1 or idea 2) and a rationale for the decision.

[2] https://console.groq.com/docs/model/meta-llama/llama-4-scout-17b-16e-instruct
[3] https://console.groq.com/docs/models
[4] https://www.crewai.com/

Domain	Idea 1	Idea 2	Idea 1 Count(%)	Idea 2 Count(%)
Computer Science	Prompt-based LLM	**Agent without tool**	14	**86**
	Agent without tool	**Agent with Tool**	14	**86**
NLP	Prompt-based LLM	**Agent without tool**	02	**98**
	Agent without tool	Agent with Tool	**88**	12
Material Chemistry	Prompt-based LLM	**Agent without tool**	08	**92**
	Agent without tool	Agent with Tool	**64**	38

Table 2: Evaluation of ideas generated using various approaches. We employ the LLM-as-a-Judge method to compare the ideas and report the percentage of ideas selected by the judge.

Criteria	Chemistry	CS	NLP
Tech Validity	1	2	3
Specificity	3	3	3
Need Validation	5	2	4
Market Size	5	1	1
Innovativeness	5	3	4
Competitive Advantage	2	3	3

Table 3: Human evaluation results provided by the organizers. Each row represents the rank/position of our submission "**TrustAI**" for each domain based on the scores for each criteria.

We used a high-capacity model LLaMA 3 70B [5] hosted via Groq for inference, ensuring strong reasoning and evaluation capabilities. This method of LLM-based comparative evaluation offers a scalable and cost-effective alternative to human annotation, especially in scenarios involving nuanced technical and entrepreneurial judgments. Furthermore, by leveraging LLMs that are blind to the origin of each idea, we minimize bias and ensure that comparisons focus purely on idea quality, not model provenance.

6 Discussion

Evaluation using LLM as a judge: The automated evaluation results (using LLM as judge) in Table 2 show clear performance differences between approaches. The Agent with Tool method consistently generates highly-ranked ideas in Computer Science (86%), demonstrates moderate performance in Material Chemistry (38%), but performs poorly in NLP (12%). The standalone Agent approach without tool usage shows strong performance in NLP (98%) and Material Chemistry (64%), though it is less effective in Computer Science (14%) compared to the Agent with Tool method. The basic LLM prompt method performs poorly across all domains (Computer Science: 14%, NLP: 02%, Material Chemistry: 08%), suggesting that multi-agent frameworks provide substantial benefits even without tool access.

Based on the automatic evaluation results comparing which approach generated the best ideas for each domain, we submitted the highest-performing outputs for organizer evaluation. The results of this evaluation are discussed in the following section.

Evaluation results given by the Organizers: The human evaluation rankings in Table 3 reveal important domain-specific patterns. In Chemistry, our system achieved top rankings in Innovativeness (1^{st}) but performed poorly in Technical Validity (5^{th}), indicating highly creative but potentially less feasible ideas. For Computer Science, we see balanced performance across criteria (mostly 2^{nd}-3^{rd} place), suggesting reliable but not exceptional results. The NLP domain shows our strongest overall performance, with top-3 rankings in all criteria except Market Size (5^{th}), highlighting both the technical strength and potential niche focus of generated ideas.

7 Conclusion

This paper presented our framework Agent Ideate, for generating product ideas from patents. We have conducted experiments using prompt-based LLM, multi-agent framework, and tool-augmented agents. Automated evaluation (LLM-as-judge) showed that Agent with Tool performed best in Computer Science, while standalone Agent excelled in NLP, and Material Chemistry. Our findings highlight the potential of agentic AI for structured innovation while underscoring domain-specific challenges.

[5] https://console.groq.com/docs/model/llama-3.3-70b-versatile

8 Limitations

Our study has several key limitations. First, reliance on open-source LLMs (e.g., LLama-4-17B, and LLaMA-3-70B) may restrict performance compared to state-of-the-art proprietary models. Second, the system's effectiveness varies significantly across domains, requiring domain specific models. Finally, the tool-augmented agent's performance depends heavily on external search quality, which can introduce noise. These constraints highlight the need for more robust domain adaptation, hybrid evaluation methods, and improved tool integration in future work.

