

# 000 IPBENCH: BENCHMARKING THE KNOWLEDGE OF 001 LARGE LANGUAGE MODELS IN INTELLECTUAL PROP- 002 ERTY 003 004

006 **Anonymous authors**

007 Paper under double-blind review  
008  
009  
010  
011

## ABSTRACT

013 Intellectual Property (IP) is a highly specialized domain that integrates technical and legal knowledge, making it inherently complex and knowledge-intensive.  
014 Recent advancements in LLMs have demonstrated their potential to handle IP-  
015 related tasks, enabling more efficient analysis, understanding, and generation of  
016 IP-related content. However, existing datasets and benchmarks focus narrowly  
017 on patents or cover limited aspects of the IP field, lacking alignment with real-  
018 world scenarios. To bridge this gap, we introduce **IPBench**, the first compre-  
019 hensive IP task taxonomy and a large-scale bilingual benchmark encompassing  
020 **8 IP mechanisms and 20 distinct tasks**, designed to evaluate LLMs in real-  
021 world IP scenarios. We benchmark **17 main LLMs**, ranging from general purpose  
022 to domain-specific, including chat-oriented and reasoning-focused models, under  
023 zero-shot, few-shot, and chain-of-thought settings. Our results show that even  
024 the top-performing model, DeepSeek-V3, achieves only 75.8% accuracy, indicat-  
025 ing significant room for improvement. Notably, open-source IP and law-oriented  
026 models lag behind closed-source general-purpose models. To foster future re-  
027 search, we publicly release IPBench, and will expand it with additional tasks to  
028 better reflect real-world complexities and support model advancements in the IP  
029 domain. We provide the data and code in the supplementary materials.  
030  
031

## 1 INTRODUCTION

032 Intellectual property (IP) is the embodiment of human creativity and innovation (WIPO, 2020a)  
033 protected through legal frameworks such as patents, copyrights, and trademarks. Owing to its  
034 intersection of technical and legal domains, IP-related tasks are inherently knowledge-intensive,  
035 highly applicable to real-world scenarios, and hold substantial practical value. Beyond domain-  
036 specific expertise, these tasks demand robust capabilities in information processing, logical reason-  
037 ing, decision-making, and creative generation.

038 With the advancement of large language models (LLMs) (Achiam et al., 2023; DeepSeek-AI et al.,  
039 2024), there is increasing potential to automate tasks across domains, including those in IP. LLMs  
040 offer a generalizable framework for understanding, processing, and generating complex content,  
041 paving the way for more efficient IP information management and decision support. Nowadays,  
042 NLP researchers have been paying increasing attention to the field of intellectual property. This has  
043 spurred growing interest among NLP researchers in IP applications. For example, Jiang & Goetz  
044 (2024) provide a comprehensive survey of patent-related NLP tasks, classifying them into analysis  
045 and generation categories, but their focus is limited to patent text.

046 Recent efforts have introduced datasets such as HUPD (Suzgun et al., 2023), which compiles a  
047 corpus of patent and defines tasks including subject classification, language modeling, and summa-  
048 rization. While practically useful, HUPD emphasizes linguistic attributes and neglects the deeper  
049 technical and legal aspects essential to IP evaluation. Similarly, benchmarks like PatentEval (Zuo  
050 et al., 2024), MoZIP (Ni et al., 2024), and IPEval (Wang et al., 2024b) concentrate on narrow and  
051 specific IP task scopes. Moreover, most existing benchmarks center exclusively on patents, leaving  
052 other critical IP mechanisms-such as trademarks and copyrights-largely unaddressed. Despite the  
053

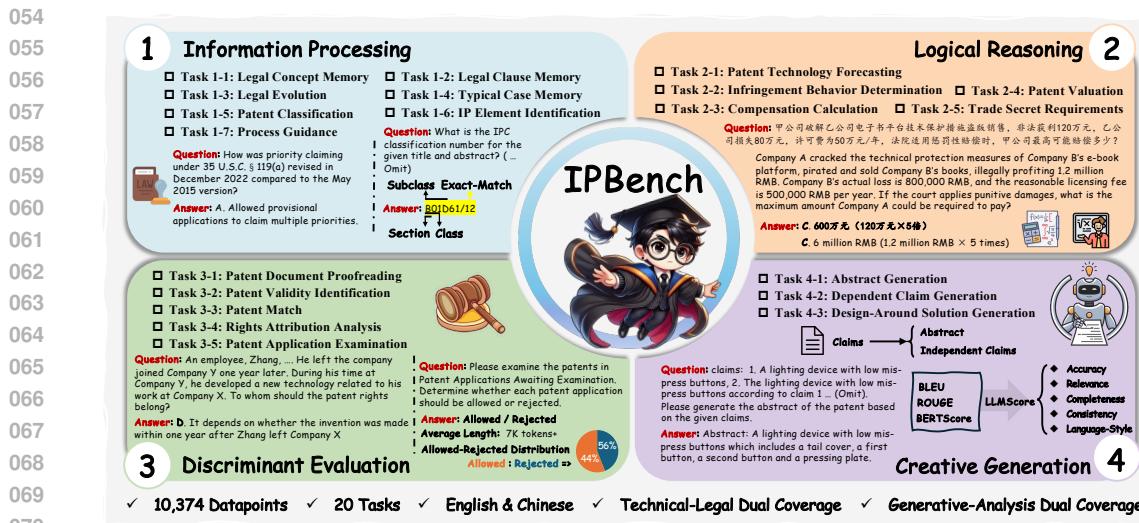


Figure 1: Overview of the comprehensive IP task taxonomy and IPBench.

field's real-world impact, there remains a lack of a comprehensive task taxonomy and benchmark that fully captures the breadth and complexity of IP scenarios.

To address the growing demand for effective LLMs applications in IP, we introduce the first comprehensive task taxonomy tailored to real-world IP challenges, as illustrated in Figure 1. It is grounded in Webb's Depth of Knowledge (DOK) theory (Webb, 2002) and extended to capture four hierarchical levels: **Information Processing**, **Logical Reasoning**, **Discriminant Evaluation**, and **Creative Generation**. These levels reflect the cognitive complexity inherent in IP tasks and provide a structured framework to assess the depth of LLMs understanding. Our taxonomy incorporates intrinsic knowledge evaluation and in-depth textual analysis from both point-wise and pairwise perspectives, covering the interplay between technical and legal reasoning.

Building on this taxonomy, we present **IPBench**, the first large-scale, comprehensive benchmark for evaluating LLMs on IP knowledge. **IPBench** comprises 10,374 data points across 20 diverse tasks, aligned with 8 core IP mechanisms. Our benchmark is bilingual (English and Chinese), and is grounded in the legal frameworks of the United States and mainland China, allowing cross-jurisdictional evaluation. IPBench tasks are carefully designed to span a spectrum of difficulty and task formats, including classification, retrieval, and open-ended generation, enabling holistic evaluation of model capabilities in knowledge recall, reasoning, legal judgment, and creative synthesis. We evaluate 17 leading LLMs on IPBench-including general-purpose models, law-oriented models, and IP-specialized models-covering both chat and reasoning-focused architectures, under zero-shot, few-shot and chain-of-thought settings. Our key contributions and findings are as follows:

- We propose the first hierarchical taxonomy for IP domain, rooted in cognitive theory, and introduce **IPBench**, a bilingual benchmark with 10,374 examples spanning 20 tasks and 8 IP mechanisms. This enables realistic, and multi-faceted evaluation of LLMs in IP contexts.
- Our experimental results reveal that even the best-performing model achieves only 75.8% accuracy overall, indicating that current LLMs fall short in reliably handling IP tasks. Notably, closed-source general-purpose models consistently outperform domain-specific open-source models, highlighting a pressing need for more capable and interpretable IP-focused LLMs.
- We include both IPC/CPC classification and conditional generation tasks in IPBench. DeepSeek-R1 achieves the best IPC classification accuracy at 10.8%, while DeepSeek-V3 leads in CPC classification at 9.5%. For generative tasks, we introduce **LLMScore**, a novel evaluation metric based on LLM-as-a-judge methodology, which exhibits stronger alignment with human judgments than traditional automatic metrics.

108  
109  
110

- We conduct comprehensive analyses, including cross-lingual performance comparisons, variations in prompt design, and a taxonomy of 7 major error types.

111 We believe **IPBench** offers a timely and essential tool for advancing the application of LLMs in IP.  
112 From a machine learning perspective, the complexity of IP language can serve as a robust stress test  
113 for LLMs. From a legal and innovation standpoint, automation in this domain can enhance service  
114 intelligence, reduce operational costs, and ultimately accelerate global technological advancement.  
115 We plan to continuously expand IPBench by incorporating additional languages, modalities, and  
116 tasks in future iterations.

117

## 118 2 RELATED WORK

119

120 Prior to the emergence of LLMs, researchers applied NLP techniques to IP tasks, particularly within the  
121 domain of patent analysis. These efforts focused on applications such as patent classification (Lee  
122 & Hsiang, 2020), and abstract or claims generation (Sharma et al., 2019; Lee & Hsiang, 2020).  
123 However, traditional models used in these studies typically lacked generalization capabilities and  
124 required extensive task-specific adaptation, limiting their scalability and real-world applicability.  
125 With the advent of LLMs based on the decoder-only transformer architecture (Radford et al., 2019),  
126 models trained using next-token prediction have demonstrated impressive zero-shot (Kojima et al.,  
127 2022) and few-shot (Brown et al., 2020) capabilities across diverse tasks. This paradigm shift intro-  
128 duced a new approach to handling IP-related challenges using prompt-based inference, reducing the  
129 need for task-specific training and enabling more versatile applications in the IP domain.

130 Recent work has explored the adaptation of  
131 LLMs specifically for IP. Ni et al. (2024) de-  
132 veloped MoZi, a multilingual IP-oriented LLM  
133 based on BLOOMZ and ChatGLM. Bai et al.  
134 (2024) proposed a cost-efficient training frame-  
135 work to fine-tune LLMs for IP tasks, claim-  
136 ing performance on par with human experts.  
137 Other studies, such as Pap2Pat (Knappich et al.,  
138 2024), AutoPatent (Wang et al., 2024b), and  
139 PatentFormer (Wang et al., 2024a), focus on  
140 long-context generation for patent documents  
141 using LLMs. These works predominantly em-  
142 phasize the technical aspects of patent lan-  
143 guage and overlook broader IP mechanisms,  
144 such as trademarks, trade secrets, and copy-  
145 rights. Moreover, they rarely consider legal rea-  
146 soning and decision-making, which are essen-  
147 tial for real-world applications.

Table 1: Comparison of IP related benchmark with Ours. *Gen.-Ana. Dual Cover.* refers to benchmarks that encompass both text generation and analysis tasks. *Tech.-Legal Dual Cover.* refers to benchmarks that contain both aspects of technical and legal content. Meanwhile, *Compre. Taxonomy* refers to a benchmark that possesses a comprehensive taxonomy.

Benchmark	PatentEval	IPEval	MoZIP	Ours
Evaluation for LLMs	✓	✓	✓	✓
Multilingual		✓	✓	✓
Multi-IP Mechanisms		✓	✓	✓
Tech.-Legal Dual Cover.			✓	✓
Gen.-Ana. Dual Cover.				✓
Compre. Taxonomy				✓
LLMs Evaluated #	6	15	5	17 *
Task #	2	1	3	20 *
Testset Size	400	2657	3121	10374 *

148 Our work differs fundamentally in both scope and design. IPBench builds upon and expands these  
149 earlier efforts by introducing a unified, comprehensive IP task taxonomy grounded in Webb’s Depth  
150 of Knowledge (DOK) theory. Notably, we include under-explored areas such as trade secret and  
151 trademark, offering a holistic evaluation of LLM performance across the IP landscape. This makes  
152 IPBench more comprehensive than prior benchmarks like IPEval, MoZIP, and PatentEval. A detailed  
153 comparison is presented in Table 1, highlighting our benchmark’s task diversity, linguistic coverage,  
154 and legal granularity.

155

## 156 3 IPBENCH

157

### 158 3.1 TASK TAXONOMY

159 While previous patent-related benchmarks have primarily focused on textual content such as clas-  
160 sification or summarization they often overlook the broader real-world implications of IP tasks. To  
161 address this gap, we introduce the first comprehensive intellectual property task taxonomy that ex-  
162 tends beyond in-domain textual analysis to encompass the multifaceted real-world demands of the

162  
163  
164  
165  
166 Table 2: Task taxonomy of IPBench. The *EN* in the Language column indicates English, while  
167 *ZH* represents Chinese. The *AE* in the Metric column indicates Automated Evaluation, while *HE*  
168 represents Human Evaluation.

Level	Index	Task Name	Metric	Data Source	Language	Size
Information Processing	1-1	Legal Concept Memory	Accuracy	Expert Annotation	EN/ZH	500
	1-2	Legal Clause Memory	Accuracy	Expert Annotation	EN/ZH	502
	1-3	Legal Evolution	Accuracy	Expert Annotation	EN/ZH	500
	1-4	Typical Case Memory	Accuracy	USTPO / CNIPA	EN/ZH	504
	1-5-1	Patent IPC Classification	Exact Match	USTPO / CNIPA	EN/ZH	1125
	1-5-2	Patent CPC Classification	Exact Match	USTPO	EN	600
	1-6	IP Element Identification	Accuracy	Expert Annotation	EN/ZH	557
	1-7	Process Guidance	Accuracy	Expert Annotation	EN/ZH	548
Logical Reasoning	2-1	Patent Technology Forecasting	Accuracy	Expert Annotation	EN/ZH	500
	2-2	Infringement Behavior Determination	Accuracy	Expert Annotation	EN/ZH	500
	2-3	Compensation Calculation	Accuracy	Expert Annotation	EN/ZH	316
	2-4	Patent Valuation	Accuracy	Expert Annotation	EN/ZH	301
	2-5	Trade Secret Requirements	Accuracy	Expert Annotation	ZH	301
Discriminant Evaluation	3-1	Patent Document Proofreading	Accuracy	Expert Annotation	EN/ZH	300
	3-2	Patent Validity Identification	Accuracy	Expert Annotation	EN/ZH	308
	3-3	Patent Match	Accuracy	MoZIP	EN/ZH	1000
	3-4	Rights Attribution Analysis	Accuracy	Expert Annotation	EN/ZH	400
	3-5	Patent Application Examination	Accuracy	USTPO	EN	314
Creative Generation	4-1	Abstract Generation	AE & HE	USTPO / CNIPA	EN/ZH	400
	4-2	Dependent Claim Generation	AE & HE	USTPO / CNIPA	EN/ZH	400
	4-3	Design-Around Solution Generation	Accuracy	Expert Annotation	EN/ZH	499

182  
183  
184 IP field, spanning both technical and legal dimensions. Given the intrinsic complexity of IP knowl-  
185 edge, effective modeling in this domain requires more than domain-specific understanding. LLMs  
186 must be capable of integrating diverse IP mechanisms, simulating real-world procedural reasoning,  
187 and interpreting varied linguistic styles present in different IP documents and legal jurisdictions.  
188 This necessitates a structured evaluation framework that captures different levels of cognitive depth  
189 and reasoning complexity.

190 To this end, our taxonomy is grounded in the Depth of Knowledge (DOK) theory by American  
191 educator Norman L. Webb, which categorizes cognitive complexity into four levels: *Recall and Re-*  
192 *production, Skills and Concepts, Strategic Thinking, and Extended Thinking*. Originally developed  
193 to guide educational assessment, this framework aligns well with the stratified nature of IP reasoning.  
194 We adapt and reinterpret DOK into a legal and technical context, resulting in four hierarchical  
195 levels tailored for IP evaluation: **Information Processing**, **Logical Reasoning**, **Discriminant Eval-**  
196 **uation**, and **Creative Generation**, as illustrated in Figure 1. These levels enable us to map tasks to  
197 specific reasoning capacities required by LLMs, ranging from simple fact recall to complex synthe-  
198 sis and decision-making. The taxonomy provides a principled foundation for evaluating LLMs not  
199 only in terms of accuracy but also cognitive depth and functional applicability. Table 2 summarizes  
200 the 20 tasks included in IPBench and we also provide further details on the task taxonomy, along  
201 with comprehensive definitions of each task, in Appendix D.

### 202 3.2 DATA PROCESSING AND ANNOTATION

203  
204 **Data Source and Collection.** Our dataset is constructed from three primary sources: expert-  
205 curated annotations, databases maintained by national IP offices, and previously published public  
206 datasets. This diverse sourcing approach ensures broad coverage of real-world scenarios and IP  
207 mechanisms. For tasks grounded in statutory interpretation-such as *Legal Concept Memory*-data are  
208 drawn from official legal texts and documentation published on the public websites of IP offices,  
209 including the United States and China. For litigation-oriented tasks-such as *Infringement Behav-*  
210 *ior Determination*-we utilize publicly available judicial decisions, including case repositories such  
211 as China Judgements Online. Patent-related tasks leverage structured data from the USPTO and  
212 the China National Intellectual Property Administration (CNIPA). All sources used in IPBench are  
213 publicly accessible, ensuring transparency and reproducibility.

214  
215 **Data Processing and Annotation.** Our IPBench is constructed as a gold-standard benchmark  
through extensive human expert annotation. Given the highly structured nature of patent documents,

both the USPTO and CNIPA datasets offer well-organized metadata-enabling the systematic creation of paired inputs, such as sequential claim pairs that reflect logical progression in legal language. To ensure annotation quality and domain relevance, we engaged 21 trained annotators, including senior undergraduate and PhD students, all supervised by four certified and experienced patent agents. Most annotators hold academic backgrounds in IP, equipping them with foundational knowledge of both technical and legal aspects of IP. This subject matter expertise was critical to generating high-fidelity annotations across legal, technical, and procedural tasks.

The annotation team is organized into four subgroups, each dedicated to one of the hierarchical levels in our taxonomy. Each task underwent a rigorous two-stage workflow: one team conducted the initial annotation while another team reviewed and validated the results. The roles were then rotated to ensure objectivity and consistency across all data points. Following annotation, we perform automatic quality filtering using cosine similarity based on the BGE-M3 model (Chen et al., 2024). This step eliminates semantically redundant examples and further enhanced the dataset’s diversity and representativeness. Our complete annotation and examination protocol is in Appendix E.

### 3.3 FEATURE OF IPBENCH

IPBench consists of 10,374 expertly curated questions spanning 20 tasks. These tasks are systematically organized across 4 hierarchical levels and cover 8 IP mechanisms, including patents and trade secrets, etc. The benchmark integrates both technical and legal domains and includes a mix of task formats, ranging from classification and comprehension to open-ended generative reasoning. This diverse coverage enables comprehensive evaluation of LLM capabilities, including factual recall, legal reasoning, procedural understanding, and content synthesis. As shown in Table 1, IPBench surpasses existing IP benchmarks across multiple dimensions, including task diversity, jurisdictional representation, cognitive complexity, and linguistic variation.

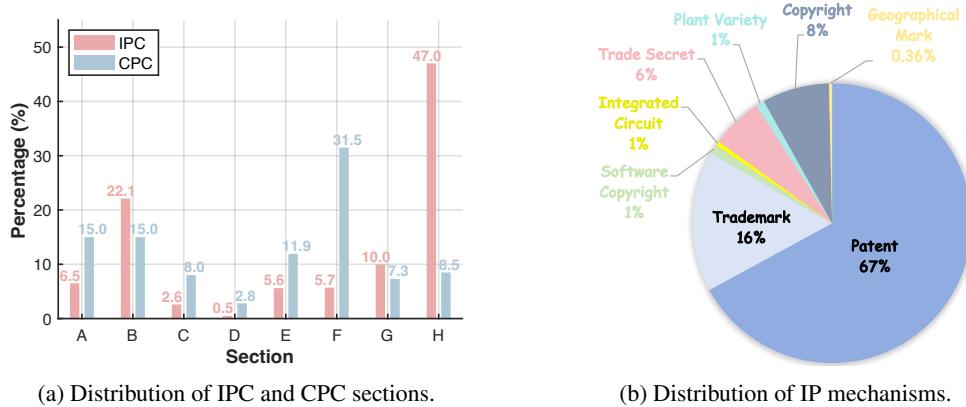


Figure 2: Distributions across IPC/CPC sections and IP mechanisms.

Given the wide scope of IP mechanisms and technical domains represented, we provide detailed statistical analysis of IPBench’s data characteristics. These include the distributions of International Patent Classification (IPC) and Cooperative Patent Classification (CPC) codes, as shown in Figure 2a, and IP mechanisms, as shown in Figure 2b. More feature details of IPBench can be found in Appendix F including the distributions of IPC/CPC classification codes, text lengths, and domain coverage.

## 4 BENCHMARKING RESULTS

### 4.1 EVALUATION SETUP

**Evaluated Models.** We evaluate 17 language models covering a broad range of sizes, architectures, and domain specializations, with details provided in Appendix I. Among them, 14 are general-purpose large language models, 2 are law-oriented models specifically fine-tuned for legal tasks, and one is an IP-oriented model developed for intellectual property applications.