References

Karan Girotra, Lennart Meincke, Christian Terwiesch, and Karl T Ulrich. 2023. Ideas are dimes a dozen: Large language models for idea generation in innovation. *The Wharton School Research Paper Forthcoming*.

Mizuki Hoshino, Shun Shramatsu, and Fuminori Nagasawa. 2025. A business idea generation framework based on creative multi-agent discussions. In *The 2nd Workshop on Agent AI For Scenario Planning (AGENTSCEN2025)*.

Lekang Jiang and Stephan Goetz. 2024. Artificial intelligence exploring the patent field. *arXiv e-prints*, pages arXiv–2403.

Lekang Jiang, Caiqi Zhang, Pascal A. Scherz, and Stefan Goetz. 2025. Can large language models generate high-quality patent claims? In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 1272–1287, Albuquerque, New Mexico. Association for Computational Linguistics.

Long Li, Weiwen Xu, Jiayan Guo, Ruochen Zhao, Xingxuan Li, Yuqian Yuan, Boqiang Zhang, Yuming Jiang, Yifei Xin, Ronghao Dang, and 1 others. 2024. Chain of ideas: Revolutionizing research via novel idea development with llm agents. *arXiv preprint arXiv:2410.13185*.

Adam Mossoff. 2000. Rethinking the development of patents: an intellectual history, 1550-1800. *Hastings Lj*, 52:1255.

Marissa Radensky, Simra Shahid, Raymond Fok, Pao Siangliulue, Tom Hope, and Daniel S Weld. 2024. Scideator: Human-llm scientific idea generation grounded in research-paper facet recombination. *arXiv preprint arXiv:2409.14634*.

Eva Sharma, Chen Li, and Lu Wang. 2019. BIGPATENT: A large-scale dataset for abstractive and coherent summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2204–2213, Florence, Italy. Association for Computational Linguistics.

Svetlana Sheremetyeva. 2003. Natural language analysis of patent claims. In *Proceedings of the ACL-2003 Workshop on Patent Corpus Processing*, pages 66–73, ,. Association for Computational Linguistics.

Masaya Shimanuki, Naoto Shimizu, Kentaro Kinugasa, and Hiroki Sugisawa. 2025. Business idea generation from patent documents: Knowledge integration and self-improvement via llm. In *The 2nd Workshop on Agent AI For Scenario Planning (AGENTSCEN2025)*.

Chenglei Si, Diyi Yang, and Tatsunori Hashimoto. 2024. Can llms generate novel research ideas? a large-scale human study with 100+ nlp researchers. *arXiv preprint arXiv:2409.04109*.

Achille Souili, Denis Cavallucci, and François Rousselot. 2015. Natural language processing (nlp)–a solution for knowledge extraction from patent unstructured data. *Procedia engineering*, 131:635–643.

Shoko Suzuki and Hiromichi Takatsuka. 2016. Extraction of keywords of novelties from patent claims. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1192–1200, Osaka, Japan. The COLING 2016 Organizing Committee.

Yasunori Terao and Yuuki Tachioka. 2025. Collaborative invention: Refining patent-based product ideation via llm-guided selection and rewriting. In *The 2nd Workshop on Agent AI For Scenario Planning (AGENTSCEN2025)*.

Ozan Tonguz, Yiwei Qin, Yimeng Gu, and Hyun Hannah Moon. 2021. Automating claim construction in patent applications: The CMUmine dataset. In *Proceedings of the Natural Legal Language Processing Workshop 2021*, pages 205–209, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ashok Urlana, Charaka Vinayak Kumar, Ajeet Kumar Singh, Bala Mallikarjunarao Garlapati, Srinivasa Rao Chalamala, and Rahul Mishra. 2024. Llms with industrial lens: Deciphering the challenges and prospects–a survey. *arXiv preprint arXiv:2402.14558*.

Qingyun Wang, Doug Downey, Heng Ji, and Tom Hope. 2024. SciMON: Scientific inspiration machines optimized for novelty. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 279–299, Bangkok, Thailand. Association for Computational Linguistics.