270 Table 3: Main results of IPBench. The best-performing model in each task is in **darker red**, and  
 271 the second best is in **lighter red**. The model DS-Qwen refers to DeepSeek-R1-Distill-Qwen.  
 272

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	3-5	4-3
GPT-4o	75.3	96.0	<b>92.0</b>	82.2	<b>83.7</b>	64.2	<b>71.9</b>	54.8	62.6	63.9	78.5	84.1	71.0	70.1	81.3	<b>83.5</b>	<b>50.0</b>	75.4
GPT-4o-mini	72.6	94.4	87.5	80.2	82.1	58.8	67.5	50.2	64.0	59.5	76.7	83.4	67.3	75.0	<b>81.6</b>	78.5	44.0	66.3
DeepSeek-V3	<b>75.8</b>	<b>96.6</b>	90.2	<b>88.4</b>	82.8	<b>66.1</b>	69.9	56.8	64.2	66.1	76.7	84.1	<b>72.0</b>	75.0	78.9	<b>83.5</b>	44.6	<b>78.8</b>
Qwen3	70.6	94.4	83.1	75.0	76.6	60.9	66.8	51.4	<b>66.8</b>	60.4	75.1	82.7	69.7	74.4	70.5	78.0	44.0	67.9
Qwen2.5-72B-it	74.7	96.0	90.4	84.2	83.5	61.3	69.2	54.4	66.6	63.0	<b>80.4</b>	82.1	71.7	73.4	79.9	80.7	43.3	75.3
Qwen2.5-7B-it	68.0	92.4	83.3	77.2	77.2	58.4	62.0	49.4	64.4	57.3	74.4	77.1	67.7	71.1	65.8	78.2	38.9	58.9
Llama3.1-70B-it	70.5	93.8	85.3	77.6	79.8	59.3	67.0	53.0	64.8	53.5	74.8	81.1	70.3	74.4	67.1	78.0	45.2	71.3
Llama3.1-8B-it	61.7	90.4	75.9	68.2	71.3	53.0	60.4	47.6	57.5	44.6	71.4	75.7	60.0	61.7	50.6	77.2	41.7	52.3
Gemma-2-27B-it	68.1	90.6	80.5	73.2	77.6	54.5	61.3	53.4	65.0	56.0	76.4	81.1	69.3	66.2	57.2	80.2	–	66.9
Gemma-2-9B-it	64.9	91.6	78.3	73.0	61.5	58.8	59.3	51.2	63.6	46.8	70.4	80.4	66.0	66.9	51.9	76.0	–	62.1
Mistral-7B-it	54.7	79.6	63.9	60.6	60.1	40.5	54.0	43.6	56.0	42.4	64.1	67.0	56.0	45.8	43.9	65.1	43.9	54.5
MoZi-qwen	64.9	93.8	83.3	77.0	66.1	58.2	64.2	50.6	58.0	41.8	67.8	76.4	68.0	64.3	56.1	79.0	43.9	57.1
DISC-LawLLM	52.8	79.0	65.3	67.6	60.1	54.5	52.0	40.8	60.4	31.3	60.1	64.8	53.7	45.1	28.2	71.2	–	35.3
Hanfei	40.1	63.0	46.4	51.8	45.4	39.8	47.3	30.8	45.6	33.9	40.9	49.2	42.7	28.6	18.9	48.8	–	29.5
DeepSeek-R1	73.9	96.0	<b>92.0</b>	87.6	80.8	64.9	<b>71.7</b>	53.6	64.6	<b>71.8</b>	78.1	<b>85.4</b>	63.3	78.2	67.2	82.0	<b>47.5</b>	74.3
DS-Qwen-7B	57.0	77.8	59.0	53.8	57.1	49.8	50.7	43.8	51.2	46.2	67.1	65.5	54.0	62.0	63.7	63.7	43.6	54.9
QwQ-32B	73.5	95.2	91.0	81.8	77.8	65.1	71.5	<b>57.4</b>	66.6	70.6	80.1	<b>85.4</b>	69.7	<b>82.1</b>	67.3	77.0	47.1	69.7

286  
 287 **Experimental Settings.** Inspired by previous benchmarks Team et al. (2025), we adopt five dis-  
 288 tinct evaluation settings for chat models: zero-shot, 1-shot, 2-shot, 3-shot, and Chain-of-Thought  
 289 (CoT). For reasoning models, we use only the zero-shot setting to ensure a fair comparison given  
 290 their limited prompt-handling flexibility. In few-shot settings, we randomly sample one to three  
 291 in-context examples (excluding the current test instance) using a fixed seed to ensure reproducibil-  
 292 ity. To ensure consistency and reproducibility, we set the temperature to 0.0 across all experiments.  
 293 The maximum input token limit is capped at 32k for reasoning models and 8k for chat models; for  
 294 models with shorter context windows, we use the maximum supported length. All the prompts used  
 295 are provided in Appendix G.

296 **Metrics.** We use accuracy as the primary evaluation metric for the most tasks. For IPC and CPC  
 297 classification tasks, we follow the evaluation strategy of HELM (Liang et al., 2022), using Ex-  
 298 act Match at **different granularity levels**: *Section*, *Class*, and *Subclass*. For generative tasks  
 299 such as abstract and claim generation, we evaluate model outputs using the F1 score of metrics:  
 300 BLEU (Papineni et al., 2002), ROUGE-L (Lin, 2004), and BERTScore (Zhang et al., 2019). Ad-  
 301 ditionally, inspired by the fine-grained error taxonomy in PatentEval, we propose **LLMScore**, a  
 302 multi-dimensional, automatic evaluation metric aligned with the LLM-as-a-judge paradigm (Liu  
 303 et al., 2023; Li et al., 2025a). LLMScore is used to assess the semantic and structural quality of  
 304 generated outputs, and we validate its consistency against human judgment. Details of metrics are  
 305 provided in Appendix H.

## 306 4.2 MAIN RESULTS

307 As shown in Table 3, 4, and 5, we present the main results under the zero-shot setting, while results  
 308 for the few-shot and CoT setting are provided in Figure 4a. More comprehensive results of IPBench  
 309 can be found in Appendix K.

## 312 4.3 ANALYSIS

314 **Disparity between IP-oriented and general-purpose models.** Surprisingly, general-purpose  
 315 models consistently outperform both law-oriented and IP-oriented models on IPBench. Although  
 316 MoZi-qwen, an IP-oriented model, outperforms the 2 law-oriented models DISC-LawLLM and  
 317 Hanfei, it still trails Qwen2.5-7B-it, by 3.1%. These results underscore a recurring issue in vertical  
 318 domain models: despite being optimized for specific applications, they tend to underperform on  
 319 domain-specific evaluations (Wang et al., 2024b; Hou et al., 2024; Li et al., 2024). This suggests  
 320 that domain-specific models must adopt improved strategies for learning domain knowledge without  
 321 sacrificing general-purpose capabilities.

322 **Model performance across different languages.** Model performance correlates strongly with  
 323 the primary training language of the model. Results across the Chinese and English subsets of IP-

324 Table 4: Main results of IPC/CPC Classification tasks. The best-performing model is in  
 325 **darker purple**, and the second best is in **lighter purple**.

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o	4.8	81.6	71.3	55.1	3.3	82.7	69.7	62.0
GPT-4o-mini	1.0	80.5	66.8	50.1	0.5	79.0	64.5	52.7
DeepSeek-V3	10.6	83.7	73.3	58.3	9.5	84.0	73.3	65.2
Qwen3	2.8	80.6	64.8	48.0	0.5	62.7	48.3	38.7
Qwen2.5-72B-it	4.9	82.4	70.4	55.2	2.5	81.5	69.5	60.7
Qwen2.5-7B-it	1.9	76.8	63.0	46.6	0.2	65.5	44.8	34.8
Llama3.1-70B-it	3.5	80.4	65.6	50.0	1.0	79.5	64.3	52.7
Llama3.1-8B-it	0.9	71.8	56.2	35.8	0.0	63.8	45.0	30.7
Gemma-2-27B-it	1.2	72.9	57.4	41.5	0.2	70.5	56.7	44.3
Gemma-2-9B-it	0.3	73.7	55.6	37.2	0.2	56.2	39.0	26.7
Mistral-7B-it	0.1	67.2	42.8	26.8	0.0	39.0	21.5	10.3
MoZi-qwen	0.6	38.8	29.6	20.3	0.0	8.5	3.1	1.8
DISC-LawLLM	0.0	68.2	47.2	28.3	0.0	31.0	23.4	11.5
Hanfei	0.0	11.7	2.0	0.1	0.0	0.8	0.0	0.0
DeepSeek-R1	10.8	85.8	74.7	59.3	8.5	82.5	71.2	63.2
DS-Qwen-7B	0.0	20.5	6.9	1.4	0.0	5.1	0.5	0.2
QwQ-32B	2.9	83.8	70.4	53.8	0.5	76.0	62.3	51.3

344  
 345 Table 5: Main results of generation tasks. The best-performing model is in **darker blue**, and  
 346 the second best is in **lighter blue**. R-L refers to ROUGE-L, BS refers to BERTScore, Tokens #  
 347 denotes the average number of tokens in the generated text, and DC # indicates the average number  
 348 of generated dependent claims.

Model	Abstract Generation (4-1)					Dependent Claim Generation (4-2)					
	BLEU	R-L	BS	LLMScore (1-10)	Tokens # (148.5)	BLEU	R-L	BS	LLMScore (1-10)	Tokens # (437.6)	DC # (5.2)
GPT-4o	17.7	31.1	89.3	8.42	271.4	18.9	26.5	88.8	6.63	647.8	6.5
GPT-4o-mini	23.4	31.9	89.6	8.05	218.1	20.3	28.3	88.4	6.37	478.1	6.5
DeepSeek-V3	19.6	28.3	89.0	8.38	246.1	19.1	26.8	89.0	7.45	691.7	14.9
Qwen2.5-72B-it	21.0	30.6	89.5	8.33	326.0	10.0	17.1	89.2	6.30	3790.9	69.1
Qwen2.5-7B-it	27.3	35.7	90.2	8.18	209.2	15.1	22.3	89.2	5.67	3511.3	45.7
Llama3.1-70B-it	31.0	38.2	90.4	7.98	226.5	16.0	23.8	88.1	5.67	2294.4	28.3
Llama3.1-8B-it	20.1	28.4	89.2	7.47	457.3	8.1	13.9	88.4	3.86	6287.9	90.8
Gemma-2-27B-it	19.7	27.5	88.9	7.64	193.3	15.2	22.6	87.3	5.98	582.3	3.3
Gemma-2-9B-it	21.6	29.4	89.0	7.91	219.3	14.7	23.2	87.1	5.55	511.9	6.4
Mistral-7B-it	20.2	27.4	89.4	7.49	361.7	7.2	11.7	88.0	3.42	6543.1	96.3
MoZi-qwen	31.2	51.0	90.4	7.73	316.4	16.3	34.4	89.0	4.81	5121.5	47.7
DeepSeek-R1	13.8	27.8	87.5	7.72	642.3	16.6	29.3	71.4	7.18	1302.9	19.1
DS-Qwen-7B	9.7	22.9	83.6	7.58	802.5	11.7	32.4	69.0	4.16	6096.9	54.1
QwQ-32B	16.6	32.0	87.9	8.51	1126.6	12.6	25.8	71.9	7.10	4997.7	41.8

364  
 365 Bench are provided in Figure 3a. DeepSeek-V3 achieves the highest accuracy on the Chinese subset  
 366 (78.7%), while GPT-4o leads on the English subset (73.2%). These findings highlight the impact  
 367 of legal system discrepancies across jurisdictions and the need for language models to recognize  
 368 and adapt to structural and contextual differences during inference, consistent with the observations  
 369 reported in IPEval (Wang et al., 2024b).

370  
 371 **Disparity between Chat Model and Reasoning Model.** In addition to chat models, we evaluate  
 372 3 reasoning-focused models, notably DeepSeek-R1. While these models do not achieve the highest  
 373 overall scores, they demonstrate superior performance on logically intensive tasks. For example,  
 374 in Task 2-3 (compensation calculation), DeepSeek-R1 surpasses the best-performing chat model,  
 375 DeepSeek-V3, by 5.7%. This task requires not only domain knowledge but also strong arithmetic  
 376 and logical reasoning skills. These findings highlight the need for future models to integrate both  
 377 intuitive (*System 1*) and analytical (*System 2*) capabilities, particularly in high-stakes, knowledge-  
 378 intensive domains such as IP.

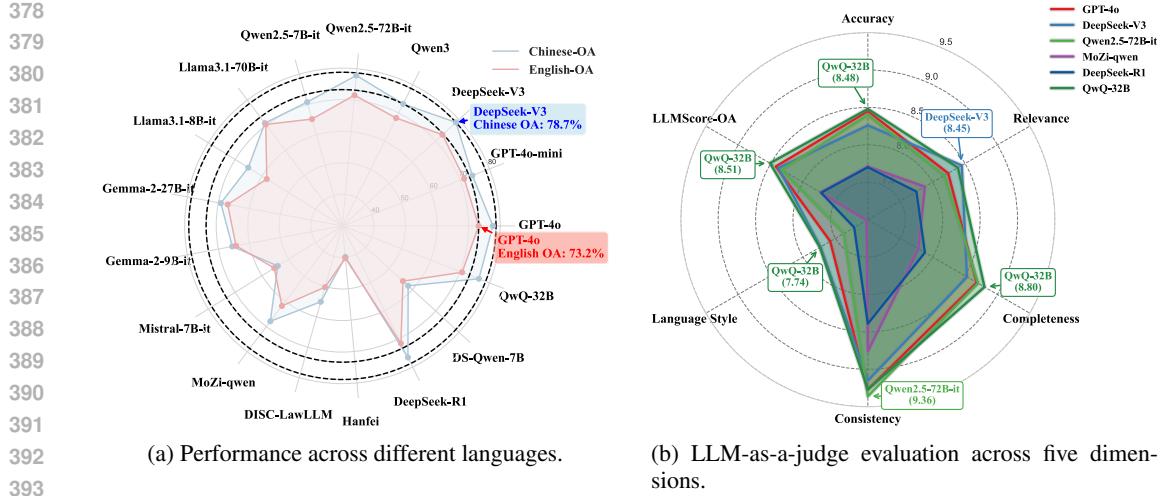


Figure 3: Comparative results: (a) performance across languages; (b) evaluation across fine-grained dimensions.

**Disaster in IPC/CPC exact match performance.** Performance on IPC/CPC classification tasks remains particularly weak. DeepSeek-R1 achieves the highest Exact Match score at 10.8%, followed by DeepSeek-V3 at 9.5%, while several models score as low as 0.0%. As the classification granularity increases -from Section to Class to Subclass to Exact Match -the difficulty also rises, given the increasingly specific technical distinctions required. These results reveal substantial limitations in current models' abilities to perform fine-grained classification and highlight the complexity of capturing structured taxonomies in patent law. Since IPC/CPC classification underpins many foundational applications in patent management, this represents a critical area for model improvement.

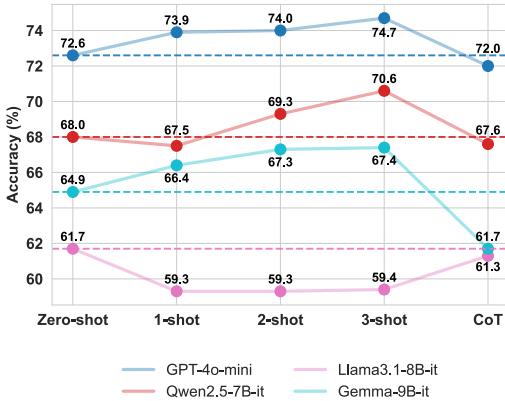
**Lack of fine-grained, interpretable automatic evaluation for IP-related generative tasks.** For these two generative tasks, there is a lack of fine-grained, interpretable automatic evaluation methods to provide more reliable results. Traditional metrics such as BLEU, ROUGE-L, and BERTScore are limited in their effectiveness and exhibit low consistency. To address this issue, we adopt an LLM-as-a-judge approach with five fine-grained dimensions, inspired by PatentEval's error taxonomy, and introduce **LLMScore** for more reliable evaluation. As shown in Table 6, LLMScore demonstrates significantly higher consistency with human judgments than other metrics, which is reflected in its higher Kendall, Spearman, and Pearson correlation coefficients, and lower  $p$ -values. We present detailed LLM-as-a-judge evaluations of generative tasks across five dimensions: *Accuracy*, *Relevance*, *Completeness*, *Consistency*, and *Language Style*, as illustrated in Figure 3b. Detailed LLMScore results are provided in Appendix K.4.

Table 6: Correlation of LLMScore with human judgments on Task 4-1 and Task 4-2 ( $p$ -value in parentheses).  $\uparrow$  Correlation coefficients,  $\downarrow p$ -value.

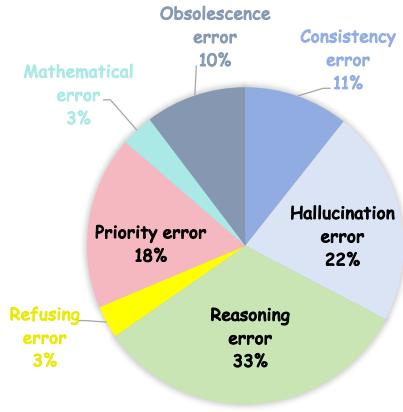
Metric	Task 4-1			Task 4-2		
	Kendall	Pearson	Spearman	Kendall	Pearson	Spearman
LLMScore	<b>0.22 (0.0005)</b>	<b>0.29 (0.0011)</b>	<b>0.32 (0.0003)</b>	<b>0.40 (0.0000)</b>	<b>0.65 (0.0000)</b>	<b>0.58 (0.0000)</b>
BLEU	0.17 (0.0042)	0.22 (0.0068)	0.23 (0.0046)	<b>0.40 (0.0000)</b>	0.47 (0.0000)	0.54 (0.0000)
ROUGE-L	0.15 (0.0123)	0.18 (0.0317)	0.20 (0.0154)	0.37 (0.0000)	0.51 (0.0000)	0.50 (0.0000)
BERTScore	0.10 (0.0746)	0.16 (0.0519)	0.14 (0.0847)	0.05 (0.3680)	0.09 (0.2950)	0.08 (0.3494)

**Results and analysis of few-shot prompting.** As shown in Figure 4a, the performance of models on IPBench generally improves as the number of shots increases, reflecting a positive correlation between in-context learning and task performance -except for Llama3.1-8B, which does not exhibit this trend. This observation is consistent with prior studies (Li et al., 2024; Wang et al., 2024b), which show that the effectiveness of few-shot prompting varies significantly across model architec-

432 tures. These findings suggest that few-shot learning may not be a universally effective strategy for  
 433 injecting domain-specific knowledge for complex domains.  
 434



(a) Performance under different prompt settings.



(b) Error distribution of GPT-4o-mini's responses.

Figure 4: Few-shot performance and error distribution.

453 **Results and analysis of CoT prompting.** As shown in Figure 4a, all models experience a slight  
 454 performance decline when using CoT prompting. Upon deeper analysis of the error cases, we observe that models generate not only the final answer but also a reasoning  
 455 trajectory. This additional reasoning, while intended to aid logical flow, often introduces new sources  
 456 of error or distracts from more intuitive solutions. These results align with recent findings (Zheng  
 457 et al., 2025; Fan et al., 2025), which suggest that CoT prompting may conflict with the natural inferential  
 458 preferences of language models -especially in tasks relying more on memorization or domain  
 459 recall than on abstract reasoning. This is further reflected in the observation that reasoning models  
 460 do not outperform chat models on IPBench, despite conducting longer reasoning sequences during  
 461 inference.  
 462

#### 464 4.4 ERROR ANALYSIS

466 To gain deeper insight into model limitations, we perform a qualitative error analysis. We randomly  
 467 selected 300 incorrect responses generated by GPT-4o-mini under the CoT setting across all IPBench  
 468 tasks. These samples were manually reviewed and annotated by expert evaluators. As shown in  
 469 Figure 4b, the errors are categorized into seven types: **Consistency Error**, **Hallucination Error**,  
 470 **Reasoning Error**, **Refusing Error**, **Priority Error**, **Mathematical Error**, and **Obsolescence Error**.  
 471 Among these, **Reasoning Error** is the most frequent, accounting for 33% of the total. This error  
 472 analysis is crucial for gaining deeper insights into the model's capabilities in the IP domain and for  
 473 revealing potential directions for future research. More details of error analysis and case study are  
 474 in Appendix L, M and N. We also provide more discuss and limitations in Appendix J, B.  
 475

## 476 5 CONCLUSION

478 We introduce the first comprehensive IP task taxonomy and present IPBench, a bilingual bench-  
 479 mark comprising 20 tasks and 10,374 test instances, covering both technical-legal and generation-  
 480 comprehension evaluations. Our experiments show that even the best-performing model, DeepSeek-  
 481 V3, achieves only a 75.8% score. We observe that current models, including IP-oriented ones, still  
 482 lag significantly behind powerful closed-source models, highlighting the need for improved domain-  
 483 specific learning approaches. Our extensive performance analysis, error analysis and case study  
 484 provide a comprehensive insight in models' IP knowledge and capabilities. We are committed to  
 485 continuously expanding IPBench to foster advancements in both the IP domain and NLP research,  
 providing meaningful guidance for the integration of LLMs into specialized vertical fields.

486 REFERENCES  
487

488 OpenAI Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Floren-  
489 cia Leoni Aleman, and Diogo Almeida et al. Gpt-4 technical report. 2023. URL <https://api.semanticscholar.org/CorpusID:257532815>.

490

491 Zilong Bai, Ruiji Zhang, Linqing Chen, Qijun Cai, Yuan Zhong, Cong Wang, Yan Fang, Jie Fang,  
492 Jing Sun, Weikuan Wang, et al. Patentgpt: A large language model for intellectual property. *arXiv*  
493 *preprint arXiv:2404.18255*, 2024.

494

495 Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal,  
496 Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are  
497 few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

498

499 Jianlyu Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. M3-embedding:  
500 Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge  
501 distillation. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Findings of the Asso-  
502 ciation for Computational Linguistics: ACL 2024*, pp. 2318–2335, Bangkok, Thailand, August  
503 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.137. URL  
<https://aclanthology.org/2024.findings-acl.137>.

504

505 DeepSeek-AI, Aixin Liu, Bei Feng, Bing Xue, and Bingxuan Wang et al. Deepseek-v3 technical  
506 report. *CoRR*, abs/2412.19437, 2024. doi: 10.48550/ARXIV.2412.19437. URL <https://doi.org/10.48550/arXiv.2412.19437>.

507

508 DeepSeek-AI, Daya Guo, Dejian Yang, and Haowei Zhang et al. Deepseek-r1: Incentivizing  
509 reasoning capability in llms via reinforcement learning. *ArXiv*, abs/2501.12948, 2025. URL  
<https://api.semanticscholar.org/CorpusID:275789950>.

510

511 Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, and Abhishek Kadian et al. The llama 3 herd  
512 of models. *ArXiv*, abs/2407.21783, 2024. URL <https://api.semanticscholar.org/CorpusID:271571434>.

513

514 EPO. Guidelines for examination in the european patent office. pp. 1 volume ;, 1994. URL <http://tind.wipo.int/record/17847>.

515

516

517 Caspar J Fall, Atilla Törcsvári, Karim Benzineb, and Gabor Karetka. Automated categorization in  
518 the international patent classification. In *Acm Sigir Forum*, volume 37, pp. 10–25. ACM New  
519 York, NY, USA, 2003.

520

521 Chenrui Fan, Ming Li, Lichao Sun, and Tianyi Zhou. Missing premise exacerbates overthinking:  
522 Are reasoning models losing critical thinking skill?, 2025. URL <https://arxiv.org/abs/2504.06514>.

523

524 Wanwei He, Jiabao Wen, Lei Zhang, Hao Cheng, Bowen Qin, Yunshui Li, Feng Jiang, Junying Chen,  
525 Benyou Wang, and Min Yang. Hanfei-1.0. <https://github.com/siat-nlp/HanFei>,  
526 2023.

527

528 Jinchang Hou, Chang Ao, Haihong Wu, Xiangtao Kong, Zhigang Zheng, Daijia Tang, Chengming  
529 Li, Xiping Hu, Ruijing Xu, Shiwen Ni, and Min Yang. E-EVAL: A comprehensive Chinese  
530 k-12 education evaluation benchmark for large language models. In Lun-Wei Ku, Andre Martins,  
531 and Vivek Srikumar (eds.), *Findings of the Association for Computational Linguistics: ACL  
532 2024*, pp. 7753–7774, Bangkok, Thailand, August 2024. Association for Computational Linguistics.  
533 doi: 10.18653/v1/2024.findings-acl.462. URL <https://aclanthology.org/2024.findings-acl.462>.

534

535 OpenAI Aaron Hurst, Adam Lerer, and Adam P. Goucher et al. Gpt-4o system card. *ArXiv*,  
536 abs/2410.21276, 2024. URL <https://api.semanticscholar.org/CorpusID:273662196>.

537

538 Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec  
539 Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. Openai o1 system card. *arXiv  
preprint arXiv:2412.16720*, 2024.

540 Fengqing Jiang. Identifying and mitigating vulnerabilities in llm-integrated applications. Master's  
 541 thesis, University of Washington, 2024.

542

543 Lekang Jiang and Stephan Goetz. Natural language processing in patents: A survey. *arXiv preprint*  
 544 *arXiv:2403.04105*, 2024.

545 Lekang Jiang, Pascal A Scherz, and Stephan Goetz. Patent-cr: A dataset for patent claim re-  
 546 vision. *ArXiv*, abs/2412.02549, 2024. URL <https://api.semanticscholar.org/CorpusID:274445928>.

547

548 Valentin Knappich, Simon Razniewski, Anna Hätty, and Annemarie Friedrich. Pap2pat: Towards  
 549 automated paper-to-patent drafting using chunk-based outline-guided generation. *arXiv preprint*  
 550 *arXiv:2410.07009*, 2024.

551

552 Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large  
 553 language models are zero-shot reasoners. *Advances in neural information processing systems*,  
 554 35:22199–22213, 2022.

555

556 Jieh-Sheng Lee and Jieh Hsiang. Patent classification by fine-tuning bert language model. *World*  
 557 *Patent Information*, 61:101965, 2020.

558

559 Haitao Li, You Chen, Qingyao Ai, WU Yueyue, Ruizhe Zhang, and LIU Yiqun. Lexeval: A com-  
 560 prehensive chinese legal benchmark for evaluating large language models. In *The Thirty-eight*  
 561 *Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024.

562

563 Haitao Li, Jiaying Ye, Yiran Hu, Jia Chen, Qingyao Ai, Yueyue Wu, Junjie Chen, Yifan Chen, Cheng  
 564 Luo, Quan Zhou, et al. Casegen: A benchmark for multi-stage legal case documents generation.  
 565 *arXiv preprint arXiv:2502.17943*, 2025a.

566

567 Shaobo Li, Jie Hu, Yuxin Cui, and Jianjun Hu. Deeppatent: patent classification with convolutional  
 568 neural networks and word embedding. *Scientometrics*, 117(2):721–744, 2018.

569

570 Xiaoxi Li, Jiajie Jin, Yujia Zhou, Yuyao Zhang, Peitian Zhang, Yutao Zhu, and Zhicheng Dou.  
 571 From matching to generation: A survey on generative information retrieval. *ACM Trans. Inf.*  
 572 *Syst.*, March 2025b. ISSN 1046-8188. doi: 10.1145/3722552. URL <https://doi.org/10.1145/3722552>. Just Accepted.

573

574 Percy Liang, Rishi Bommasani, and Tony Lee et al. Holistic evaluation of language models. *CoRR*,  
 575 abs/2211.09110, 2022. doi: 10.48550/ARXIV.2211.09110. URL <https://doi.org/10.48550/arXiv.2211.09110>.

576

577 Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization*  
 578 *Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguis-  
 579 *tics*. URL <https://aclanthology.org/W04-1013/>.

580

581 Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. G-eval: NLG  
 582 evaluation using gpt-4 with better human alignment. In Houda Bouamor, Juan Pino, and Ka-  
 583 lika Bali (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language*  
 584 *Processing*, pp. 2511–2522, Singapore, December 2023. Association for Computational Linguis-  
 585 *tics*. doi: 10.18653/v1/2023.emnlp-main.153. URL <https://aclanthology.org/2023.emnlp-main.153/>.

586

587 Qiang Lu, Amanda F. Myers, and Scott Beliveau. USPTO Patent Prosecution Research Data:  
 588 *Unlocking Office Action Traits*. Technical report, United States Patent and Trademark Office  
 589 (USPTO), 2017.

590

591 Shiwen Ni, Minghuan Tan, Yuelin Bai, Fuqiang Niu, Min Yang, Bowen Zhang, Ruifeng Xu, Xiao-  
 592 jun Chen, Chengming Li, and Xiping Hu. Mozip: A multilingual benchmark to evaluate large  
 593 language models in intellectual property. In *Proceedings of the 2024 Joint International Confer-  
 594 ence on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*,  
 595 pp. 11658–11668, 2024.

594 Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic  
 595 evaluation of machine translation. In Pierre Isabelle, Eugene Charniak, and Dekang Lin (eds.),  
 596 *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp.  
 597 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguis-  
 598 tics. doi: 10.3115/1073083.1073135. URL <https://aclanthology.org/P02-1040/>.

599 Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language  
 600 models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

602 Gemma Team Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, and Cassidy Hardin et al.  
 603 Gemma 2: Improving open language models at a practical size. *ArXiv*, abs/2408.00118, 2024.  
 604 URL <https://api.semanticscholar.org/CorpusID:270843326>.

605 Eva Sharma, Chen Li, and Lu Wang. Bigpatent: A large-scale dataset for abstractive and coherent  
 606 summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational  
 607 Linguistics*, pp. 2204–2213, 2019.

609 Mirac Suzgun, Luke Melas-Kyriazi, Suproteem Sarkar, Scott D Kominers, and Stuart Shieber. The  
 610 harvard uspto patent dataset: A large-scale, well-structured, and multi-purpose corpus of patent  
 611 applications. *Advances in neural information processing systems*, 36:57908–57946, 2023.

613 M-A-P Team, Xinrun Du, Yifan Yao, and Kaijing Ma et al. Supergpqa: Scaling llm evaluation  
 614 across 285 graduate disciplines, 2025. URL <https://arxiv.org/abs/2502.14739>.

615 Qwen Team. Qwen3, April 2025a. URL <https://qwenlm.github.io/blog/qwen3/>.

617 Qwen Team. Qwq-32b: Embracing the power of reinforcement learning, March 2025b. URL  
 618 <https://qwenlm.github.io/blog/qwq-32b/>.

619 USTPO. Manual of patent examining procedure. pp. 4 v. (loose-leaf), 2024b. URL <http://tind.wipo.int/record/41638>. This resource was extracted from USPTO.gov.

622 Juanyan Wang, Sai Krishna Reddy Mudhiganti, and Manali Sharma. Patentformer: A novel method  
 623 to automate the generation of patent applications. In *Proceedings of the 2024 Conference on  
 624 Empirical Methods in Natural Language Processing: Industry Track*, pp. 1361–1380, 2024a.

626 Qiyao Wang, Jianguo Huang, Shule Lu, Yuan Lin, Kan Xu, Liang Yang, and Hongfei Lin. Ipeval:  
 627 A bilingual intellectual property agency consultation evaluation benchmark for large language  
 628 models. *arXiv preprint arXiv:2406.12386*, 2024b.

629 Qiyao Wang, Shiwen Ni, Huaren Liu, Shule Lu, Guhong Chen, Xi Feng, Chi Wei, Qiang Qu,  
 630 Hamid Alinejad-Rokny, Yuan Lin, et al. Autopatent: A multi-agent framework for automatic  
 631 patent generation. *arXiv preprint arXiv:2412.09796*, 2024b.

633 Norman L Webb. Depth-of-knowledge levels for four content areas. *Language Arts*, 28(March):  
 634 1–9, 2002.

635 WIPO. What is intellectual property? pp. 1 PDF (24 pages) :, 2020a. doi: <https://doi.org/10.34667/tind.42176>. URL <http://tind.wipo.int/record/42176>.

638 An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, and Bowen Yu et al. Qwen2.5  
 639 technical report. *CoRR*, abs/2412.15115, 2024. doi: 10.48550/ARXIV.2412.15115. URL  
 640 <https://doi.org/10.48550/arXiv.2412.15115>.

641 Shengbin Yue, Wei Chen, Siyuan Wang, Bingxuan Li, Chenchen Shen, Shujun Liu, Yuxuan Zhou,  
 642 Yao Xiao, Song Yun, Wei Lin, Xuanjing Huang, and Zhongyu Wei. Disc-lawllm: Fine-tuning  
 643 large language models for intelligent legal services, 2023.

645 Shengbin Yue, Shujun Liu, Yuxuan Zhou, Chenchen Shen, Siyuan Wang, Yao Xiao, Bingxuan Li,  
 646 Yun Song, Xiaoyu Shen, Wei Chen, et al. Lawllm: Intelligent legal system with legal reasoning  
 647 and verifiable retrieval. In *International Conference on Database Systems for Advanced Applica-  
 tions*, pp. 304–321. Springer, 2024.

648 Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. Bertscore:  
 649 Evaluating text generation with bert. *ArXiv*, abs/1904.09675, 2019. URL <https://api.semanticscholar.org/CorpusID:127986044>.  
 650

651  
 652 Tianshi Zheng, Yixiang Chen, Chengxi Li, Chunyang Li, Qing Zong, Haochen Shi, Baixuan Xu,  
 653 Yangqiu Song, Ginny Y. Wong, and Simon See. The curse of cot: On the limitations of chain-of-  
 654 thought in in-context learning, 2025. URL <https://arxiv.org/abs/2504.05081>.  
 655

656  
 657 You Zuo, Kim Gerdes, Éric Clergerie, and Benoît Sagot. PatentEval: Understanding errors in patent  
 658 generation. In Kevin Duh, Helena Gomez, and Steven Bethard (eds.), *Proceedings of the 2024*  
 659 *Conference of the North American Chapter of the Association for Computational Linguistics: Human*  
 660 *Language Technologies (Volume 1: Long Papers)*, pp. 2687–2710, Mexico City, Mexico,  
 661 June 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.naacl-long.147.  
 662 URL <https://aclanthology.org/2024.naacl-long.147/>.  
 663  
 664

## 665 APPENDIX

### 666 A THE USE OF LARGE LANGUAGE MODELS (LLMs)

667  
 668 In accordance with the policy on the use of Large Language Models, we clarify that in this work  
 669 LLMs were employed exclusively for improving the presentation of the manuscript, such as cor-  
 670 recting grammatical errors, enhancing clarity, and refining writing. The research design, conceptual  
 671 development, and analytical contributions were made solely by the authors.  
 672

### 673 B LIMITATIONS

674  
 675 While IPBench represents a significant step forward in evaluating large language models for intel-  
 676 lectual property tasks, several limitations remain.  
 677

678 First, due to the jurisdiction-specific nature of intellectual property law, the current version of IP-  
 679 Bench focuses primarily on the legal frameworks of the United States and mainland China. This  
 680 restricts its global applicability, as key differences in legal definitions, procedural structures, and  
 681 enforcement standards exist across countries. Expanding the benchmark to include legal systems  
 682 from jurisdictions such as the European Union, Japan, and Korea would enhance its cross-cultural  
 683 robustness and relevance.  
 684

685 Second, resource constraints limited our evaluation to four reasoning models. While these include  
 686 some of the most advanced publicly available systems, we were unable to include proprietary models  
 687 such as OpenAI’s o1 (Jaech et al., 2024) and o3 series due to prohibitive costs. As pricing structures  
 688 evolve and research access improves, future iterations of IPBench will aim to incorporate a wider  
 689 array of state-of-the-art reasoning models.  
 690

691 Third, intellectual property remains an underexplored vertical domain in large language model re-  
 692 search. Currently, MoZi is the only publicly available IP-specific model, and thus the only one  
 693 benchmarked in this study. The development and release of more open-source IP-oriented models  
 694 will be essential for driving progress in this field and enabling more comprehensive comparisons in  
 695 future studies.  
 696

697 Lastly, although we propose *LLMScore*, a fine-grained, interpretable, and high-consistency evalua-  
 698 tion metric grounded in the LLM-as-a-judge paradigm, there is still room for improvement. Future  
 699 work should focus on minimizing bias and improving the robustness of automatic evaluation meth-  
 700 ods across diverse model architectures, task types, and cultural contexts.  
 701

We view these limitations not only as constraints, but also as valuable directions for extending the  
 scope, depth, and impact of IPBench in future work.

702 **C DATA USAGE STATEMENT**  
 703

704 In developing IPBench, all data are collected exclusively from open and publicly available sources.  
 705 We strictly adhered to all relevant copyright and licensing regulations. Any data originating from  
 706 websites or platforms that prohibit copying, redistribution, or automated crawling are explicitly  
 707 excluded from use. Furthermore, we confirm that all data are used solely for academic and research  
 708 purposes, and not for any commercial applications. We are committed to upholding responsible data  
 709 usage and transparency in our research practices. Future updates of IPBench will continue to follow  
 710 the same principles and remain fully open to academic scrutiny and community feedback.  
 711

712 **D TAXONOMY AND TASK DETAILS**  
 713

714 **D.1 TAXONOMY DETAILS**  
 715

716 **Information Processing.** In the first level of the taxonomy, we replace Recall and Reproduction  
 717 with Information Processing, which encompasses the legal concepts, clauses, evolution, and typical  
 718 case knowledge of various IP mechanisms. It also includes real-world applications such as patent  
 719 classification, IP element identification, and process guidance, requiring models to memorize dif-  
 720 ferent concepts, along with the procedures executed in real-world scenarios. Our expert-annotated  
 721 memory-type tasks are inspired by those in LexEval (Li et al., 2024) but differ significantly from it  
 722 across various IP mechanisms, using accuracy as the evaluation metric. The IP element identification  
 723 task focuses on identifying key elements in a case, such as claim coverage in patent infringement.  
 724 Previous work has made significant progress in patent classification (Li et al., 2018; Lee & Hsiang,  
 725 2020; Fall et al., 2003), which has been adopted by IP offices in many countries. However, these  
 726 models are task-specific and lack the strong generalization ability of LLMs. Our patent classification  
 727 task consists of two types: International Patent Classification (IPC) and Cooperative Patent Classi-  
 728 fication (CPC). We aim to accomplish these tasks within a single model, enabling it to distinguish  
 729 both differences within the same classification system and across different classification systems.  
 730 We adopt the top-prediction scheme, following Fall et al. (Fall et al., 2003), to compare the top  
 731 predicted category with the label for an Exact Match (Liang et al., 2022) in the main IPC symbol,  
 732 and CPC. This setup increases the task difficulty for LLMs, requiring models to be familiar with  
 733 classification rules.  
 734

735 **Logical Reasoning.** At the second level of the taxonomy, we focus on examining a model’s ability  
 736 to apply memorized concepts and utilize logical reasoning to provide insights into both text analysis  
 737 and mathematical calculations. One of the important roles of IP is to protect inventors’ rights from  
 738 infringement. Therefore, we define the tasks of Infringement Behavior Determination and Compen-  
 739 sation Calculation. To complete these two tasks, models need to analyze the background of cases  
 740 to identify infringement behavior and apply relevant laws to determine the appropriate penalties.  
 741 Apart from the legal aspect, we introduce Patent Technology Forecasting, Patent Valuation, and Fact  
 742 Checking to evaluate models’ ability in information mining and conditional reasoning. As we men-  
 743 tioned, our IPBench consists of different IP mechanisms. We specifically introduce a novel task  
 744 called Trade Secret Requirements, which differs from Infringement Behavior Determination. This  
 745 task focuses on trade secret rights, requiring the model to determine whether a situation meets the  
 746 confidentiality requirements of trade secrets.  
 747

748 **Discriminant Evaluation.** At this level, we focus on evaluating models’ understanding of IP in-  
 749 domain texts, particularly patent documents, as well as their ability to perform discriminative tasks  
 750 such as rights attribution. As an important part of IP management activities, as mentioned before,  
 751 IP offices face a massive volume of patent applications. Determining the quality of an application  
 752 requires assessing its patentability based on four aspects outlined in the Manual of Patent Examining  
 753 Procedure (MPEP) (USTPO, 2024b; EPO, 1994): utility, non-obviousness, statutory subject matter,  
 754 and novelty. We aim to evaluate whether current LLMs can assist patent examiners in reducing time  
 755 costs within a single model. To achieve this, we introduce three tasks: Patent Document Proofread-  
 756 ing, Patent Validity Identification, and Patent Match. LLMs’ output mechanisms are not well-suited  
 757 for retrieval-based approaches, and Li et al. (2025b) introduced a novel perspective on matching  
 758 documents through a generative approach. Based on this insight, the Patent Match task draws in-

756 spiration from the corresponding task in MoZIP (Ni et al., 2024). We sample 1000 datapoints from  
 757 MoZIP in both English and Chinese and require expert annotation for detailed examination.  
 758

759 Apart from the evaluation of in-domain text, we introduce one real-world common task for evaluating  
 760 models’ discrimination ability: Rights Attribution Analysis. The Rights Attribution Analysis  
 761 task requires the model to infer the legal rights holder of a specific intellectual property based on the  
 762 context of IP creation, legal agreements such as contract terms and confidentiality agreements, and  
 763 judicial precedents within the legal framework. At last, we extend HUPD’s (Suzgun et al., 2023)  
 764 Patent Acceptance Prediction task into a more comprehensive Patent Application Examination task,  
 765 leveraging the USPTO Office Action Dataset (Lu et al., 2017). In this task, the model is required  
 766 to determine whether a given patent application should be accepted or rejected. Additionally, we  
 767 provide stepwise examination actions for an interpretable examination process, which can be used  
 768 in future work to construct a reliable examination system.

769 **Creative Generation.** At the final level of our IPBench, we focus on evaluating the models’ ability  
 770 to extract critical information, convert between different linguistic styles, and generate new content.  
 771 Previous works such as BigPatent (Sharma et al., 2019), Patent-CR (Jiang et al., 2024), and  
 772 PatentEval (Zuo et al., 2024) focus on specific types of content for patent generation. We draw inspiration  
 773 from some of their tasks and extend their scope to include both Chinese and English. All the  
 774 data used in Abstract Generation, Claim Generation, Sequential Claim Generation are sourced from  
 775 the latest patents, ensuring no data leakage and distinguishing our dataset from existing ones. At  
 776 last, we introduce a novel task called Design-Around Solution Generation, which evaluates whether  
 777 models can generate innovative solutions that avoid duplication of existing ones. This capability is  
 778 crucial in strategic patent planning. Given the distinct characteristics of the tasks at this level, we  
 779 use accuracy as the metric for Language Simplification and Design-Around Solution Generation.  
 780 For the other three generative tasks, we note that PatentEval (Zuo et al., 2024) provides an LLM-  
 781 based evaluation method for claim generation. However, this approach relies on the assumption  
 782 that the employed LLMs are sufficiently capable. Moreover, for other types of content, no superior  
 783 evaluation method currently exists. We adopt a combination of automated evaluation and human  
 784 assessment. For automated evaluation, we use n-gram-based metrics such as BLEU (Papineni et al.,  
 785 2002) and ROUGE (Lin, 2004), along with the semantic metric BERTScore (Zhang et al., 2019), and  
 786 analyze their consistency with human evaluation to enhance result interpretation. We will explore  
 787 better evaluation methods in future work, especially for patent generation, which involves complex  
 788 technical and legal content.

788 It is important to note that the abstract generation evaluation in BigPatent (Sharma et al., 2019) is  
 789 based on converting only the first 400 words of a patent’s description into an abstract, a limitation  
 790 imposed by the context length of language models at the time. In our IPBench, we evaluate models  
 791 on their ability to generate abstracts from the entire description, assessing their long-context  
 792 understanding and summarization capabilities for complex patent documents.

## 793 D.2 TASK DEFINITION

### 794 D.2.1 INFORMATION PROCESSING

795 **Task 1-1: Legal Concept Memory** Legal Concept Memory refers to the ability to precisely memorize  
 796 and recall foundational definitions within the intellectual property domain. These definitions,  
 797 such as those of patents, copyrights, trademarks, and trade secrets, are grounded in authoritative  
 798 legal frameworks and scholarly interpretations that constitute the foundation of intellectual  
 799 property law. When given a concept name or contextual description, LLMs must retrieve the precise  
 800 legal definition, scope, and jurisdictional boundaries as codified in statutes such as China’s Patent  
 801 Law and Copyright Law, as well as relevant international agreements, purely from their intrinsic  
 802 knowledge without relying on external databases or tools.

803 **Task 1-2: Legal Clause Memory** Legal Clause Memory requires the precise memorization and  
 804 retrieval of specific legal provisions, including their exact article numbers and textual content. These  
 805 clauses, drawn from authoritative legal codes such as China’s Criminal Law, Civil Code, and In-  
 806 tellectual Property Law, define rights, obligations, penalties, or procedural rules within statutory  
 807 frameworks. When provided with an article number (e.g., Article 217 of China’s Copyright Law) or

810 a contextual description of a legal scenario, LLMs must accurately recall the verbatim wording and  
 811 scope of the corresponding clause.  
 812

813 **Task 1-3: Legal Evolution** Legal Evolution refers to the ability to accurately memorize and an-  
 814alyze the revision history of legal texts, including the tracking of changes in specific clauses across  
 815 different versions of statutes, regulations, or international treaties. This capability requires models  
 816 to retain knowledge of amendments, such as updates to China’s Patent Law, and to systematically  
 817 compare the wording, scope, and intent of clauses before and after revisions.  
 818

819 **Task 1-4: Typical Case Memory** Typical Case Memory requires the memorization of landmark  
 820 intellectual property cases, including their judicial outcomes, factual details, and legal reasoning.  
 821 These cases, such as high-profile patent disputes, copyright infringement rulings, or trademark op-  
 822 position decisions, establish precedents that shape the interpretation and enforcement of IP law.  
 823 When provided with a case name, jurisdiction, or factual scenario, models must accurately recall  
 824 the judgment summary, key legal arguments, cited statutes, and contextual factors, without using an  
 825 external database or retrieval tool.  
 826

826 **Task 1-5: Patent Classification** Patent Classification involves the capability to automatically as-  
 827 sign International Patent Classification (IPC) or Cooperative Patent Classification (CPC) codes based  
 828 on the technical content of patent documents. This task requires models to analyze patent texts, in-  
 829 cluding titles and abstracts to identify the core inventions, technological domains, and functional  
 830 features, then map them to hierarchical classification codes. This task evaluates the model’s capabili-  
 831 ties across three hierarchical levels: Section, Class, and Subclass. A distribution table for the section  
 832 level as shown in Table 7.  
 833

Table 7: International Patent Classification (IPC) Sections

Section	Content
A	Human Necessities
B	Performing Operations; Transporting
C	Chemistry; Metallurgy
D	Textiles; Paper
E	Fixed Constructions
F	Mechanical Engineering; Lighting; Heating; Weapons; Blasting
G	Physics
H	Electricity

Table 8: Cooperative Patent Classification (CPC) Sections

Section	Content
A	Human Necessities
B	Operations and Transport
C	Chemistry and Metallurgy
D	Textiles and Paper
E	Fixed Constructions
F	Mechanical Engineering and Lighting
G	Physics
H	Electricity
Y	Emerging Technologies

857 **Task 1-6: IP Element Identification** IP Element Identification entails detecting and categorizing  
 858 intellectual property components such as patent claims, trademark-protected assets, copyrighted ma-  
 859 terial, or trade secret identifiers within legal disputes, technical specifications, or commercial con-  
 860 tracts. This task requires models to analyze textual data to identify legally protected innovations,  
 861 distinctive brand assets, and ownership claims, while ensuring alignment with statutory definitions.  
 862

863 **Task 1-7: Process Guidance** Process Guidance focuses on delivering structured knowledge of  
 864 intellectual property application procedures, covering legal requirements, technical documentation

864 standards, and jurisdictional workflows. This task requires models to provide step-by-step guidance  
 865 on processes such as conducting patent or trademark searches, drafting application materials,  
 866 navigating submission procedures, and ensuring compliance with examination regulations.  
 867

### 868 D.2.2 LOGICAL REASONING

870 The Logical Reasoning level is designed to evaluate the capability of large language models (LLMs)  
 871 to perform multi-dimensional legal and technical reasoning within the complex framework of  
 872 intellectual property (IP) law and textual analysis. This layer tests the model’s ability to analyze,  
 873 interpret, and apply intersecting legal rules. It focuses on assessing whether models can synthesize  
 874 statutory provisions, case law precedents, and technical domain knowledge to reach legally sound  
 875 conclusions such as identifying infringement risks, resolving conflicts between overlapping rights,  
 876 or predicting litigation outcomes based on factual scenarios.  
 877

**878 Task 2-1: Patent Technology Forecasting** Patent Technology Forecasting involves analyzing the  
 879 technical features of patents such as claims, innovation summaries, and domain-specific terminology  
 880 to predict future technological trajectories and potential application areas. This task requires  
 881 models to identify emerging trends, interconnected technical fields, and latent innovation pathways  
 882 within patent datasets, enabling the projection of how core inventions might evolve or intersect with  
 883 adjacent industries.