Guihua Wen, Lijun Jiang, Jun Wen, and Nigel R Shadbolt. 2006. Generating creative ideas through patents. In *Proceedings of the 9th Pacific Rim international conference on Artificial intelligence*, pages 681–690.

Yuzheng Xu, Tosho Hirasawa, Seiya Kawano, Shota Kato, and Tadashi Kozuno. 2025. Mk2 at pbig competition: A prompt generation solution. In *The*

2nd Workshop on Agent AI For Scenario Planning (AGENTSCEN2025).

Hayato Yoshiyasu. 2025. Team ns_nlp at the agentscen shared task: Structured ideation using divergent and convergent thinking. In *The 2nd Workshop on Agent AI For Scenario Planning (AGENTSCEN2025)*.

A Appendix

We present the description of each agent's role, goal, backstory, and the tools they can access in Table 4 and Table 5. This also includes the task descriptions and expected outputs for each agent. Additionally, we provide the evaluation criteria used to compare the ideas generated by various methods using the LLM-as-a-judge approach in Table 6 and Table 7.

Agent Name	Role	Goal	Backstory / Tools Used
Patent Analyst	Reader Agent	Extract and summarize key features from patents	Specializes in understanding complex patent documents and identifying key technological aspects.
Keyword Extractor	Keyword Agent	Generate essential keywords from patent summary	NLP expert identifying core technologies to support product discovery.
Researcher	Search Agent	Search for relevant products/tools using keywords and synthesize results	Enthusiast in discovering tools/products relevant to keywords with clear and concise summaries. Tools Used: DuckDuckGo Tool
Idea Generator	Business Idea Agent	Generate innovative product ideas from patent content	Creative entrepreneur skilled in mapping technology to business ideas.
Business Validator	Validator Agent	Validate ideas for structure and uniqueness	Ensures business ideas are well-formatted, feasible, and differentiated from existing solutions.

Table 4: Description of each agent's role, goal, backstory, and tool usage

Task Name	Performed By	Task Description	Expected Output
Patent Analysis	Patent Analyst	Read and extract core information from patent sections	Structured summary of key patent features.
Keyword Generation	Keyword Extractor	Generate two keywords representing the patent's core technological concepts	List of keywords: `["keyword1", "keyword2"]`
Product Research	Researcher	Use keywords to search web using DuckDuckGo Tool for related products and synthesize findings	Text summary with relevant products/tools and short descriptions.
Idea Generation	Idea Generator	Based on findings and patent, generate an innovative product/business idea	JSON object with below fields: `product_title`, `product_description`, `implementation`, `differentiation`.
Idea Validation	Business Validator	Review the generated idea for adherence to format and uniqueness	Validated JSON output with feedback on issues if any.

Table 5: Description of each agent's task, and expected output for each task.

Aspect	Description
Evaluator Role	LLM-as-a-Judge: A large language model is prompted to objectively compare two product ideas derived from a common patent.
Input Provided	1. Patent description 2. Two distinct product/business ideas using the patent
Evaluation Goal	Select the better idea based on well-defined business and technical criteria.
Prompt Structure	Multi-section prompt including: • <patent>: full patent description • <idea_1>, <idea_2>: structured product ideas • Explicit list of 6 evaluation criteria (refer Table 7)
LLM Output Format	JSON: `{"output": "idea_1 or idea_2", "reason": "reason for the choice"}`
Use Case	Used for comparative evaluation of generated product ideas, testing how well different agents or models transform patent knowledge into viable business ideas.

Table 6: Evaluation (LLM-as-a-Judge) Setup Overview

Criterion	Explanation
Technical Validity	Is the patent technology appropriate and realistically implementable within 3 years?
Innovativeness	Does the idea utilize the patent in a novel way? Does it stand out in terms of technological creativity?
Specificity	Is the idea clearly and narrowly defined (e.g., “manage references” vs. “do research”)?
Need Validity	Is there a clear and valid user need addressed by the product idea?
Market Size	Is the target market large enough to make the product viable? Are there many potential users?
Competitive Advantage	Does the use of the patented technology offer a unique advantage over competitors?

Table 7: Description of evaluation criteria of generated ideas using LLM as a judge.