**884 Task 2-2: Infringement Behavior Determination** Infringement Behavior Determination focuses  
 885 on identifying acts that constitute violations of intellectual property rights. It involves analyzing  
 886 the legally protected scope of patents, copyrights, trademarks, or other IP types, and comparing  
 887 them with suspected infringing products, services, or content to determine whether an intellectual  
 888 property infringement has occurred. This task requires models to evaluate technical equivalence,  
 889 trademark similarity, or substantial similarity in copyrighted works, while accurately applying the  
 890 relevant statutory criteria to determine whether an intellectual property infringement has occurred.  
 891

**892 Task 2-3: Compensation Calculation** Compensation Calculation focuses on determining statutory  
 893 damages for intellectual property infringement by analyzing the severity, scope, and economic  
 894 impact of the violation. This task requires models to perform mathematical reasoning and calculation,  
 895 taking into account factors such as the rights holder’s actual losses, reasonable licensing fees,  
 896 and statutory limits. Additionally, models must incorporate contextual elements such as the duration  
 897 of infringement, geographic scope, and the presence of malicious intent to arrive at a legally  
 898 grounded and quantitatively sound compensation estimate.

**899 Task 2-4: Patent Valuation** Patent Valuation entails evaluating the value trajectory of a patent by  
 900 synthesizing its technical merit, market viability, and legal robustness. This task requires models to  
 901 analyze technical claims, market analysis reports, and legal histories to project trends such as value  
 902 appreciation, obsolescence risks, or licensing potential.  
 903

**904 Task 2-5: Trade Secret Requirements** Trade Secret Requirements assesses whether a given sce-  
 905 nario satisfies the legal criteria for trade secret protection under statutory frameworks such as China’s  
 906 Anti-Unfair Competition Law and the U.S. Defend Trade Secrets Act (DTSA). This task requires  
 907 models to verify three core elements: the existence of secrecy, the presence of commercial value,  
 908 and the implementation of reasonable confidentiality measures.  
 909

### 910 D.2.3 DISCRIMINANT EVALUATION

**911 Task 3-1: Patent Document Proofreading** Patent Document Proofreading involves identifying  
 912 formatting deviations and logical inconsistencies within patent specifications, claims, and technical  
 913 descriptions to ensure compliance with statutory drafting standards. This task requires models to  
 914 detect issues such as mismatched section numbering, non-compliant claim dependencies, contradic-  
 915 tory technical descriptions, and deviations from jurisdiction-specific filing guidelines.  
 916

**917 Task 3-2: Patent Validity Identification** Patent Validity Identification involves assessing whether  
 918 a patent satisfies the statutory criteria of novelty, inventiveness (non-obviousness), and practical  
 919 applicability (utility) by analyzing its technical disclosures in light of relevant prior art. This task

918 requires models to evaluate patent texts, including claims and specifications, against existing tech-  
 919 nologies to determine if the invention is new, involves an inventive step, and has industrial applica-  
 920 bility.

921  
 922 **Task 3-3: Patent Match** Patent Match involves identifying the most relevant patents from a candi-  
 923 date pool based on technical, legal, and contextual alignment with a query patent. This task requires  
 924 models to analyze technical features and semantic similarity to rank patents by relevance. This task  
 925 is inspired by MoZIP (Ni et al., 2024).

926  
 927 **Task 3-4: Rights Attribution Analysis** Rights Attribution Analysis involves determining the le-  
 928 gitimate rights holder in intellectual property ownership disputes by analyzing legal documents,  
 929 contractual agreements, and contextual evidence. This task requires models to evaluate factors such  
 930 as invention ownership under employment relationships, joint authorship claims in copyright cases,  
 931 or trademark transfer agreements, while reconciling conflicting claims based on statutory frame-  
 932 works.

933  
 934 **Task 3-5: Patent Application Examination** Patent Application Examination involves conduct-  
 935 ing compliance reviews of patent documents to ensure adherence to statutory and administrative re-  
 936 quirements. This task requires models to verify the accuracy, completeness, and legal sufficiency of  
 937 patent applications, including claims, specifications, and drawings, against jurisdictional standards.  
 938 Key checks include clarity of technical disclosure, consistency between claims and descriptions,  
 939 proper support for embodiments, and alignment with formalities. The data for this task is sourced  
 940 from the USPTO Office Action Dataset (Lu et al., 2017).

#### 941 D.2.4 CREATIVE GENERATION

942  
 943 **Task 4-1: Abstract Generation** Abstract Generation assesses a model’s ability to automatically  
 944 extract core elements from intellectual property (IP) texts, such as patent claims, and synthesize  
 945 them into concise, structured, and legally compliant summaries. This task requires models to distill  
 946 technical innovations, legal scopes, and critical details while adhering to jurisdictional formatting  
 947 rules and avoiding oversimplification that misrepresents legal or technical nuances.

948  
 949 **Task 4-2: Dependent Claim Generation** Dependent Claim Generation involves automatically  
 950 drafting legally compliant and technically precise dependent claims based on the core inventions  
 951 described in patent disclosures. This task requires models to analyze technical descriptions and  
 952 generate claims that refine or limit the scope of independent claims by incorporating additional  
 953 technical features, while ensuring logical dependency and alignment with jurisdictional formalities.  
 954 This task is inspired by PatentEval (Zuo et al., 2024).

955  
 956 **Task 4-3: Design-Around Solution Generation** Design-Around Solution Generation focuses on  
 957 creating non-infringing technical alternatives by analyzing existing patent claims and identifying  
 958 opportunities to circumvent key protected elements. This task requires models to deconstruct patent  
 959 claims and propose modifications that avoid literal or equivalent infringement, while maintaining  
 960 technical feasibility.

961  
 962 Table 9: Data language distribution of IPBench.

Language	1-1	1-2	1-3	1-4	1-5-1	1-5-2	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	3-5	4-1	4-2	4-3	Sum
Chinese	259	276	294	252	525	0	338	308	250	228	156	139	301	160	159	500	217	0	200	200	328	5090
English	241	226	206	252	600	219	240	250	272	160	162	0	140	149	500	183	314	200	199	171	5284	
Total	500	502	500	504	1125	600	557	548	500	500	316	301	301	300	308	1000	400	314	400	399	499	10374

## 965 966 E DATA ANNOTATION AND EXAMINATION PROTOCOL

### 967 E.1 DATA COLLECTION

970 We list the primary websites from which we collected the raw data as follows:  
 971

- 972 • USTPO’s Open Data Portal: <https://data.uspto.gov/home>

972           • CNIPA’s Official Website: <https://www.cnipa.gov.cn/>  
 973           • China Judgements Online: <https://wenshu.court.gov.cn/>  
 974

975  
 976 **Ethical considerations.** The data we collected come from open and public sources, and we con-  
 977 firm that they are not used for any commercial purposes. We strictly comply with all copyright and  
 978 licensing regulations. Data originating from sources that do not allow copying or redistribution are  
 979 deliberately excluded.

980           E.2 ANNOTATION AND EXAMINATION GUIDELINES  
 981

982           We provide detailed data annotation guidelines to ensure the quality, correctness, and difficulty  
 983 of our benchmark. Notably, most of our human expert annotators, who come from backgrounds in  
 984 intellectual property and public management, range from senior undergraduates to Ph.D. candidates.  
 985 They are included as co-authors of this paper as a non-monetary form of acknowledgment for their  
 986 efforts. They possess deep knowledge of intellectual property.

987  
 988 **Preparation before annotation.** We divide our 21 human expert annotators into four groups and  
 989 assign them to different tasks, including data annotation and annotation review. Each group is re-  
 990 quired to thoroughly understand their assigned task and formulate a comprehensive annotation plan  
 991 accordingly. This involves understanding the task definition, relevant legal concepts, and technical  
 992 terminologies related to intellectual property.

993  
 994 **General principles and process of annotation.** Firstly, all raw data or information must be col-  
 995 lected from official websites that are publicly accessible. For websites that prohibit copying, anno-  
 996 tators are instructed not to use them. Secondly, all annotators are required to ensure the accuracy of  
 997 their annotated questions and to ensure that the difficulty level is appropriate. For data containing  
 998 mathematical equations or special notations, we ask annotators to convert them into LaTeX format.  
 999 For other typographical errors, human expert annotators will correct them manually. Thirdly, all data  
 1000 will be examined by switching roles between annotation teams to verify and ensure their quality. For  
 1001 each datapoint, after the quality check, human expert annotators are required to label the language,  
 1002 the type of IP mechanism, and the data source.

1003  
 1004 **Specific principles of examination.** To ensure data quality, we assign a different annotation team  
 1005 to double-check and cross-validate the results. In cases where errors, inconsistencies, or misun-  
 1006 derstandings are identified, human examiners must provide detailed explanations and determine  
 1007 whether the data can be corrected and preserved. After the annotator corrects the question, the ex-  
 1008 aminer will re-evaluate the data until it passes the review with mutual agreement. This strict process  
 1009 ensures the reliability of our data, with each datapoint undergoing an average of three rounds of  
 1010 review to form IPBench.

1011  
 1012           F MORE DETAILS ABOUT DATA STATICS  
 1013

1014           In this section, we provide additional details about the data. Further statistical information can be  
 1015 found in Section F.

1016           Our IPBench comprises 10,374 datapoints spanning 20 tasks, including multiple-choice questions,  
 1017 classification tasks, and generation tasks. In this section, we provide additional data statistics, cover-  
 1018 ing language distribution, IP mechanism distribution, IPC/CPC classification distribution, text length  
 1019 distribution, and the distribution of option counts in multiple-choice questions.

1020  
 1021           F.1 DATA LANGUAGE DISTRIBUTION  
 1022

1023           Our IPBench is constrained to the legal frameworks of the United States and mainland China; there-  
 1024 fore, the dataset includes both English and Chinese languages. We present the language distribution  
 1025 for each task, as well as for the entire dataset, in Table 9.

1026 Table 10: Intellectual property mechanisms distribution of IPBench. TD: Trademark, SC: Software  
 1027 Copyright, TS: Trade Secret, PV: Plant Variety, CR: Copyright, IC: Integrated Circuit, GM:  
 1028 Geographical Mark.

Task	Patent	TD	SC	TS	PV	CR	IC	GM	Total
1-1	225	157	13	25	13	34	24	9	500
1-2	221	95	21	0	0	141	6	18	502
1-3	237	116	1	1	1	143	0	1	500
1-4	325	37	12	33	29	58	8	2	504
1-5-1	525	600	0	0	0	0	0	0	1125
1-5-2	600	0	0	0	0	0	0	0	600
1-6	159	103	22	107	1	157	1	7	557
1-7	190	358	0	0	0	0	0	0	548
2-1	320	21	9	77	39	24	10	0	500
2-2	183	105	16	49	3	144	0	0	500
2-3	101	94	11	10	0	100	0	0	316
2-4	301	0	0	0	0	0	0	0	301
2-5	0	0	0	301	0	0	0	0	301
3-1	300	0	0	0	0	0	0	0	300
3-2	308	0	0	0	0	0	0	0	308
3-3	1000	0	0	0	0	0	0	0	1000
3-4	353	0	8	13	18	5	3	0	400
3-5	314	0	0	0	0	0	0	0	314
4-1	400	0	0	0	0	0	0	0	400
4-2	399	0	0	0	0	0	0	0	399
4-3	497	1	0	1	0	0	0	0	499
<b>Total</b>	<b>6958</b>	<b>1687</b>	<b>113</b>	<b>617</b>	<b>104</b>	<b>806</b>	<b>52</b>	<b>37</b>	<b>10374</b>

Table 11: Distribution of intellectual property mechanisms in the English portion of IPBench.

Task	Patent	TD	SC	TS	PV	CR	IC	GM	Total
1-1	150	57	0	1	10	3	20	0	241
1-2	92	52	0	0	0	82	0	0	226
1-3	53	64	0	1	1	86	0	1	206
1-4	202	13	4	4	0	29	0	0	252
1-5-1	0	600	0	0	0	0	0	0	600
1-5-2	600	0	0	0	0	0	0	0	600
1-6	65	54	8	26	0	66	0	0	219
1-7	58	182	0	0	0	0	0	0	240
2-1	170	21	4	22	9	24	0	0	250
2-2	101	58	4	29	0	80	0	0	272
2-3	52	45	6	2	0	55	0	0	160
2-4	162	0	0	0	0	0	0	0	162
2-5	0	0	0	0	0	0	0	0	0
3-1	140	0	0	0	0	0	0	0	140
3-2	149	0	0	0	0	0	0	0	149
3-3	500	0	0	0	0	0	0	0	500
3-4	175	0	0	4	0	4	0	0	183
3-5	314	0	0	0	0	0	0	0	314
4-1	200	0	0	0	0	0	0	0	200
4-2	199	0	0	0	0	0	0	0	199
4-3	169	1	0	1	0	0	0	0	171
<b>Total</b>	<b>3551</b>	<b>1147</b>	<b>26</b>	<b>90</b>	<b>20</b>	<b>429</b>	<b>20</b>	<b>1</b>	<b>5284</b>

## F.2 INTELLECTUAL PROPERTY MECHANISMS DISTRIBUTION

Our IPBench covers eight intellectual property mechanisms, including Patent, Trademark, Software Copyright, Trade Secret, New Plant Variety, Copyright, Integrated Circuit Layout Design, and Geographical Indication. We present a detailed distribution of these intellectual property mechanisms in our benchmark, as shown in Table 10, Table 11 (English section), and Table 12 (Chinese section).

Table 12: Distribution of Intellectual Property Mechanisms in the Chinese Portion of IPBench.

Task	Patent	TD	SC	TS	PV	CR	IC	GM	Total
1-1	75	100	13	24	3	31	4	9	259
1-2	129	43	21	0	0	59	6	18	276
1-3	184	52	1	0	0	57	0	0	294
1-4	123	24	8	29	29	29	8	2	252
1-5-1	525	0	0	0	0	0	0	0	525
1-5-2	0	0	0	0	0	0	0	0	0
1-6	94	49	14	81	1	91	1	7	338
1-7	132	176	0	0	0	0	0	0	308
2-1	150	0	5	55	30	0	10	0	250
2-2	82	47	12	20	3	64	0	0	228
2-3	49	49	5	8	0	45	0	0	156
2-4	139	0	0	0	0	0	0	0	139
2-5	0	0	0	301	0	0	0	0	301
3-1	160	0	0	0	0	0	0	0	160
3-2	159	0	0	0	0	0	0	0	159
3-3	500	0	0	0	0	0	0	0	500
3-4	178	0	8	9	18	1	3	0	217
3-5	0	0	0	0	0	0	0	0	0
4-1	200	0	0	0	0	0	0	0	200
4-2	200	0	0	0	0	0	0	0	200
4-3	328	0	0	0	0	0	0	0	328
<b>Total</b>	<b>3407</b>	<b>540</b>	<b>87</b>	<b>527</b>	<b>84</b>	<b>377</b>	<b>32</b>	<b>36</b>	<b>5090</b>

1080 F.3 IPC AND CPC CLASSIFICATION DISTRIBUTION  
10811082 We present the IPC Section classification distribution in Table 13a and the CPC Section classification  
1083 distribution in Table 13b.1084  
1085 Table 13: Distribution of IPC and CPC sections.

Section	Count	Percentage (%)	Section	Count	Percentage (%)
A	72	6.5	A	90	15.0
B	249	22.1	B	90	15.0
C	29	2.6	C	48	8.0
D	6	0.5	D	17	2.8
E	63	5.6	E	71	11.9
F	64	5.7	F	189	31.5
G	113	10.0	G	44	7.3
H	529	47.0	H	51	8.5
<b>All</b>	<b>1125</b>	<b>100</b>	<b>All</b>	<b>600</b>	<b>100</b>

(a) Distribution of IPC sections

(b) Distribution of CPC sections

1100 Table 14: Text length statistics (in tokens) for each task across three dimensions: average, minimum,  
1101 and maximum length; each further split by language (EN/CH). Missing values are denoted by "-".

Task	Avg-All	Avg-EN	Avg-CH	Min-All	Min-EN	Min-CH	Max-All	Max-EN	Max-CH
1-1	83.9	68.9	97.8	46	46	66	258	112	258
1-2	81.2	71.3	89.3	47	47	66	135	105	135
1-3	102.2	80.3	117.6	55	55	76	208	129	208
1-4	116.2	112.8	119.6	61	61	73	195	151	195
1-5-1	216.7	163.6	277.4	49	49	110	305	305	455
1-5-2	165.1	165.1	-	50	50	-	337	337	-
1-6	89.4	77.0	97.5	55	55	70	146	146	128
1-7	41.7	-	74.2	40	40	53	107	101	107
2-1	161.6	103.1	220.1	66	66	177	310	140	310
2-2	109.4	97.4	123.6	59	59	71	211	189	211
2-3	122.8	107.9	138.1	70	70	80	263	171	263
2-4	99.8	88.1	113.5	66	66	87	144	125	144
2-5	112.1	-	112.1	51	-	51	302	-	302
3-1	158.2	145.8	169.1	-	-	-	-	-	-
3-2	91.4	76.0	105.9	53	53	78	150	121	150
3-3	1239.5	1231.8	1247.3	575	581	575	1956	1845	1956
3-4	166.1	169.8	163.0	60	92	60	297	297	327
3-5	7460.4	7460.4	60.5	1428	1428	-	10219	10219	-
4-1	1636.7	2199.3	1074.0	-	-	285	-	8064	5675
4-2	448.5	534.1	363.3	68	89	68	1861	1485	1861
4-3	121.2	111.8	126.1	56	56	84	218	183	218

1121  
1122 Table 15: Aggregated text length statistics (in tokens) by task type. PE refers to Patent Examination  
1123 (Task 3-5), MCQA refers to Multiple-choice Question Answering.

Type	Avg-All	Avg-EN	Avg-CH	Min-All	Min-EN	Min-CH	Max-All	Max-EN	Max-CH
MCQA	181.0	181.6	194.7	90.7	96.2	111.1	326.7	272.5	327.5
PE	7460.4	7460.4	60.5	1428.0	1428.0	-	10219.0	10219.0	-
Classification	190.9	164.4	277.4	49.5	49.5	110.0	321.0	321.0	455.0
Generation	1042.6	1366.7	718.7	68.0	89.0	176.5	1861.0	4774.5	3768.0

1130 F.4 TEXT LENGTH DISTRIBUTION  
11311132 We provide detailed statistics on the text length distribution for each task, across the three question  
1133 types, in both Chinese and English. In all text length computations presented in this paper, we

1134  
1135 Table 16: Distribution of answer choices by  
1136 task.  
1137

Task	A	B	C	D	Total
1-1	125	129	126	120	500
1-2	117	162	117	106	502
1-3	125	126	126	123	500
1-4	126	126	127	125	504
1-6	151	137	142	127	557
1-7	132	194	124	98	548
2-1	155	130	124	91	500
2-2	101	177	154	68	500
2-3	74	108	85	49	316
2-4	74	83	75	69	301
2-5	57	165	63	16	301
3-1	114	82	59	45	300
3-2	72	76	76	84	308
3-3	240	256	230	278	1000
3-4	76	141	128	55	400
4-3	170	144	111	74	499
<b>Total</b>	<b>1909</b>	<b>2236</b>	<b>1867</b>	<b>1528</b>	<b>7536</b>

1134  
1135 Table 17: Distribution of English questions' an-  
1136 swers by task.  
1137

Task	A	B	C	D	Total
1-1	64	53	65	59	241
1-2	65	69	48	44	226
1-3	44	47	58	57	206
1-4	70	71	57	54	252
1-6	39	70	40	70	219
1-7	41	70	64	65	240
2-1	92	65	64	29	250
2-2	70	68	77	57	272
2-3	38	68	26	28	160
2-4	39	45	40	38	162
3-1	54	38	33	15	140
3-2	38	35	39	37	149
3-3	120	126	115	139	500
3-4	44	64	50	25	183
4-3	44	65	42	20	171
<b>Total</b>	<b>862</b>	<b>954</b>	<b>818</b>	<b>737</b>	<b>3371</b>

1134  
1135 adopt the tokenizer of GPT-4o for consistency and comparability. Table 14 and Table 15 present the  
1136 distribution of text lengths from different perspectives: the former provides statistics by task, while  
1137 the latter summarizes the data by question type.

## 1138 F.5 MULTI-CHOICE QUESTION OPTION COUNT DISTRIBUTION.

1139 In this section, we present the distribution of multiple-choice question option counts, as shown in  
1140 Table 16, Table 17, and Table 18, along with the examination option distribution for Task 3-5, as  
1141 shown in Table 19. For multiple-choice questions, each question has four options: A, B, C, and D.  
1142 In contrast, for Task 3-5, each question has two options: allowed and rejected.

1143  
1144 Table 18: Distribution of Chinese questions' answers by task.  
1145

Task	A	B	C	D	Total
1-1	61	76	61	61	259
1-2	52	93	69	62	276
1-3	81	79	68	66	294
1-4	56	55	70	71	252
1-6	112	67	102	57	338
1-7	91	124	60	33	308
2-1	63	65	60	62	250
2-2	31	109	77	11	228
2-3	36	40	59	21	156
2-4	35	38	35	31	139
2-5	57	165	63	16	301
3-1	60	44	26	30	160
3-2	34	41	37	47	159
3-3	120	126	115	139	500
3-4	32	77	78	30	217
4-3	126	79	69	54	328
<b>Total</b>	<b>1047</b>	<b>1278</b>	<b>1049</b>	<b>791</b>	<b>4165</b>

1188

Table 19: Examination outcome distribution for Task 3-5.

1189

1190

1191

1192

1193

1194

1195

1196

1197

1198

1199

1200

## G PROMPTS

1201

1202

1203

1204

### G.1 ZERO-SHOT AND FEW-SHOT PROMPT

1205

1206

1207

We adapt four types of zero-shot prompts and few-shot prompts for our experiment, corresponding to different task types: choice questions, classification, examination, and generation, across both English and Chinese languages. The Chinese version uses the same content as the English version.

1211

#### Zero-shot Prompt for Choice Question Task

1212

1213

1214

Please answer the following question thoughtfully and provide your final answer at the end in the format 'Answer: \*\*option\*\*',

1215

1216

{ Question }

1217

1218

#### Zero-shot Prompt for IPC/CPC Classification Task (1-5)

1219

1220

1221

Please answer the following question thoughtfully and provide your final answer at the end in the format 'Answer: \*\*corresponding IPC number\*\*',

1222

1223

{Question}

1224

1225

#### Zero-shot Prompt for Generation Task (4-1, 4-2)

1226

##### Abstract Generation based on Claims (4-1):

1227

# Claims

1228

{Claims Text}

1229

Please generate the abstract of the patent based on the given claims.

1230

1231

##### Dependent Claim Generation (1-5-2):

1232

# Independent Claim

1233

{Claim Text}

1234

Please generate all dependent claims corresponding to the given independent claim.

1235

1236

#### Zero-shot Prompt for Patent Application Examination Task (3-5)

1237

1238

1239

Please examine the patents in # Patent Applications Awaiting Examination. Determine whether each patent application should be allowed or rejected.

Return your decision in the following format:

1240

1241

Answer: allowed / rejected

1242                    Few-shot Prompt for Choice Question Task  
 1243  
 1244        # There are k examples  
 1245        ## Example {1}  
 1246        Question: {1-shot-question}  
 1247        Answer:{1-shot-answer}  
 1248        ...  
 1249        ## Example {k}  
 1250        Question: {k-shot-question}  
 1251        Answer:{k-shot-answer}  
 1252        Please answer the following question thoughtfully and provide your final answer at the end in  
 1253        the format 'Answer: \*\*option\*\*',  
 1254        { Question }  
 1255

1256                    Few-shot Prompt for IPC/CPC Classification Task  
 1257  
 1258        # There are k examples  
 1259        ## Example {1}  
 1260        Question: {1-shot-question}  
 1261        Answer:{1-shot-answer}  
 1262        ...  
 1263        ## Example {k}  
 1264        Question: {k-shot-question}  
 1265        Answer:{k-shot-answer}  
 1266        Please answer the following question thoughtfully and provide your final answer at the end in  
 1267        the format 'Answer: \*\*corresponding IPC/CPC number\*\*',  
 1268        {Question}  
 1269

## G.2 CHAIN-OF-THOUGHT PROMPT

1270  
 1271  
 1272                    Chain-of-Thought Prompt for Choice Question Task  
 1273  
 1274  
 1275        Please answer the following question thoughtfully and provide your final answer at the end in the  
 1276        format 'Answer: \*\*option\*\*',  
 1277  
 1278        { Question }  
 1279  
 1280        **Let's think step by step.**  
 1281

1282                    Chain-of-Thought Prompt for IPC/CPC Classification Task  
 1283  
 1284        Please answer the following question thoughtfully and provide your final answer at the end in the  
 1285        format 'Answer: \*\*corresponding IPC/CPC number\*\*',  
 1286  
 1287        {Question}  
 1288        **Let's think step by step.**  
 1289

## H METRICS

1290  
 1291  
 1292        In this section, we provide the details of the metrics used in our IPBench. The details of the multiple-  
 1293        choice question metric are in Section [H.1](#), the details of the classification task metric are in Section [H.2](#), and the details of the generation task are in Section [H.3](#).  
 1294

1296 Table 20: The overview of evaluated models. Max Context refers to the maximum context length of  
 1297 the model without length extrapolation for all models.

Model	Size	Max Context	Type	Orientation	Access
GPT-4o (Hurst et al., 2024)	–	128k	Chat Model	General	OpenAI API
GPT-4o-mini (Hurst et al., 2024)	–	128k	Chat Model	General	OpenAI API
DeepSeek-V3 (DeepSeek-AI et al., 2024)	671B	128k	Chat Model	General	DeepSeek API
Qwen3 (Team, 2025a)	8B	32k	Chat Model	General	Weights
Qwen2.5-Instruct (Yang et al., 2024)	7/72B	32k	Chat Model	General	Weights
Llama3.1-Instruct (Dubey et al., 2024)	8/70B	32k	Chat Model	General	Weights
Gemma-2-Instruct (Riviere et al., 2024)	9/27B	8k	Chat Model	General	Weights
Mistral-7B-Instruct (Jiang, 2024)	7B	32k	Chat Model	General	Weights
MoZi-qwen (Ni et al., 2024)	7B	32k	Chat Model	IP	Weights
DISC-LawLLM (Yue et al., 2023; 2024)	6B	2048	Chat Model	Law	Weights
HanFei (He et al., 2023)	7B	2048	Chat Model	Law	Weights
DeepSeek-R1 (DeepSeek-AI et al., 2025)	671B	128k	Reasoning Model	General	DeepSeek API
Deepseek-R1-Distill-Qwen (DeepSeek-AI et al., 2025)	7B	32k	Reasoning Model	General	Weights
QwQ (Team, 2025b)	32B	32k	Reasoning Model	General	Weights

## H.1 MULTI-CHOICE QUESTION METRIC

For multiple-choice questions, we use accuracy as the metric due to the straightforward nature of the judgment process. Each multiple-choice question has four options: A, B, C, and D. We use the same extraction method for each model’s response, compare the selected answer with the ground-truth option, and then compute the average accuracy. The average score ranges from 0 to 100, and is computed as shown in the Equation 1.

$$\text{Accuracy} = \frac{\text{Number of Correct Answers \#}}{\text{Total Number of Questions \#}} \quad (1)$$

## H.2 IPC/CPC CLASSIFICATION TASK METRIC

For IPC/CPC classification task, we use exact-match as the metric. For example, in the IPC code A01B00/66, ‘A’ represents the Section, ‘01’ the Class, and ‘B’ the Subclass. If the model predicts ‘A’, it earns one point for the Section; if it predicts ‘A01’, it earns one point for the Class; and if it predicts ‘A01B’, it earns one point for the Subclass. If the entire code is predicted correctly, one point is awarded for the Exact Match. We evaluate all the test data to calculate the average exact-match score across these four levels. The difficulty increases as the model is required to make correct predictions at more levels.

## H.3 GENERATION TASK METRIC

In this section, we provide the details of the LLM-as-a-judge approach used for LLMScore and analyze its consistency with human evaluation.

We design five evaluation dimensions for LLM-as-a-judge: Accuracy, Relevance, Completeness, Consistency, and Language-Style. The detailed definitions are provided in the prompts below. Each dimension is scored on a scale from 1 to 10 points. We use DeepSeek-V3 as the judge model because it achieves relatively better performance on the multiple-choice tasks, indicating solid knowledge in the intellectual property domain. In addition to the LLM-as-a-judge evaluation, we further sample 50 responses each from GPT-4o, DeepSeek-V3 and LLaMA3.1-8B-Instruct for the two tasks. These responses are assessed by three human experts using the same criteria as the LLM-as-a-judge framework. The results and the corresponding consistency between the LLM and human analysis are presented in Table 6.

We provide a consistency analysis between different metrics and human evaluations, including Kendall, Pearson, and Spearman coefficients. The higher the consistency coefficient, the better, indicating stronger consistency; the smaller the p-value, the better, indicating statistical significance. A smaller p-value, typically less than 0.05, indicates that the observed correlation is statistically significant.

1350 **LLMScore for Generation Task.** For Task 4-1 and 4-2, we draw inspiration from the error tax-  
 1351 onomy for abstract generation and dependent claims generation proposed in PatentEval (Zuo et al.,  
 1352 2024), and used five dimensions to evaluate the quality of the generated abstract. The specific prompt  
 1353 we use for LLM-as-a-judge in evaluating generation task are provided in code.  
 1354

## 1355 I DETAILS ABOUT EVALUATED MODELS

1356 We provide details of the evaluated models, including their size, context length, type, and access  
 1357 method, as shown in Table 20.

## 1361 J MORE DISCUSSION

1363 The growing integration of LLMs into high-stakes domains demands rigorous, domain-specific eval-  
 1364 uation frameworks. Among these domains, IP presents unique challenges that remain largely unad-  
 1365 dressed in existing NLP benchmarks. IP tasks operate at the intersection of technical innovation and  
 1366 legal regulation, requiring precise reasoning over structured taxonomies (e.g., IPC/CPC classifica-  
 1367 tions), formal legal constructs (e.g., claim scope and infringement logic), and high-stakes decisions  
 1368 (e.g., patentability, damages, licensing). Yet most LLM benchmarks either omit this domain or  
 1369 reduce it to surface-level tasks like summarization or basic classification.

1370 This oversight poses real risks. As LLMs begin to influence decision-making pipelines in patent  
 1371 examination, IP analytics, or IP litigation support, the lack of tailored evaluation may lead to mis-  
 1372 leading conclusions about model capabilities. Moreover, the complexity of IP, spanning multiple  
 1373 jurisdictions, languages, legal doctrines, and technical fields, makes it an ideal stress test for mea-  
 1374 suring LLMs’ reasoning, memory, and generation under constraint.

1376 Our work addressed this critical gap by introducing IPBench, a bilingual, multi-dimensional bench-  
 1377 mark grounded in real-world IP tasks. The benchmark is built on a four-level task taxonomy adapted  
 1378 from Webb’s DoK theory, ranging from low-level recall to high-level creative synthesis. These lev-  
 1379 els are aligned not just with educational psychology but with actual workflows in patent offices, IP  
 1380 law firms, and technology transfer environments. Unlike prior benchmarks such as PatentEval (Zuo  
 1381 et al., 2024), which focus narrowly on a few patent tasks, IPBench spans 20 tasks across 8 IP mech-  
 1382 anisms and includes both comprehension-based and generative formats.

1383 Our empirical findings revealed several important trends and limitations in current LLMs. First,  
 1384 general-purpose models such as GPT-4o and DeepSeek-V3 consistently outperform law- and IP-  
 1385 specific models. This may seem counterintuitive, as vertical models like MoZi-qwen are explicitly  
 1386 trained on legal corpora. However, this underperformance likely results from a combination of  
 1387 overfitting, insufficient general reasoning capabilities, and inadequate coverage of the procedural  
 1388 and generative aspects of IP workflows. Vertical fine-tuning strategies may inadvertently narrow  
 1389 the model’s inferential space or induce catastrophic forgetting, degrading performance on multi-step  
 1390 reasoning tasks.

1391 Second, our evaluation of reasoning-oriented models such as DeepSeek-R1 and QwQ-32B revealed  
 1392 that while they do not top the overall leaderboard, they outperform chat-based models on specific  
 1393 tasks requiring arithmetic logic, legal thresholds, or rule-based evaluation (e.g., compensation es-  
 1394 timation or damages calculation). This supports the hypothesis that architecture matters: models  
 1395 with symbolic reasoning capabilities have a structural advantage in tasks where correctness hinges  
 1396 on numerical precision or multi-condition rule satisfaction.

1397 Third, our analysis of prompting techniques showed mixed outcomes. Few-shot prompting improves  
 1398 performance on some models and tasks, particularly in instruction-following or retrieval-based sce-  
 1399 narios. However, models like Llama3.1-8B-it show no consistent improvement, suggesting sensitiv-  
 1400 ity to prompt design or training data mismatches. The CoT prompting, often touted as a reasoning  
 1401 enhancer, surprisingly leads to performance drops (0.40.6%) across models. Our error analysis  
 1402 attributes this to the injection of spurious reasoning paths and overthinkinga phenomenon also ob-  
 1403 served in prior work (Zheng et al., 2025; Fan et al., 2025). In domains like IP, where many tasks  
 1404 hinge on memorized definitions or hierarchical rule structures, CoT may actually degrade perfor-  
 1405 mance by introducing incorrect logic.

1404  
 1405 Fourth, the performance on IPC/CPC classification is alarmingly low. Even the best model,  
 1406 DeepSeek-R1, achieves only a 10.8% Exact Match rate. These classification systems are essential  
 1407 for patent analytics, prior art search, and innovation tracking, and failure to resolve them accurately  
 1408 reflects fundamental limitations in LLMs’ ability to represent domain hierarchies, align semantic  
 1409 cues with technical structure, and disambiguate overlapping categories. These failures underscore  
 1410 a broader issue in LLM design: current architectures are not optimized for structured symbolic  
 1411 taxonomies or discrete label hierarchies that are common in regulatory domains.

1411 We also introduced LLMScore, an automatic evaluation metric tailored to generative tasks in IP. Un-  
 1412 like traditional metrics such as BLEU and ROUGE, which are inadequate for legal text due to their  
 1413 lack of semantic granularity, LLMScore is based on the LLM-as-a-judge paradigm and evaluates  
 1414 responses across four human-aligned dimensions. Empirical results demonstrate that LLMScore  
 1415 correlates more strongly with human judgments and supports nuanced evaluation of claim and ab-  
 1416 stract generation, which are central to both patent drafting and retrieval.

1417 Collectively, these results highlight not only current limitations in model generalization and prompt-  
 1418 ing strategies but also the inherent complexity of the IP domain. This complexity arises from its  
 1419 hybrid nature: legal and technical, deterministic and interpretive, global and jurisdiction-specific.  
 1420 Benchmarks like IPBench are thus essential not only for evaluation but for guiding the next phase  
 1421 of model development.

1422 Looking forward, we envision several extensions to IPBench and its applications. The current ver-  
 1423 sion focuses on U.S. and Chinese legal frameworks; future iterations will incorporate additional  
 1424 jurisdictions such as the EU and Japan, enabling cross-legal evaluation and comparative reasoning.  
 1425 Moreover, as more IP-specific models become available, IPBench can serve as a testbed for  
 1426 fine-tuning strategies, prompt engineering, and hybrid symbolicneural architectures. More broadly,  
 1427 IPBench offers a blueprint for evaluating LLMs in other complex verticalssuch as medicine, fi-  
 1428 nance, or regulatory compliance, where task diversity, interpretability, and factual correctness are  
 1429 non-negotiable. By operationalizing cognitive depth and legal realism in benchmark design, we  
 1430 hope to catalyze the development of trustworthy, capable, and domain-aligned LLMs.

## 1431 1432 K MORE RESULTS

1435 In Section Section K, we present additional results under various experimental settings, covering  
 1436 both Chinese and English. Specifically, Section K.1 reports the overall results on IPBench, Sec-  
 1437 tion K.2 presents the results for Chinese questions, Section K.3 covers the results for English ques-  
 1438 tions, and Section K.4 provides detailed results of the LLM-as-a-judge evaluation along with its  
 1439 consistency with human judgments.

### 1441 1442 K.1 OVERALL RESULTS

1443 We provide the results of overall performance under the few-shot setting (1-shot, 2-shot, and 3-shot)  
 1444 in Section K.1.1, and the results under the chain-of-thought setting in Section K.1.2. We provide a  
 1445 model performance heatmap as shown in Figure 5, where models are sorted by their overall perfor-  
 1446 mance. A redder color indicates that the model on the x-axis outperforms the corresponding model  
 1447 on the y-axis.

#### 1450 1451 K.1.1 FEW-SHOT RESULTS

1452 The 1-shot results of IPBench are presented in Table 21 and Table 22, the 2-shot results in Table 23  
 1453 and Table 24, and the 3-shot results in Table 25 and Table 26.

#### 1454 1455 K.1.2 CHAIN-OF-THOUGHT RESULTS

1456 The chain-of-thought results of IPBench are presented in Table 27 and Table 28.

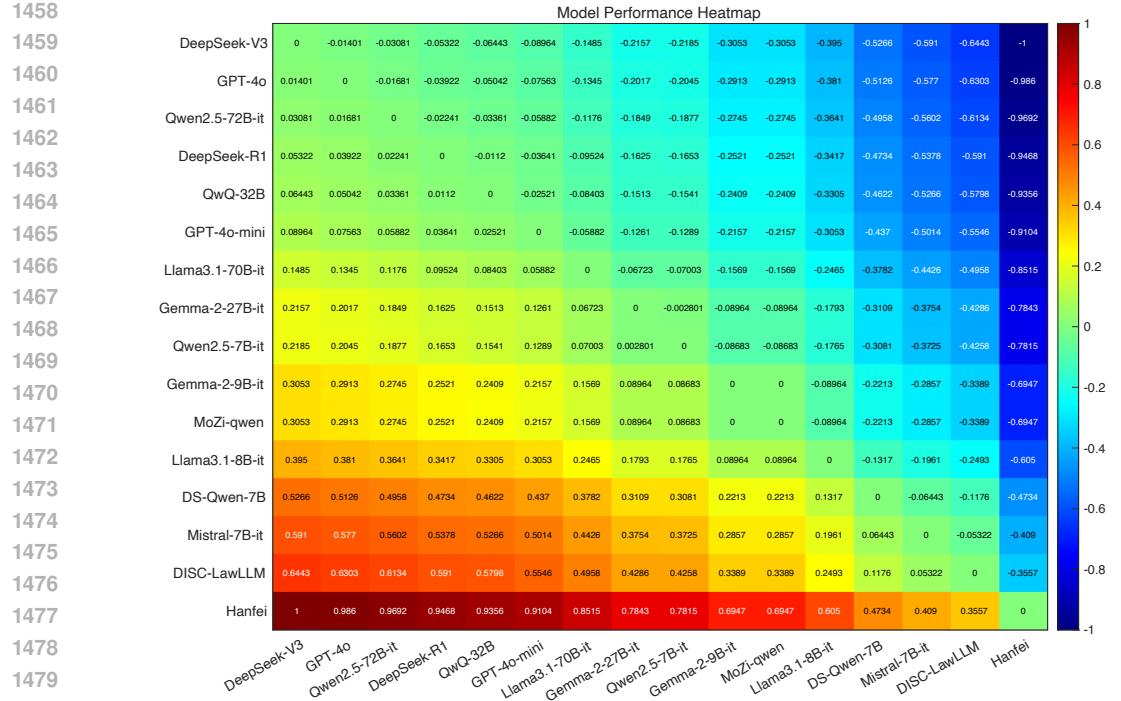


Figure 5: Model performance heatmap.

Table 21: Results of IPBench with 1-shot setting. The best-performing model in each task is in **darker red**, and the second best is in **lighter red**.

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>73.9</b>	93.8	<b>86.3</b>	<b>78.6</b>	<b>79.8</b>	<b>61.1</b>	<b>66.6</b>	<b>51.6</b>	62.6	<b>61.4</b>	<b>77.1</b>	81.4	69.3	<b>76.9</b>	<b>81.1</b>	<b>79.8</b>	<b>70.3</b>
Qwen2.5-7B-it	67.5	94.2	82.5	76.4	72.8	60.8	63.5	48.8	62.0	52.2	70.4	76.1	69.0	69.5	55.6	79.5	62.1
Llama3.1-8B-it	59.3	87.0	69.7	67.4	72.0	50.7	64.6	46.0	57.6	43.7	69.4	45.9	60.0	69.8	36.8	73.2	57.1
Gemma-2-9B-it	66.4	89.2	74.3	71.0	73.2	55.6	60.8	50.8	<b>65.0</b>	50.0	72.8	<b>82.7</b>	<b>70.0</b>	70.8	53.9	75.7	69.5
Mistral-7B-it	54.8	79.0	61.6	63.4	59.9	43.4	52.9	44.6	57.4	36.4	61.8	62.1	60.3	48.4	39.4	67.2	57.1
MoZi-qwen	63.0	<b>95.2</b>	81.5	76.2	66.9	59.9	64.4	47.2	<b>62.6</b>	40.2	72.1	77.4	69.0	58.8	36.6	71.8	57.9
DISC-LawLLM	47.7	78.8	66.9	66.4	65.5	45.0	52.2	40.0	54.6	32.0	51.8	70.8	58.7	39.6	–	67.3	30.7
Hanfei	28.0	37.0	29.3	28.8	31.2	33.9	43.6	26.6	37.2	35.8	29.9	24.6	37.0	22.4	–	31.0	30.1

Table 22: Results of Patent IPC/CPC Classification tasks (1-5-1 and 1-5-2) with 1-shot setting. The best-performing model in each task is in **darker purple**, and the second best is in **lighter purple**.

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o-mini	2.2	81.8	67.0	50.8	0.5	74.3	59.1	49.1
DeepSeek-V3	<b>15.1</b>	<b>86.3</b>	<b>75.5</b>	<b>60.5</b>	<b>7.3</b>	<b>86.2</b>	<b>74.3</b>	<b>65.0</b>
Qwen2.5-7B-it	2.2	73.8	57.9	42.2	0.3	67.5	48.8	37.2
Llama3.1-8B-it	0.7	64.1	49.7	33.7	0.0	45.2	35.2	22.2
MoZi-qwen	0.4	47.0	34.6	21.9	0.0	16.5	7.8	4.3

## K.2 CHINESE QUESTIONS RESULTS

In Section Section K.2, we focus on the IPBench results for Chinese questions. We provide the zero-shot results for the Chinese portion of IPBench in Section K.2.1, the few-shot results in Section K.2.2, and the chain-of-thought results in Section K.2.3.

1512 Table 23: Results of IPBench with 2-shot setting. The best-performing model in each task is in  
 1513 **darker red**, and the second best is in **lighter red**.  
 1514

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>74.0</b>	94.2	<b>87.9</b>	77.0	<b>80.2</b>	60.6	<b>66.6</b>	<b>52.0</b>	59.6	<b>63.3</b>	<b>79.4</b>	83.4	68.0	<b>78.6</b>	<b>80.7</b>	76.7	<b>72.5</b>
Qwen2.5-7B-it	69.3	94.8	83.3	76.0	77.0	59.9	65.0	47.8	60.4	56.0	71.8	83.7	67.7	69.5	65.9	<b>81.5</b>	57.7
Llama3.1-8B-it	59.3	85.8	67.3	68.4	69.6	52.3	61.7	48.0	56.6	45.6	71.4	66.8	60.0	61.4	37.8	69.8	53.9
Gemma-2-9B-it	67.3	89.0	76.3	71.4	72.2	56.3	61.1	51.4	<b>62.8</b>	52.2	74.7	<b>84.7</b>	66.7	72.4	57.4	78.0	70.3
Mistral-7B-it	57.2	79.8	64.1	64.6	61.5	44.4	53.8	47.0	56.6	38.0	65.1	73.1	61.8	48.7	47.1	67.5	56.5
MoZi-qwen	66.8	<b>96.0</b>	84.1	<b>77.2</b>	75.0	<b>63.6</b>	64.2	48.4	61.2	43.0	75.1	82.4	<b>72.0</b>	64.6	49.4	80.8	55.1
DISC-LawLLM	56.7	79.2	67.1	66.6	65.3	50.2	51.3	39.6	53.8	31.3	58.1	78.1	57.7	42.2	–	71.5	37.1
Hanfei	32.5	36.0	24.7	35.8	33.7	35.3	33.0	25.4	37.0	32.6	26.6	49.2	42.3	22.1	–	32.2	25.7

1524  
 1525 Table 24: Results of Patent IPC/CPC Classification tasks (1-5-1 and 1-5-2) with 2-shot setting. The  
 1526 best-performing model in each task is in **darker purple**, and the second best is in **lighter purple**.  
 1527

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o-mini	2.3	82.2	68.0	51.5	0.2	76.3	61.6	51.6
DeepSeek-V3	<b>15.1</b>	<b>86.7</b>	<b>76.1</b>	<b>60.6</b>	<b>7.2</b>	<b>86.5</b>	<b>73.3</b>	<b>65.7</b>
Qwen2.5-7B-it	2.5	78.2	62.4	46.2	0.3	68.5	51.2	38.8
Llama3.1-8B-it	1.1	59.7	44.7	29.2	0.0	63.3	45.7	26.7
MoZi-qwen	0.6	56.6	41.7	26.8	0.2	32.3	17.3	9.3

1536  
 1537 Table 25: Results of IPBench with 3-shot setting. The best-performing model in each task is in  
 1538 **darker red**, and the second best is in **lighter red**.  
 1539

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>74.7</b>	94.4	<b>87.5</b>	<b>79.6</b>	<b>80.0</b>	<b>63.3</b>	<b>68.8</b>	<b>52.4</b>	58.6	<b>63.6</b>	<b>80.1</b>	82.7	70.3	<b>77.9</b>	<b>80.0</b>	80.0	<b>75.0</b>
Qwen2.5-7B-it	70.6	94.2	83.3	73.6	76.0	62.2	68.2	50.8	<b>62.2</b>	51.3	74.8	84.7	70.7	72.4	68.5	<b>82.2</b>	60.3
Llama3.1-8B-it	59.4	87.4	67.0	66.4	69.8	52.7	63.7	45.6	55.4	43.0	66.4	75.4	61.0	62.7	36.7	70.5	56.3
Gemma-2-9B-it	67.4	89.4	76.1	70.6	70.6	56.6	62.5	51.2	<b>62.2</b>	51.6	76.4	<b>85.1</b>	68.3	69.5	58.7	76.3	70.7
Mistral-7B-it	56.5	80.6	63.9	62.8	61.7	45.9	54.4	47.4	56.0	36.1	64.1	73.8	63.3	50.3	40.1	67.8	58.1
MoZi-qwen	65.3	<b>96.2</b>	83.5	77.2	76.0	62.4	65.1	49.4	61.8	40.5	75.8	80.4	<b>72.7</b>	62.0	38.3	79.0	57.3
DISC-LawLLM	57.4	83.8	67.3	64.6	66.7	53.6	52.0	41.2	54.8	29.4	62.1	74.4	60.0	41.2	–	67.5	38.5
Hanfei	29.9	32.0	28.9	26.2	28.4	31.7	31.6	23.6	36.2	29.1	22.6	30.6	41.0	24.4	–	26.3	34.5

1548  
 1549 Table 26: Results of Patent IPC/CPC Classification tasks (1-5-1 and 1-5-2) with 3-shot setting. The  
 1550 best-performing model in each task is in **darker purple**, and the second best is in **lighter purple**.  
 1551

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o-mini	2.0	82.5	67.8	50.9	0.3	80.1	65.3	54.1
DeepSeek-V3	<b>15.6</b>	<b>87.1</b>	<b>76.2</b>	<b>61.2</b>	<b>7.7</b>	<b>85.7</b>	<b>73.3</b>	<b>64.7</b>
Qwen2.5-7B-it	2.3	78.8	62.7	46.8	0.5	68.3	50.8	38.8
Llama3.1-8B-it	1.2	65.6	48.9	32.8	0.0	64.8	45.8	29.8
MoZi-qwen	1.0	70.6	51.3	34.2	0.0	24.2	12.8	7.7

### K.2.1 ZERO-SHOT RESULTS

1562 The zero-shot results for the Chinese portion of IPBench are shown in Table 29, Table 30 and  
 1563 Table 31. Since the Patent CPC Classification task (1-5-2) only includes English questions, Table 30  
 1564 does not include it.

1566 Table 27: Results of IPBench with chain-of-thought setting. The best-performing model in each  
 1567 task is in **darker red**, and the second best is in **lighter red**.

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	3-5	4-3
GPT-4o-mini	<b>72.0</b>	<b>94.4</b>	<b>85.9</b>	<b>78.0</b>	<b>80.4</b>	<b>59.9</b>	<b>67.3</b>	<b>51.4</b>	62.6	<b>62.0</b>	<b>74.8</b>	<b>80.4</b>	65.7	<b>71.1</b>	<b>81.1</b>	<b>78.8</b>	<b>44.9</b>	<b>66.9</b>
Qwen2.5-7B-it	67.6	89.0	82.9	75.2	76.2	57.3	63.0	48.0	64.2	58.2	73.8	79.7	<b>66.7</b>	70.1	65.1	78.5	44.3	58.9
Llama3.1-8B-it	61.3	84.1	69.5	67.6	70.2	53.6	59.5	49.4	60.6	45.9	66.4	71.8	62.3	59.4	54.8	73.0	43.6	54.1
Gemma-2-9B-it	61.7	87.0	72.3	65.4	66.3	54.1	55.7	51.0	<b>64.8</b>	47.8	71.1	76.4	66.3	67.9	57.9	73.5	–	65.1
Mistral-7B-it	54.3	80.6	63.3	63.6	62.5	43.6	54.0	42.4	54.6	44.3	64.1	65.8	56.7	51.0	41.0	66.5	36.3	47.9
MoZi-qwen	60.2	<b>93.0</b>	79.9	72.0	65.3	50.2	61.9	45.2	52.4	45.2	66.8	72.4	58.3	62.7	49.2	71.0	43.9	44.1
DISC-LawLLM	37.3	65.4	57.4	48.6	39.3	42.8	41.4	25.4	34.8	25.9	32.2	62.8	26.7	25.7	17.5	39.0	–	30.5
Hanfei	29.9	42.0	28.5	32.4	34.1	30.6	26.8	28.8	29.6	21.5	25.9	24.3	34.3	26.6	24.6	31.7	–	35.3

1578  
 1579 Table 28: Results of Patent IPC/CPC Classification tasks (1-5-1 and 1-5-2) with chain-of-thought  
 1580 setting. The best-performing model in each task is in **darker purple**, and the second best is in  
 1581  
 1582 lighter purple .

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o-mini	0.2	80.4	67.4	51.6	0.0	76.8	63.0	52.5
DeepSeek-V3	1.3	<b>82.3</b>	<b>72.0</b>	<b>57.4</b>	<b>1.0</b>	<b>83.3</b>	<b>70.7</b>	<b>63.0</b>
Qwen2.5-7B-it	<b>1.8</b>	74.3	60.4	42.0	0.5	60.2	46.0	35.7
Llama3.1-8B-it	0.9	67.0	50.5	32.4	0.2	64.0	44.8	29.5
MoZi-qwen	0.3	22.4	17.2	12.6	0.0	7.7	2.8	1.8

1593 Table 29: Chinese questions results of IPBench. The best-performing model in each task is in  
 1594 **darker red**, and the second best is in **lighter red**. The model DS-Qwen refers to DeepSeek-R1-  
 1595 Distill-Qwen, while the suffix *it* indicates the Instruct version of the model. OA denotes the overall  
 1596 accuracy on the choice tasks.

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	4-3
GPT-4o	77.7	95.0	92.4	82.0	80.6	73.1	<b>73.7</b>	<b>50.8</b>	73.7	<b>67.3</b>	70.0	84.1	66.9	63.5	80.2	82.9	79.0
GPT-4o-mini	74.2	91.9	86.2	76.9	80.6	65.1	66.6	44.4	66.2	57.7	66.9	83.4	58.8	<b>80.5</b>	<b>82.2</b>	77.9	72.6
DeepSeek-V3	<b>78.7</b>	<b>97.3</b>	90.2	<b>87.1</b>	<b>83.0</b>	76.3	72.1	48.0	74.1	65.4	70.5	84.1	<b>67.5</b>	71.1	78.2	83.9	<b>84.8</b>
Qwen3	73.2	95.4	85.5	75.9	79.8	68.3	68.2	43.6	74.1	55.8	65.5	82.7	62.5	73.0	67.4	78.3	77.4
Qwen2.5-72B-it	77.9	96.5	92.4	82.0	81.8	69.8	70.5	48.0	<b>81.1</b>	62.2	<b>74.8</b>	82.1	<b>67.5</b>	68.6	81.8	79.3	82.6
Qwen2.5-7B-it	70.8	93.8	85.9	72.4	77.1	64.5	63.6	47.6	76.3	51.9	66.9	77.1	60.6	66.7	67.6	78.3	68.0
Llama3.1-70B-it	70.8	91.5	81.9	71.4	76.2	66.6	66.9	47.2	71.9	48.7	66.9	81.1	66.3	66.7	63.0	78.3	77.4
Llama3.1-8B-it	65.1	88.4	73.6	65.7	78.6	63.6	58.4	50.4	69.3	41.7	62.6	75.7	58.8	57.2	55.8	79.3	55.8
Gemma-2-27B-it	69.2	88.8	78.3	66.0	75.4	60.4	60.1	47.2	<b>77.2</b>	53.2	66.2	81.1	66.3	61.6	62.8	80.2	76.8
Gemma-2-9B-it	65.5	91.5	75.4	68.4	63.5	67.5	57.5	45.2	73.7	45.5	59.7	80.4	62.5	57.9	51.4	73.7	68.9
Mistral-7B-it	54.1	74.1	58.3	55.4	58.7	44.1	51.6	45.6	60.5	40.4	48.2	67.0	53.1	29.6	47.4	60.1	59.5
MoZi-qwen	67.9	93.1	86.2	73.1	59.1	69.2	64.6	48.4	64.0	41.0	56.8	76.4	60.0	57.9	64.0	82.0	65.5
DISC-LawLLM	55.0	86.9	69.2	64.6	63.5	64.5	49.7	38.0	66.7	40.4	46.8	64.8	50.6	37.7	28.4	72.8	41.5
Hanfei	39.9	65.3	46.7	50.3	53.2	45.9	51.3	28.4	43.4	26.9	36.7	49.2	41.9	20.1	10.0	45.2	35.1
DeepSeek-R1	76.6	95.8	91.3	84.4	79.4	74.0	73.1	44.8	76.8	66.0	67.6	<b>85.4</b>	59.4	78.6	64.8	<b>84.3</b>	82.9
DS-Qwen-7B	58.2	79.2	59.1	50.7	63.9	51.2	49.4	44.0	53.5	38.5	49.6	65.5	50.6	64.8	63.6	63.6	63.7
QwQ-32B	76.4	94.6	<b>93.1</b>	79.3	79.8	<b>76.9</b>	<b>73.7</b>	50.0	77.2	66.7	<b>74.8</b>	<b>85.4</b>	64.4	<b>80.5</b>	65.6	77.4	75.0

### K.2.2 FEW-SHOT RESULTS

1614  
 1615  
 1616  
 1617  
 1618  
 1619 The 1-shot results for the Chinese portion of IPBench are shown in Table 32 and Table 33, the 2-shot  
 results in Table 34 and Table 35 and the 3-shot results in Table 36 and Table 37.

1620 Table 30: Results of Chinese Patent IPC Classification task (1-5-1). The best-performing model in  
 1621 each task is in **darker purple**, and the second best is in **lighter purple**.  
 1622

Model	IPC Classification (1-5-1)			
	Exact-Match	Section	Class	Subclass
GPT-4o	8.0	76.4	70.5	62.1
GPT-4o-mini	1.0	76.0	66.1	53.7
DeepSeek-V3	<b>20.2</b>	80.4	72.8	66.3
Qwen3	3.4	76.8	61.0	47.6
Qwen2.5-72B-it	8.0	79.8	71.0	62.3
Qwen2.5-7B-it	2.3	68.6	57.9	46.3
Llama3.1-70B-it	5.0	77.9	64.8	55.2
Llama3.1-8B-it	0.4	65.1	51.8	34.1
Gemma-2-27B-it	1.0	72.4	56.4	45.9
Gemma-2-9B-it	0.0	66.3	51.0	34.9
Mistral-7B-it	0.0	49.9	26.1	16.8
MoZi-qwen	0.4	34.3	25.5	16.8
DISC-LawLLM	0.0	51.2	30.9	15.5
Hanfei	0.0	17.2	4.2	0.2
DeepSeek-R1	19.6	<b>83.2</b>	<b>75.4</b>	<b>67.6</b>
DS-Qwen-7B	0.0	29.3	8.5	1.2
QwQ-32B	3.9	80.9	71.5	60.7

1644  
 1645 Table 31: Results of Chinese generation tasks (4-1 and 4-2). The best-performing model in each  
 1646 task is in **darker blue**, and the second best is in **lighter blue**. R-L refers to ROUGE-L, BS refers  
 1647 to BERTScore, LLMScore refers to GPT-4 judge score (1-10), Avg Tokens # denotes the average  
 1648 number of tokens in the generated text, and Avg DC # indicates the average number of generated  
 1649 dependent claims.  
 1650

Model	Abstract Generation (4-1)					Dependent Claim Generation (4-2)					
	BLEU	R-L	BS	LLMScore (1-10)	Tokens # (167.9)	BLEU	R-L	BS	LLMScore (1-10)	Tokens # (457.8)	DC # (4.1)
GPT-4o	17.7	34.8	91.0	8.77	278.7	12.7	25.0	90.3	6.30	658.3	6.8
GPT-4o-mini	17.9	35.2	90.9	8.51	224.9	<b>15.0</b>	28.6	90.3	6.09	497.9	11.8
DeepSeek-V3	12.4	29.9	90.5	<b>8.92</b>	273.8	10.8	23.4	90.0	<b>7.36</b>	799.7	14.9
Qwen2.5-72B-it	13.7	30.8	90.8	8.50	379.6	9.6	20.3	<b>90.6</b>	6.60	1374.2	17.4
Qwen2.5-7B-it	20.5	36.8	91.1	8.29	190.4	11.0	21.5	<b>90.6</b>	5.64	3453.3	43.5
Llama3.1-70B-it	24.9	40.3	91.1	7.89	261.2	7.3	19.8	89.7	4.99	4045.2	43.0
Llama3.1-8B-it	11.9	26.7	90.2	7.16	554.9	4.7	13.1	90.1	3.09	4932.2	36.1
Gemma-2-27B-it	14.7	30.6	90.2	7.74	215.5	6.5	19.1	88.6	5.46	678.5	3.1
Gemma-2-9B-it	17.7	33.9	90.6	8.07	247.8	5.7	20.8	88.3	5.15	577.1	5.8
Mistral-7B-it	11.8	26.2	90.4	7.24	479.5	3.5	10.7	88.8	2.13	4968.3	44.0
MoZi-qwen	<b>31.3</b>	<b>53.4</b>	<b>91.6</b>	7.91	335.9	7.7	28.6	90.3	4.28	8306.6	59.8
DeepSeek-R1	8.9	28.0	89.3	7.89	671.0	9.3	26.8	81.8	7.17	1374.2	15.8
DS-Qwen-7B	12.4	36.1	90.3	7.80	918.2	5.4	<b>32.5</b>	81.8	3.69	9878.2	89.8
QwQ-32B	11.1	33.6	90.2	8.84	1403.8	5.4	22.6	80.8	7.05	5360.0	37.8

1667  
 1668  
 1669  
 1670 K.2.3 CHAIN-OF-THOUGHT RESULTS  
 1671  
 1672  
 1673 The chain-of-thought results for the Chinese portion of IPBench are presented in Table 38 and Ta-  
 ble 39.

1674 Table 32: Chinese questions results of IPBench with 1-shot setting. The best-performing model in  
 1675 each task is in **darker red**, and the second best is in **lighter red**.

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>73.1</b>	91.5	85.9	72.5	<b>76.6</b>	67.2	65.9	43.6	68.9	<b>53.2</b>	66.9	81.4	63.1	<b>74.2</b>	<b>79.2</b>	77.0	76.5
Qwen2.5-7B-it	69.9	94.6	86.2	72.5	73.8	<b>71.0</b>	<b>66.6</b>	43.6	71.5	50.0	59.7	76.1	58.1	64.2	60.0	<b>82.5</b>	70.1
Llama3.1-8B-it	58.9	81.5	64.1	60.9	73.0	59.2	61.7	43.6	69.3	38.4	61.2	45.9	56.3	64.8	38.0	73.2	65.6
Gemma-2-9B-it	68.1	86.9	74.3	67.4	69.4	64.8	60.7	<b>46.8</b>	<b>79.4</b>	45.5	<b>67.6</b>	<b>82.7</b>	<b>64.4</b>	64.8	56.8	74.7	<b>79.0</b>
Mistral-7B-it	53.7	71.0	56.2	58.5	59.5	47.0	52.3	38.8	66.2	42.3	43.2	62.1	53.8	35.2	41.4	65.4	61.6
MoZi-qwen	67.5	<b>95.0</b>	<b>87.3</b>	<b>73.8</b>	61.9	<b>71.0</b>	63.0	44.8	71.9	39.4	64.0	77.4	63.1	60.4	48.2	80.7	71.7
DISC-LawLLM	55.7	79.2	68.8	62.9	69.1	51.5	44.8	38.8	56.6	35.9	46.0	70.8	55.6	38.4	–	75.1	31.1
Hanfei	31.6	34.0	23.9	23.1	33.7	33.1	49.0	24.0	49.1	35.9	29.5	24.6	38.1	18.9	–	33.6	27.4

1686  
 1687 Table 33: Results of Chinese Patent IPC Classification task (1-5-1) with 1-shot setting. The best-  
 1688 performing model in each task is in **darker purple**, and the second best is in **lighter purple**.

Model	IPC Classification (1-5-1)			
	Exact-Match	Section	Class	Subclass
GPT-4o-mini	3.6	76.8	67.1	54.9
DeepSeek-V3	<b>30.2</b>	<b>82.4</b>	<b>75.4</b>	<b>68.9</b>
Qwen2.5-7B-it	3.8	59.6	50.9	39.6
Llama3.1-8B-it	0.8	45.9	36.4	25.5
MoZi-qwen	0.0	11.6	8.2	3.6

1700 Table 34: Chinese questions results of IPBench with 2-shot setting. The best-performing model in  
 1701 each task is in **darker red**, and the second best is in **lighter red**.

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>73.5</b>	91.9	<b>87.3</b>	69.1	76.2	66.3	67.2	44.8	64.5	<b>58.3</b>	<b>69.8</b>	83.4	60.0	<b>73.0</b>	<b>81.0</b>	73.7	78.7
Qwen2.5-7B-it	70.8	95.0	86.2	70.8	76.2	69.5	<b>68.8</b>	39.6	68.9	51.9	63.3	83.7	54.4	60.4	67.2	83.0	65.6
Llama3.1-8B-it	58.9	79.5	63.4	63.6	69.8	58.3	59.1	46.8	64.0	36.5	65.5	66.8	53.1	54.7	38.0	66.4	61.3
Gemma-2-9B-it	68.7	86.5	76.1	66.7	70.6	65.4	60.4	46.8	<b>75.9</b>	46.8	68.4	<b>84.7</b>	61.3	63.5	58.6	77.9	<b>79.0</b>
Mistral-7B-it	56.6	73.0	58.0	58.5	63.1	49.1	52.6	44.4	63.2	42.3	51.1	73.1	55.6	37.1	48.6	66.4	61.0
MoZi-qwen	70.3	<b>95.8</b>	<b>87.3</b>	<b>75.2</b>	<b>76.6</b>	<b>74.9</b>	65.3	<b>47.6</b>	71.1	43.6	67.6	82.4	<b>68.1</b>	61.0	52.2	<b>83.4</b>	70.1
DISC-LawLLM	57.6	76.5	66.3	62.9	68.3	56.8	45.8	38.0	56.1	36.5	51.1	78.1	54.4	34.6	–	74.7	43.9
Hanfei	31.7	30.5	19.6	40.1	30.2	35.2	34.4	25.6	35.5	22.4	25.9	49.2	44.4	12.0	–	32.7	32.3

1714 Table 35: Results of Chinese Patent IPC Classification task (1-5-1) with 2-shot setting. The best-  
 1715 performing model in each task is in **darker purple**, and the second best is in **lighter purple**.

Model	IPC Classification (1-5-1)			
	Exact-Match	Section	Class	Subclass
GPT-4o-mini	2.5	75.4	65.9	54.5
DeepSeek-V3	<b>29.4</b>	<b>82.8</b>	<b>76.0</b>	<b>69.1</b>
Qwen2.5-7B-it	3.6	70.5	60.0	47.6
Llama3.1-8B-it	1.0	34.9	28.4	18.9
MoZi-qwen	0.4	29.9	23.1	15.4

Table 36: Chinese questions results of IPBench with 3-shot setting. The best-performing model in each task is in **darker red**, and the second best is in **lighter red**.

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>74.7</b>	91.5	<b>87.3</b>	73.5	<b>77.0</b>	70.1	67.9	47.2	64.0	<b>59.6</b>	<b>71.9</b>	82.7	62.5	<b>73.6</b>	<b>80.6</b>	79.3	<b>79.3</b>
Qwen2.5-7B-it	72.1	95.0	84.4	69.4	<b>77.0</b>	71.6	<b>69.8</b>	45.6	73.7	43.6	65.5	84.7	58.1	65.4	69.6	<b>83.4</b>	68.9
Llama3.1-8B-it	59.3	81.1	60.9	59.5	68.7	60.9	64.0	47.6	57.5	32.1	58.3	75.4	53.1	55.3	37.8	67.7	64.6
Gemma-2-9B-it	68.7	87.6	75.0	66.0	69.0	64.5	63.3	47.6	<b>75.0</b>	48.1	69.8	<b>85.1</b>	63.8	61.6	60.0	76.5	77.4
Mistral-7B-it	56.3	72.2	59.8	56.8	65.1	51.2	53.6	45.6	62.3	41.0	52.5	73.8	58.1	39.0	43.4	65.4	60.7
MoZi-qwen	69.2	<b>95.8</b>	87.0	<b>74.5</b>	75.8	<b>72.5</b>	65.3	<b>51.2</b>	70.6	37.8	68.4	80.4	<b>67.5</b>	55.4	47.4	81.6	72.9
DISC-LawLLM	58.0	81.9	69.2	60.9	70.6	61.0	47.4	40.4	53.5	35.9	53.2	74.4	58.1	34.0	–	67.7	43.9
Hanfei	30.1	27.8	22.8	28.9	25.8	34.3	34.7	27.2	41.7	24.4	20.9	30.6	43.1	22.0	–	27.2	39.9

Table 37: Results of Chinese Patent IPC Classification task (1-5-1) with 3-shot setting. The best-performing model in each task is in **darker purple**, and the second best is in **lighter purple**.

Model	IPC Classification (1-5-1)			
	Exact-Match	Section	Class	Subclass
GPT-4o-mini	2.3	76.4	66.3	55.6
DeepSeek-V3	<b>29.4</b>	<b>83.8</b>	<b>76.9</b>	<b>70.8</b>
Qwen2.5-7B-it	3.6	71.1	60.2	48.0
Llama3.1-8B-it	0.4	46.1	33.9	23.2
MoZi-qwen	0.6	59.4	42.1	31.4

Table 38: Chinese questions results of IPBench with chain-of-thought setting. The best-performing model in each task is in **darker red**, and the second best is in **lighter red**.

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	2-5	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>72.4</b>	92.3	84.1	70.8	77.8	<b>65.4</b>	<b>66.9</b>	45.2	68.9	<b>59.6</b>	60.4	<b>80.4</b>	57.5	66.0	<b>81.00</b>	75.1	<b>71.3</b>
Qwen2.5-7B-it	70.9	<b>93.4</b>	<b>86.2</b>	<b>72.8</b>	<b>79.4</b>	63.0	65.3	46.0	74.1	55.1	<b>61.9</b>	79.7	<b>59.4</b>	<b>69.8</b>	67.00	<b>78.3</b>	67.1
Llama3.1-8B-it	62.3	83.4	63.8	62.9	66.7	58.3	55.8	<b>52.4</b>	66.7	39.1	51.8	71.8	58.8	51.6	62.20	71.4	61.0
Gemma-2-9B-it	63.1	85.3	69.9	57.8	63.1	59.8	52.3	46.8	<b>75.0</b>	49.4	59.7	76.4	<b>59.4</b>	61.0	53.80	67.3	70.7
Mistral-7B-it	54.5	76.1	55.8	57.8	60.3	48.8	50.0	45.6	59.7	45.5	51.1	65.8	56.9	39.6	44.80	61.8	54.3
MoZi-qwen	60.5	91.5	80.4	64.6	57.9	55.3	60.7	39.2	54.8	44.9	54.0	72.4	45.0	55.3	58.00	65.9	45.4
DISC-LawLLM	52.5	83.8	66.3	62.9	63.5	61.0	49.0	37.2	58.8	25.6	48.2	62.8	43.8	37.7	30.40	60.4	41.5
Hanfei	29.1	44.0	25.0	35.0	37.7	34.9	26.3	24.8	25.9	22.4	25.2	24.3	35.0	17.6	23.00	24.4	38.4

Table 39: Results of Chinese Patent IPC Classification task (1-5-1) with chain-of-thought setting. The best-performing model in each task is in **darker purple**, and the second best is in **lighter purple**.

Model	IPC Classification (1-5-1)			
	Exact-Match	Section	Class	Subclass
GPT-4o-mini	0.4	75.4	67.6	56.0
DeepSeek-V3	<b>2.5</b>	<b>77.1</b>	<b>70.9</b>	<b>64.2</b>
Qwen2.5-7B-it	1.7	58.7	50.1	36.4
Llama3.1-8B-it	0.2	56.6	41.3	26.7
MoZi-qwen	0.0	3.6	2.9	2.1

### K.3 ENGLISH QUESTIONS RESULTS

In Section [Section K.3](#), we focus on the IPBench results for English questions. We provide the zero-shot results for the English portion of IPBench in [Section K.3.1](#), the few-shot results in [Section K.3.2](#), and the chain-of-thought results in [Section K.3.3](#).

## K.3.1 ZERO-SHOT RESULTS

The zero-shot results for the English portion of IPBench are shown in Table 40, Table 41 and Table 42.

Table 40: English questions results of IPBench. The best-performing model in each task is in **darker red**, and the second best is in **lighter red**. The model DS-Qwen refers to DeepSeek-R1-Distill-Qwen, while the suffix *it* indicates the Instruct version of the model. OA denotes the overall average accuracy on the choice tasks.

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4	3-5	4-3
GPT-4o	<b>73.2</b>	<b>97.1</b>	91.6	82.5	<b>86.9</b>	50.7	69.6	58.8	53.3	60.6	85.8	75.7	77.2	<b>82.4</b>	<b>84.2</b>	<b>50.0</b>	<b>68.4</b>
GPT-4o-mini	71.4	<b>97.1</b>	88.9	85.0	83.7	49.3	68.8	56.0	62.1	61.3	85.2	77.1	69.2	81.0	79.2	44.0	54.4
DeepSeek-V3	72.9	95.9	90.3	90.3	82.5	50.7	67.1	<b>65.6</b>	55.9	66.9	82.1	77.1	79.2	79.6	83.1	44.6	67.3
Qwen3	68.2	93.4	80.1	73.8	73.4	49.8	65.0	59.2	<b>60.7</b>	65.0	83.3	<b>77.9</b>	75.8	73.6	77.6	44.0	49.7
Qwen2.5-72B-it	71.6	95.4	88.1	87.4	85.3	48.4	67.5	60.8	54.4	63.8	85.2	76.4	78.5	78.0	82.5	43.3	61.4
Qwen2.5-7B-it	65.2	90.9	80.1	84.0	77.4	49.3	60.0	51.2	54.4	62.5	80.9	75.7	75.8	64.0	78.1	38.9	41.5
Llama3.1-70B-it	70.3	96.3	89.4	86.4	83.3	48.4	67.1	58.8	58.8	58.1	81.5	75.0	<b>82.6</b>	71.2	77.6	45.2	59.7
Llama3.1-8B-it	58.2	92.5	78.8	71.8	64.0	37.0	62.9	44.8	47.6	47.5	79.0	61.4	66.4	45.4	74.9	41.7	45.6
Gemma-2-27B-it	67.0	92.5	83.2	83.5	79.8	45.7	62.9	59.6	54.8	58.8	85.2	72.9	71.1	51.6	80.3	–	48.0
Gemma-2-9B-it	64.3	91.7	81.9	79.6	59.5	45.7	61.7	57.2	55.1	48.1	79.6	70.0	76.5	52.4	78.7	–	49.1
Mistral-7B-it	55.4	85.5	70.8	68.0	61.5	35.2	57.1	41.6	52.2	44.4	77.8	59.3	63.1	40.4	71.0	43.9	45.0
MoZi-qwen	61.8	94.6	79.6	82.5	73.0	41.6	63.8	52.8	52.9	42.5	77.2	<b>77.1</b>	71.1	48.2	75.4	43.9	40.9
DISC-LawLLM	50.2	70.5	60.6	71.8	56.8	39.3	55.0	43.6	55.2	22.5	71.6	57.1	53.0	28.0	69.4	–	23.4
Hanfei	40.3	60.6	46.0	53.9	37.7	30.6	42.1	33.2	47.4	40.6	44.4	43.6	37.6	27.8	53.0	–	18.7
DeepSeek-R1	71.6	96.3	<b>92.9</b>	<b>92.2</b>	82.1	<b>51.1</b>	<b>70.0</b>	62.4	54.4	<b>77.5</b>	<b>87.0</b>	67.9	77.9	69.6	79.2	47.5	57.9
DS-Qwen-7B	56.0	76.4	58.9	58.3	50.4	48.0	52.5	43.6	49.3	53.8	82.1	57.9	59.1	63.8	63.8	43.6	38.0
QwQ-32B	70.7	95.9	88.5	85.4	75.8	47.0	68.8	64.8	57.7	74.4	84.6	75.7	<b>83.9</b>	69.0	76.5	47.1	59.7

Table 41: Results of English Patent IPC/CPC Classification tasks (1-5-1 and 1-5-2). The best-performing model in each task is in **darker purple**, and the second best is in **lighter purple**.

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o	2.0	86.2	72.0	49.0	3.3	82.7	69.7	62.0
GPT-4o-mini	1.0	84.5	67.5	47.0	0.5	79.0	64.5	52.7
DeepSeek-V3	2.3	86.7	73.8	51.3	<b>9.5</b>	<b>84.0</b>	<b>73.3</b>	<b>65.2</b>
Qwen3	2.2	84.0	68.2	48.3	0.5	62.7	48.3	38.7
Qwen2.5-72B-it	2.2	84.7	69.8	49.0	2.5	81.5	69.5	60.7
Qwen2.5-7B-it	1.5	84.0	67.5	46.8	0.2	65.5	44.8	34.8
Llama3.1-70B-it	2.2	82.7	66.3	45.3	1.0	79.5	64.3	52.7
Llama3.1-8B-it	1.3	77.7	60.0	37.3	0.0	63.8	45.0	30.7
Gemma-2-27B-it	1.3	73.3	58.3	37.7	0.2	70.5	56.7	44.3
Gemma-2-9B-it	0.5	80.2	59.7	39.2	0.2	56.2	39.0	26.7
Mistral-7B-it	0.2	82.3	57.3	35.5	0.0	39.0	21.5	10.3
MoZi-qwen	0.8	42.8	33.2	23.3	0.0	8.5	3.1	1.8
DISC-LawLLM	0.0	83.0	61.3	39.5	0.0	31.0	23.4	11.5
Hanfei	0.0	6.9	0.0	0.0	0.0	0.9	0.0	0.0
DeepSeek-R1	<b>3.2</b>	<b>88.0</b>	<b>74.0</b>	<b>52.0</b>	8.5	82.5	71.2	63.2
DS-Qwen-7B	0.0	12.9	5.5	1.7	0.0	5.1	0.5	0.2
QwQ-32B	2.0	86.3	69.5	47.7	0.5	76.0	62.3	51.3

## K.3.2 FEW-SHOT RESULTS

The 1-shot results for the English portion of IPBench are shown in Table 43 and Table 44, the 2-shot results in Table 45 and Table 46 and the 3-shot results in Table 47 and Table 48.

Table 42: Results of English generation tasks (4-1 and 4-2). The best-performing model in each task is in **darker blue**, and the second best is in **lighter blue**. R-L refers to ROUGE-L, BS refers to BERTScore, LLMScore refers to GPT-4 judge score (1-10), Avg Tokens # denotes the average number of generated tokens, and Avg DC # denotes the average number of generated dependent claims.

Model	Abstract Generation (4-1)					Dependent Claim Generation (4-2)								
	BLEU	R-L	BS	LLMScore	Tokens #	BLEU	R-L	BS	LLMScore	Tokens #	DC #			
	(1-10)					(129.0)					(417.4)		(13.1)	
GPT-4o	17.7	27.3	87.7	8.07	264.2	25.2	28.0	87.4	6.97	637.4	6.2			
GPT-4o-mini	28.8	28.6	88.4	7.59	211.3	25.5	28.0	86.5	6.66	458.4	1.1			
DeepSeek-V3	26.7	26.8	87.4	7.84	218.3	<b>27.4</b>	30.1	<b>88.0</b>	<b>7.54</b>	583.7	14.8			
Qwen2.5-72B-it	28.3	30.4	88.3	<b>8.17</b>	272.5	10.3	13.9	87.8	6.01	6207.6	120.8			
Qwen2.5-7B-it	34.2	34.6	89.3	8.07	227.9	19.2	23.1	87.8	5.71	3569.3	48.0			
Llama3.1-70B-it	<b>37.1</b>	36.1	<b>89.7</b>	8.07	191.8	24.8	27.8	86.5	6.36	543.6	13.6			
Llama3.1-8B-it	28.4	30.1	88.1	7.79	359.7	11.4	14.6	86.7	4.64	7643.6	145.4			
Gemma-2-27B-it	24.7	24.4	87.6	7.54	171.2	23.8	26.1	86.0	6.49	486.1	3.4			
Gemma-2-9B-it	25.5	24.9	87.4	7.76	190.7	23.8	25.6	86.0	5.95	446.7	7.0			
Mistral-7B-it	28.6	28.6	88.4	7.75	243.9	10.9	12.8	87.3	4.72	8117.8	148.5			
MoZi-qwen	31.1	<b>48.6</b>	89.1	7.56	296.8	24.8	<b>40.3</b>	87.7	5.34	1936.5	35.5			
DeepSeek-R1	18.7	27.6	85.7	7.55	613.6	23.8	31.8	61.1	7.19	1231.6	22.4			
DS-Qwen-7B	7.0	9.7	76.8	7.36	686.9	17.9	32.2	56.2	4.62	2315.7	18.4			
QwQ-32B	22.2	30.4	85.6	<b>8.17</b>	849.5	19.8	29.1	63.1	7.14	4635.4	45.7			

Table 43: English questions results of IPBench with 1-shot setting. The best-performing model in each task is in **darker red**, and the second best is in **lighter red**.

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>75.4</b>	<b>96.3</b>	<b>86.7</b>	<b>87.4</b>	<b>82.9</b>	<b>52.1</b>	67.5	<b>59.6</b>	<b>57.4</b>	<b>69.4</b>	<b>85.8</b>	63.1	<b>79.9</b>	<b>83.0</b>	<b>83.1</b>	<b>58.5</b>
Qwen2.5-7B-it	65.1	93.8	77.9	82.0	71.8	45.2	59.6	54.0	54.0	54.4	79.6	58.1	75.2	51.2	76.0	46.8
Llama3.1-8B-it	59.9	93.0	<b>76.5</b>	76.7	71.0	37.9	<b>68.3</b>	48.4	47.8	48.8	76.5	56.3	75.2	35.6	73.2	40.9
Gemma-2-9B-it	64.6	91.7	74.3	76.2	77.0	41.6	60.8	54.8	52.9	54.4	77.2	<b>64.4</b>	77.2	51.0	77.0	51.5
Mistral-7B-it	56.2	87.6	68.1	70.4	60.3	37.9	53.8	50.4	50.0	30.6	77.8	53.8	62.4	37.4	69.4	48.5
MoZi-qwen	57.6	95.4	74.4	79.6	71.8	42.9	66.3	49.6	54.8	39.4	79.0	63.1	57.1	25.0	61.2	31.6
DISC-LawLLM	54.1	78.4	64.6	71.4	61.9	35.2	61.7	41.2	52.9	28.1	56.8	55.6	40.9	–	57.9	29.8
Hanfei	32.9	40.3	35.8	36.9	28.6	35.2	36.7	29.2	27.2	35.6	30.3	38.1	26.2	–	27.9	35.1

Table 44: Results of English Patent IPC/CPC Classification tasks (1-5-1 and 1-5-2) with 1-shot setting. The best-performing model in each task is in **darker purple**, and the second best is in **lighter purple**.

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o-mini	1.0	86.2	67.0	47.3	0.5	74.3	59.1	49.1
DeepSeek-V3	<b>2.0</b>	<b>89.7</b>	<b>75.7</b>	<b>53.2</b>	<b>7.3</b>	<b>86.2</b>	<b>74.3</b>	<b>65.0</b>
Qwen2.5-7B-it	0.8	86.2	64.0	44.5	0.3	67.5	48.8	37.2
Llama3.1-8B-it	0.6	80.0	61.3	40.8	0.0	45.2	35.2	22.2
MoZi-qwen	0.7	78.0	57.7	37.8	0.0	16.5	7.8	4.3

### K.3.3 CHAIN-OF-THOUGHT RESULTS

The chain-of-thought results for the English portion of IPBench are presented in Table 49 and Table 50.

1890 Table 45: English questions results of IPBench with 2-shot setting. The best-performing model in  
 1891 each task is in **darker red**, and the second best is in **lighter red**.  
 1892

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>75.2</b>	<b>96.7</b>	<b>88.5</b>	<b>88.4</b>	<b>84.1</b>	<b>52.1</b>	<b>65.8</b>	<b>59.2</b>	<b>55.5</b>	<b>68.1</b>	<b>87.7</b>	77.1	<b>84.6</b>	<b>80.4</b>	<b>80.3</b>	<b>60.8</b>
Qwen2.5-7B-it	68.3	94.6	79.7	83.5	77.8	45.2	60.0	56.0	53.3	60.0	79.0	<b>82.9</b>	79.2	64.6	79.8	42.7
Llama3.1-8B-it	60.1	92.5	72.1	75.2	69.4	43.4	65.0	49.2	50.4	54.4	76.5	67.9	68.5	37.6	73.8	39.8
Gemma-2-9B-it	66.0	91.7	76.5	78.2	73.8	42.5	62.1	<b>56.0</b>	51.8	57.5	80.2	72.9	81.9	56.2	78.1	52.6
Mistral-7B-it	58.0	87.1	71.7	73.3	59.9	37.4	55.4	49.6	51.1	33.8	77.2	68.8	61.1	45.6	68.9	48.0
MoZi-qwen	62.7	96.3	<b>80.1</b>	80.1	73.4	46.6	62.9	49.2	52.9	42.5	81.5	76.4	68.5	46.6	77.6	26.3
DISC-LawLLM	55.8	82.2	68.1	71.8	62.3	40.2	58.3	41.2	51.8	26.3	64.2	61.4	50.3	–	67.8	24.0
Hanfei	32.8	41.9	31.0	29.6	37.3	35.6	31.3	25.2	38.2	42.5	27.2	40.0	32.9	–	31.7	12.9

1902  
 1903 Table 46: Results of English Patent IPC/CPC Classification tasks (1-5-1 and 1-5-2) with 2-shot  
 1904 setting. The best-performing model in each task is in **darker purple**, and the second best is in  
 1905  
 1906 lighter purple .  
 1907

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o-mini	2.17	88.2	69.8	48.8	0.2	76.3	61.6	51.6
DeepSeek-V3	<b>2.67</b>	<b>90.2</b>	<b>76.2</b>	<b>53.2</b>	<b>7.2</b>	<b>86.5</b>	<b>73.3</b>	<b>65.7</b>
Qwen2.5-7B-it	1.50	85.0	64.5	45.0	0.3	68.5	51.2	38.8
Llama3.1-8B-it	1.30	81.5	59.0	38.2	0.0	63.3	45.7	26.7
MoZi-qwen	0.83	80.0	58.0	36.7	0.2	32.3	17.3	9.3

1917 Table 47: English questions results of IPBench with 3-shot setting. The best-performing model in  
 1918 each task is in **darker red**, and the second best is in **lighter red**.  
 1919

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4	4-3
GPT-4o-mini	<b>74.7</b>	<b>97.5</b>	<b>87.6</b>	<b>88.4</b>	<b>82.9</b>	<b>53.0</b>	<b>70.0</b>	<b>57.6</b>	54.0	<b>67.5</b>	<b>87.0</b>	79.3	<b>82.6</b>	<b>79.4</b>	<b>80.9</b>	<b>66.7</b>
Qwen2.5-7B-it	70.6	93.4	81.9	79.6	75.0	48.0	66.3	56.0	52.6	58.8	82.7	<b>85.0</b>	79.9	67.4	<b>80.9</b>	43.9
Llama3.1-8B-it	59.4	94.2	74.4	76.3	71.0	40.2	63.3	43.6	53.7	53.8	73.5	70.0	70.5	35.6	73.8	40.4
Gemma-2-9B-it	67.4	91.3	77.4	77.2	72.2	44.7	61.4	54.8	51.5	55.0	82.1	73.6	77.9	57.4	76.0	57.9
Mistral-7B-it	56.5	89.6	69.0	71.4	58.3	37.9	55.4	49.2	50.7	31.3	74.1	69.3	62.4	36.8	70.5	53.2
MoZi-qwen	65.3	96.7	79.2	81.1	76.2	47.0	65.0	47.6	54.4	43.1	82.1	78.6	69.1	29.1	76.0	27.5
DISC-LawLLM	56.6	85.9	65.0	69.9	62.7	42.5	57.9	42.0	<b>55.9</b>	23.1	69.8	62.1	49.0	–	67.2	28.1
Hanfei	28.9	36.5	36.3	22.3	31.0	27.9	27.5	20.0	31.6	33.8	24.1	38.6	26.9	–	25.1	24.0

1929 Table 48: Results of English Patent IPC/CPC Classification tasks (1-5-1 and 1-5-2) with 3-shot setting.  
 1930 The best-performing model in each task is in **darker purple**, and the second best is in  
 1931  
 1932 lighter purple .  
 1933

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o-mini	1.8	87.8	69.2	46.8	0.3	80.1	65.3	54.1
DeepSeek-V3	<b>3.5</b>	<b>90.0</b>	<b>75.7</b>	<b>52.8</b>	<b>7.7</b>	<b>85.7</b>	<b>73.3</b>	<b>64.7</b>
Qwen2.5-7B-it	1.2	85.7	64.8	45.7	0.5	68.3	50.8	38.8
Llama3.1-8B-it	1.8	82.7	62.0	41.2	0.0	64.8	45.8	29.8
MoZi-qwen	1.3	80.3	59.3	36.7	0.0	24.2	12.8	7.7

1944 Table 49: English questions results of IPBench with chain-of-thought setting. The best-performing  
 1945 model in each task is in **darker red**, and the second best is in **lighter red**.  
 1946

Model	OA	1-1	1-2	1-3	1-4	1-6	1-7	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4	3-5	4-3
GPT-4o-mini	<b>72.2</b>	<b>96.7</b>	<b>88.1</b>	<b>88.4</b>	<b>82.9</b>	<b>51.6</b>	<b>67.9</b>	<b>57.6</b>	<b>57.4</b>	<b>64.4</b>	<b>87.0</b>	<b>75.0</b>	<b>76.5</b>	<b>81.2</b>	<b>83.1</b>	<b>44.9</b>	<b>58.5</b>
Qwen2.5-7B-it	64.4	84.2	78.8	78.6	73.0	48.9	60.0	50.0	55.9	61.3	84.0	<b>75.0</b>	70.5	63.2	78.8	44.3	43.3
Llama3.1-8B-it	60.4	84.9	76.6	74.3	73.8	46.6	64.2	46.4	55.5	52.5	79.0	66.4	67.8	47.4	74.9	43.6	40.9
Gemma-2-9B-it	60.3	88.8	75.2	76.2	69.4	45.7	60.0	55.2	56.3	46.3	80.9	74.3	75.2	62.0	80.9	–	54.4
Mistral-7B-it	54.1	85.5	72.6	71.8	64.7	35.6	59.2	39.2	50.4	43.1	75.3	56.4	63.1	37.2	72.1	36.3	35.7
MoZi-qwen	60.5	94.6	79.2	82.5	72.6	42.5	63.3	51.2	50.4	45.6	77.8	73.6	70.5	40.4	77.1	43.9	41.5
DISC-LawLLM	19.5	45.6	46.5	28.2	15.1	15.1	31.7	13.6	14.7	26.3	18.5	7.1	12.8	13.6	4.6	–	9.4
Hanfei	30.5	39.8	32.7	28.6	30.6	24.2	27.5	32.8	32.7	20.6	26.5	33.6	36.2	26.2	40.4	–	29.2

1956 Table 50: Results of English Patent IPC/CPC Classification tasks (1-5-1 and 1-5-2) with chain-of-thought setting. The best-performing model in each task is in **darker purple**, and the second best is in **lighter purple**.  
 1957

Model	IPC Classification (1-5-1)				CPC Classification (1-5-2)			
	Exact-Match	Section	Class	Subclass	Exact-Match	Section	Class	Subclass
GPT-4o-mini	0.0	84.8	67.3	47.8	0.0	76.8	63.0	52.5
DeepSeek-V3	0.3	86.8	<b>73.0</b>	<b>51.5</b>	<b>1.0</b>	<b>83.3</b>	<b>70.7</b>	<b>63.0</b>
Qwen2.5-7B-it	<b>1.8</b>	<b>88.0</b>	69.5	47.0	0.5	60.2	46.0	35.7
Llama3.1-8B-it	1.5	76.2	58.5	37.5	0.2	64.0	44.8	29.5
MoZi-qwen	0.5	38.8	29.8	21.8	0.0	7.7	2.8	1.8

#### K.4 LLM-AS-A-JUDGE RESULTS

1971 We provide detailed results of the LLM-as-a-judge evaluation for the overall, Chinese, and English  
 1972 parts. The evaluation includes four dimensions and an overall score, as shown in Table 51, Table 52,  
 1973 and Table 53. The definitions of these metrics are provided in Section H.3, with all scores ranging  
 1974 from 1 to 10.  
 1975

1976 Table 51: Multi-dimension results of generation tasks (4-1 and 4-2) in LLM-as-a-judge. The best-  
 1977 performing model in each task is in **darker blue**, and the second best is in **lighter blue**. Accuracy  
 1978 (Acc.), Relevance (Rel.), Completeness (Comp.), Consistency (Cons.), L-S and LLMScore are  
 1979 generation quality metrics rated by an LLM-as-a-judge.  
 1980

Model	Abstract Generation (4-1)						Dependent Claim Generation (4-2)					
	Acc.	Rel.	Comp.	Cons.	L-S	LLMScore	Acc.	Rel.	Comp.	Cons.	L-S	LLMScore
GPT-4o	8.45	8.24	8.68	9.27	7.58	8.42	7.45	6.28	6.22	6.58	7.17	6.63
GPT-4o-mini	7.99	8.02	8.13	8.94	7.47	8.05	7.17	5.92	6.06	6.30	7.02	6.37
DeepSeek-V3	8.26	<b>8.45</b>	8.53	9.15	<b>7.73</b>	8.38	<b>7.93</b>	<b>7.30</b>	<b>7.13</b>	<b>7.38</b>	<b>7.92</b>	<b>7.45</b>
Qwen2.5-72B-it	8.40	8.18	8.70	<b>9.36</b>	7.37	8.33	7.13	5.77	6.00	6.35	6.72	6.30
Qwen2.5-7B-it	8.17	8.14	8.19	9.08	7.61	8.18	6.59	5.47	5.09	5.68	5.96	5.67
Llama3.1-70B-it	7.98	8.03	7.94	8.96	7.31	7.98	6.57	5.38	5.16	5.69	6.21	5.67
Llama3.1-8B-it	7.52	7.41	7.71	8.57	6.54	7.47	4.70	3.95	3.18	3.91	4.15	3.86
Gemma-2-27B-it	7.63	7.78	7.46	8.40	7.32	7.64	6.51	5.56	5.71	5.84	6.54	5.98
Gemma-2-9B-it	7.89	8.03	7.82	8.76	7.43	7.91	6.21	5.23	5.20	5.51	6.12	5.55
Mistral-7B-it	7.47	7.38	7.86	8.62	6.40	7.49	4.19	3.30	3.07	3.38	3.71	3.42
MoZi-qwen	7.71	7.88	7.78	8.76	7.02	7.73	5.82	4.70	4.00	4.83	5.17	4.81
DeepSeek-R1	7.70	7.75	7.88	8.39	7.21	7.72	7.73	6.76	7.00	7.16	7.69	7.18
DS-Qwen-7B	7.58	7.50	7.78	8.43	6.90	7.58	4.67	4.02	3.97	4.01	4.60	4.16
QwQ-32B	<b>8.48</b>	8.39	<b>8.80</b>	9.27	<b>7.74</b>	<b>8.51</b>	7.63	6.61	6.97	7.13	7.61	7.10

Table 52: Multi-dimension results of Chinese generation tasks (4-1 and 4-2) in LLM-as-a-judge. The best-performing model in each task is in **darker blue**, and the second best is in **lighter blue**.

Model	Abstract Generation (4-1)						Dependent Claim Generation (4-2)					
	Acc.	Rel.	Comp.	Cons.	L-S	LLMScore	Acc.	Rel.	Comp.	Cons.	L-S	LLMScore
GPT-4o	8.66	8.77	8.96	9.50	8.00	8.77	7.06	5.85	5.91	6.27	6.94	6.30
GPT-4o-mini	8.39	8.59	8.46	9.24	8.00	8.51	6.93	5.59	5.69	6.10	6.85	6.09
DeepSeek-V3	8.80	9.05	9.08	<b>9.54</b>	<b>8.27</b>	<b>8.92</b>	<b>7.90</b>	<b>6.98</b>	<b>7.09</b>	<b>7.40</b>	<b>7.93</b>	<b>7.36</b>
Qwen2.5-72B-it	8.53	8.55	8.86	9.47	7.56	8.50	7.33	6.14	6.36	6.61	7.12	6.60
Qwen2.5-7B-it	8.21	8.50	8.09	9.05	8.03	8.29	6.56	5.45	5.11	5.58	5.89	5.64
Llama3.1-70B-it	7.85	8.22	7.76	8.88	7.26	7.89	6.04	4.80	4.37	5.00	5.45	4.99
Llama3.1-8B-it	7.12	7.30	7.47	8.31	6.11	7.16	3.55	3.48	2.47	3.03	3.49	3.09
Gemma-2-27B-it	7.65	7.98	7.52	8.45	7.49	7.74	5.89	5.14	5.33	5.19	5.92	5.46
Gemma-2-9B-it	7.94	8.30	7.98	8.86	7.52	8.07	5.82	4.87	4.88	5.14	5.68	5.15
Mistral-7B-it	7.20	7.35	7.73	8.50	6.03	7.24	2.50	2.22	1.96	2.05	2.36	2.13
MoZi-qwen	7.79	8.40	7.60	8.75	7.47	7.91	5.33	4.23	3.42	4.34	4.52	4.28
DeepSeek-R1	7.76	8.01	8.05	8.30	7.42	7.89	7.74	6.74	6.98	7.16	7.69	7.17
DS-Qwen-7B	7.68	7.84	8.08	8.10	6.97	7.80	4.26	3.71	3.30	4.47	4.00	3.69
QwQ-32B	<b>8.71</b>	<b>8.82</b>	<b>9.10</b>	9.09	8.09	8.84	7.70	6.55	6.92	7.13	7.58	7.05

Table 53: Multi-dimension results of English generation tasks (4-1 and 4-2) in LLM-as-a-judge. The best-performing model in each task is in **darker blue**, and the second best is in **lighter blue**.

Model	Abstract Generation (4-1)						Dependent Claim Generation (4-2)					
	Acc.	Rel.	Comp.	Cons.	L-S	LLMScore	Acc.	Rel.	Comp.	Cons.	L-S	LLMScore
GPT-4o	8.24	7.70	8.40	9.04	7.16	8.07	7.85	6.71	6.53	6.88	7.40	6.97
GPT-4o-mini	7.59	7.45	7.81	8.64	6.94	7.59	7.42	6.26	6.43	6.49	7.20	6.66
DeepSeek-V3	7.72	<b>7.86</b>	7.99	8.76	7.20	7.84	<b>7.96</b>	<b>7.63</b>	<b>7.17</b>	<b>7.36</b>	<b>7.92</b>	<b>7.54</b>
Qwen2.5-72B-it	<b>8.28</b>	7.82	<b>8.54</b>	9.25	7.19	<b>8.17</b>	6.93	5.40	5.64	6.09	6.33	6.01
Qwen2.5-7B-it	8.12	7.78	8.30	9.10	7.20	8.07	6.63	5.48	5.07	5.78	6.03	5.71
Llama3.1-70B-it	8.12	7.84	8.12	9.03	7.36	8.07	7.10	5.96	5.95	6.38	6.97	6.36
Llama3.1-8B-it	7.93	7.51	7.96	8.82	6.98	7.79	5.84	4.43	3.89	4.79	4.80	4.64
Gemma-2-27B-it	7.61	7.58	7.40	8.35	7.15	7.54	7.13	5.97	6.10	6.48	7.17	6.49
Gemma-2-9B-it	7.85	7.75	7.67	8.65	7.34	7.76	6.60	5.59	5.53	5.88	6.56	5.95
Mistral-7B-it	7.75	7.42	7.99	8.73	6.76	7.75	5.89	4.39	4.18	4.71	5.07	4.72
MoZi-qwen	7.63	7.37	7.97	8.77	6.57	7.56	6.32	5.18	4.59	5.32	5.83	5.34
DeepSeek-R1	7.63	7.48	7.70	8.47	6.99	7.55	7.71	<b>6.77</b>	7.02	7.16	7.68	7.19
DS-Qwen-7B	7.48	7.15	7.48	8.76	6.82	7.36	5.07	4.32	4.64	3.55	5.19	4.62
QwQ-32B	8.25	<b>7.96</b>	8.50	<b>9.44</b>	<b>7.39</b>	<b>8.17</b>	7.55	6.66	7.01	7.13	7.64	7.14

## 2052 L MORE DETAILS ABOUT ERROR ANALYSIS 2053

2054 **Definition of Different Error Type.** We classify the error into 7 types: Consistency error, Hallu-  
2055 cination error, Reasoning error, Refusing error, Priority error, Mathematical error and Obsolescence  
2056 error. The detailed definitions of each error type are as follows:

- 2058 • **Consistency error:** The content in the model’s response is inherently flawed or internally  
2059 inconsistent, such as when the intermediate reasoning steps contradict the model’s final  
2060 answer.
- 2061 • **Hallucination error:** The large language model’s responses sometimes introduce fabri-  
2062 cated legal information or include statements that sound plausible but are factually incor-  
2063 rect particularly in Tasks 14, which require familiarity with typical legal cases.
- 2064 • **Reasoning error:** This type refers to flaws in the logical process used by the model to arrive  
2065 at its answer. These errors may include invalid deductions, misinterpretation of conditions,  
2066 or incorrect application of domain-specific rules. In many cases, the model’s intermediate  
2067 reasoning steps fail to logically support its final conclusion, even if the answer appears  
2068 superficially plausible. Such issues are particularly critical in the second-level tasks of  
2069 IPBench, which demand accurate multi-step and conditional reasoning within legal and  
2070 technical contexts.
- 2071 • **Refusing error:** This error typically occurs in Tasks 14, which require the model to recall  
2072 specific factual or legal cases. In these instances, some models respond by asking the  
2073 user for additional information or by explicitly refusing to provide an answer. While such  
2074 refusals may be more cautious or aligned with reliability principles, they still indicate a  
2075 limitation in the model’s ability to engage with the task as expected.
- 2076 • **Priority error:** Priority Error refers to the model’s failure to identify and prioritize the most  
2077 critical factor(s) when multiple elements jointly influence the outcome. Instead of focusing  
2078 on the decisive issue, the model may weigh secondary or irrelevant aspects equally, leading  
2079 to incorrect or misleading conclusions.
- 2080 • **Mathematical error:** This error type refers to issues related to a lack of precision in com-  
2081 plex calculations, often resulting in incorrect outcomes. These errors can arise from miscal-  
2082 culations, rounding mistakes, or failure to properly apply mathematical operations, leading  
2083 to significant discrepancies in the final result. This is particularly evident in Tasks 23,  
2084 Compensation Calculation, where both IP law knowledge and an understanding of the case  
2085 background are necessary to perform accurate calculations.
- 2086 • **Obsolescence error:** Obsolescence Error refers to the model’s failure to account for differ-  
2087 ences between current and outdated versions of legal documents or frameworks. This error  
2088 occurs when the generated answer overlooks changes in the law, leading to outdated or  
2089 inaccurate information. This is especially relevant in Tasks 13, Legal Evolution, where the  
2090 model must retain knowledge of both current and past laws and understand the differences  
2091 between them. However, some models do not update their memory, resulting in the use of  
2092 obsolete information.

2093 The most common error type is reasoning error, accounting for 33%. This is consistent with the  
2094 performance decrease observed in models using the Chain-of-Thought setting. This highlights the  
2095 importance of developing an IP-oriented model that balances both System 1 and System 2 capabili-  
2096 ties.

2097 **Case Study for Each Error Type.** We provide two examples, one in Chinese and one in English,  
2098 for each error type, as shown from Figure 6 to Figure 12. More extensive case studies for each task  
2099 can be found in Appendix N.

## 2100 M DATA EXAMPLES 2101

2103 We provide extensive data examples for each task in this section, as shown from Figure 13 to Fig-  
2104 ure 33. These examples include both English and Chinese datapoints, serving as representative  
2105 samples for each corresponding task and helping to better illustrate the task definitions.

2106	Chinese	Consistency Error	English
2107			
2108			
2109			
2110	问题: 某发明将蓝牙模块与机械锁体集成实现APP控制开锁, 现有技术中D1为带GPS的电子锁, D2公开蓝牙控制门锁。审查员应如何判断其创造性?	Question: The invention integrates a Bluetooth module with a mechanical lock body to enable app-controlled unlocking. In the prior art, D1 is an electronic lock with GPS, and D2 discloses Bluetooth-controlled door locks. How should the examiner assess its inventiveness?	Question: A lab-grown seafood patent expires in 3 years, but Congress proposes banning "cell-cultured" labeling. What drives its value erosion?
2111	A.具备创造性, 因首次用于共享单车 A. Inventive, as it is the first application in shared bicycles. B.不具备, 蓝牙应用属公知常识 B. Not inventive, as Bluetooth application is common knowledge. C.具备, 解决车辆防盗难题 C. Inventive, as it solves vehicle anti-theft challenges. D.不具备, D2给出明确技术启示 D. Not inventive, as D2 provides clear technical inspiration.	A. Loss of consumer branding clarity. B. Accelerated obsolescence from regulatory bans. C. Shift to B2B industrial ingredient patents. D. Protected by USDA Bioengineered labels.	
2112			
2113			
2114			
2115			
2116			
2117			
2118			
2119			
2120			
2121	gpt-4o-mini: ✗ D.不具备, D2给出明确技术启示 D. Not inventive, as D2 provides clear technical inspiration.	gpt-4o-mini: ✗ A. Loss of consumer branding clarity.	
2122			
2123			
2124	Error Category: <b>Consistency Error</b> Error Reason: Option A mentions "the first application in shared bicycles," but the question does not provide sufficient information to support this claim, making it ambiguous. Option B argues against inventiveness on the grounds that <b>Bluetooth applications are common general knowledge</b> . Since D2 already discloses Bluetooth-controlled door locks, this rationale is valid. Option C claims inventiveness by addressing vehicle anti-theft challenges, but it fails to directly compare with the prior art or provide sufficient evidence. Option D asserts a lack of inventiveness because D2 provides clear technical teachings, which aligns with Option B's reasoning.	Error Category: <b>Consistency Error</b> Error Reason: If Congress bans "cell-cultured" labeling, it could confuse consumers about what they are buying. This could lead to a decrease in consumer trust and interest in lab-grown seafood, <b>which would erode its market value</b> . (Option C): This option implies that the market is moving towards business-to-business applications rather than consumer products. However, it does not directly address the <b>immediate impact of the patent expiration and the labeling ban on consumer products</b> . (Option D): However, if the labeling is banned, this protection may not be sufficient to maintain consumer interest or market value.	
2125			
2126			
2127			
2128			
2129			
2130			
2131			
2132			
2133			
2134			
2135	Chinese	Reasoning Error	English
2136			
2137			
2138	问题: 采用表面码与拓扑编码结合, 逻辑错误率降至 $10^{-6}$ 。D1公开表面码方案, D2记载拓扑量子计算。该方案创造性如何评价?	Question: By combining surface codes with topological encoding, the logical error rate is reduced to $10^{-6}$ . D1 discloses a surface code scheme, and D2 documents topological quantum computing. How should the inventiveness of this solution be evaluated?	Question: In the case of "Mattel Inc. v. MGA Entertainment Inc.", What was one of the key legal issues addressed in the appeal by Mattel Inc.?
2139	A.不具备, 编码组合属理论研究 A. Not inventive, as the code combination is theoretical research. B.具备, 实现实用化突破 B. Inventive, as it achieves practical breakthroughs. C.不具备, D2提示编码改进方向 C. Not inventive, as D2 suggests improvement directions for encoding. D.具备, 解决量子退相干问题 D. Inventive, as it addresses quantum decoherence.	A. Whether MGA's Bratz dolls were substantially similar to Mattel's Barbie dolls. B. Whether the jury's verdict was supported by sufficient evidence. C. Whether MGA's advertising was false or misleading. D. Whether Mattel had a valid copyright claim over the Bratz dolls' design.	
2140			
2141			
2142			
2143			
2144			
2145			
2146			
2147			
2148			
2149	gpt-4o-mini: ✗ B.具备, 实现实用化突破 B. Inventive, as it achieves practical breakthroughs.	gpt-4o-mini: ✗ A. Whether MGA's Bratz dolls were substantially similar to Mattel's Barbie dolls.	
2150			
2151	Error Category: <b>Reasoning Error</b> Error Reason: The combination in this scheme is not only theoretically significant but also <b>holds the potential to achieve practical breakthroughs in quantum computing</b> . Both Options A and C argue that the scheme lacks inventiveness, which is clearly illogical. While Option D mentions addressing quantum decoherence, it fails to explicitly highlight the practical breakthrough. Option B, however, directly emphasizes the scheme's potential to <b>realize practical implementation breakthroughs</b> . Therefore, after comprehensive consideration, the most appropriate answer is B.	Error Category: <b>Reasoning Error</b> Error Reason: A. This option directly addresses the core of the dispute, as it involves the comparison of the two products and the potential for infringement. B. While this is a relevant legal issue in appeals, it is more procedural. C. This is a separate issue related to advertising practices and not the primary focus of the appeal regarding the dolls themselves. D. This option is also relevant, but it is more specific to copyright law rather than the broader issue of similarity and infringement. Given this analysis, the most pertinent legal issue that was addressed in the appeal by Mattel Inc. is option A.	
2152			
2153			
2154			
2155			
2156			
2157			
2158			
2159			

Figure 6: Consistency error case study.

2135	Chinese	Reasoning Error	English
2136			
2137			
2138	问题: 采用表面码与拓扑编码结合, 逻辑错误率降至 $10^{-6}$ 。D1公开表面码方案, D2记载拓扑量子计算。该方案创造性如何评价?	Question: By combining surface codes with topological encoding, the logical error rate is reduced to $10^{-6}$ . D1 discloses a surface code scheme, and D2 documents topological quantum computing. How should the inventiveness of this solution be evaluated?	Question: In the case of "Mattel Inc. v. MGA Entertainment Inc.", What was one of the key legal issues addressed in the appeal by Mattel Inc.?
2139	A.不具备, 编码组合属理论研究 A. Not inventive, as the code combination is theoretical research. B.具备, 实现实用化突破 B. Inventive, as it achieves practical breakthroughs. C.不具备, D2提示编码改进方向 C. Not inventive, as D2 suggests improvement directions for encoding. D.具备, 解决量子退相干问题 D. Inventive, as it addresses quantum decoherence.	A. Whether MGA's Bratz dolls were substantially similar to Mattel's Barbie dolls. B. Whether the jury's verdict was supported by sufficient evidence. C. Whether MGA's advertising was false or misleading. D. Whether Mattel had a valid copyright claim over the Bratz dolls' design.	
2140			
2141			
2142			
2143			
2144			
2145			
2146			
2147			
2148			
2149	gpt-4o-mini: ✗ B.具备, 实现实用化突破 B. Inventive, as it achieves practical breakthroughs.	gpt-4o-mini: ✗ A. Whether MGA's Bratz dolls were substantially similar to Mattel's Barbie dolls.	
2150			
2151	Error Category: <b>Reasoning Error</b> Error Reason: The combination in this scheme is not only theoretically significant but also <b>holds the potential to achieve practical breakthroughs in quantum computing</b> . Both Options A and C argue that the scheme lacks inventiveness, which is clearly illogical. While Option D mentions addressing quantum decoherence, it fails to explicitly highlight the practical breakthrough. Option B, however, directly emphasizes the scheme's potential to <b>realize practical implementation breakthroughs</b> . Therefore, after comprehensive consideration, the most appropriate answer is B.	Error Category: <b>Reasoning Error</b> Error Reason: A. This option directly addresses the core of the dispute, as it involves the comparison of the two products and the potential for infringement. B. While this is a relevant legal issue in appeals, it is more procedural. C. This is a separate issue related to advertising practices and not the primary focus of the appeal regarding the dolls themselves. D. This option is also relevant, but it is more specific to copyright law rather than the broader issue of similarity and infringement. Given this analysis, the most pertinent legal issue that was addressed in the appeal by Mattel Inc. is option A.	
2152			
2153			
2154			
2155			
2156			
2157			
2158			
2159			

Figure 7: Reasoning error case study.

	Chinese	Refusing Error	English
2160			
2161			
2162	 Chinese	Refusing Error	English
2163			
2164			
2165	问题：“空调专用微处理器控制芯片”开发合同案中，最高人民法院认为泰某微电子（上海）公司是否构成违约？	Question: In the case of the "Air Conditioner-Specific Microprocessor Control Chip" development contract, did the Supreme People's Court determine that TaiM Microelectronics (Shanghai) Co., Ltd. breached the contract?	Question: What limits claims covering "all sensory feedback systems integrating thalamocortical oscillations" after the inventor discloses only rodent models in the specification?
2166	A.无法确定 A. Cannot be determined. B.部分构成违约 B. Partially breached the contract. C.不构成违约 C. Did not breach the contract. D.构成违约 D. Breached the contract.	A. § 112 enablement for cross-species applicability. B. Written description requirement under Amgen for human neurophysiology. C. § 101 abstract idea exclusion. D.Obviousness-type double patenting.	
2167			
2168			
2169			
2170			
2171			
2172			
2173			
2174			
2175	gpt-4o-mini:  A. Cannot be determined.	gpt-4o-mini:  A. § 112 enablement for cross-species applicability.	
2176			
2177			
2178			
2179			
2180			
2181			
2182			
2183			
2184			
2185			
2186			
2187			
2188			
2189	 Chinese	Obsolescence Error	English
2190			
2191			
2192	问题：相较于2010年《中华人民共和国著作权法》，2020年修订版本中新增了哪项关于侵权举证的规定？	Question: Compared to the 2010 "Copyright Law of the People's Republic of China," which new provision on evidentiary rules for infringement was added in the 2020 revised version?	Question: How did 35 U.S.C. § 131 change examiner interview rules in December 2022 compared to the May 2015 version?
2193	A.网络平台承担过错推定责任 A. Online platforms bear presumed-fault liability. B.技术措施合法性举证责任倒置 B. Inversion of the burden of proof for the legality of technical measures. C.侵权复制品数量推定规则 C. Presumption rule for the quantity of infringing reproductions. D.法院可责令侵权人提供账簿资料 D. Courts may order infringers to provide accounting records.	A. Limited interviews to once per application. B. Required public disclosure of interview summaries. C. Mandated examiner approval for all amendments. D. Allowed virtual interviews by default.	
2194			
2195			
2196			
2197			
2198			
2199			
2200			
2201			
2202			
2203	gpt-4o-mini:  B.技术措施合法性举证责任倒置 B. Inversion of the burden of proof for the legality of technical measures.	gpt-4o-mini:  D. Allowed virtual interviews by default.	
2204			
2205			
2206			
2207			
2208			
2209			
2210			
2211			
2212			
2213			

Figure 8: Refusing error case study.

	Chinese	Obsolescence Error	English
2192	 Chinese	Obsolescence Error	English
2193			
2194			
2195			
2196			
2197			
2198			
2199			
2200			
2201			
2202			
2203	gpt-4o-mini:  B.技术措施合法性举证责任倒置 B. Inversion of the burden of proof for the legality of technical measures.	gpt-4o-mini:  D. Allowed virtual interviews by default.	
2204			
2205			
2206			
2207			
2208			
2209			
2210			
2211			
2212			
2213			

Figure 9: Obsolescence error case study.

2214

2215

2216

2217

2218

问题: 针对发明专利CN202588024680H记载的“再生冷却通道设计”, 其权利要求明确微通道水力直径 $1\pm0.1\text{mm}$ 、冷却剂流速 $8\pm0.5\text{m/s}$ 、热流密度 $\geq5\text{MW/m}^2$ , 且压降损失 $<10\%$ 。有效规避路径需包含?

**Question:** Compared to the 2010 "Copyright Law of the People's Republic of China," which new provision on evidentiary rules for infringement was added in the 2020 revised version?

A. 开发发汗冷却与冲击冷却复合系统  
 B. 优化通道形状将水力直径减至 $0.8\text{mm}$   
 C. Inversion of the burden of proof for the legality of technical measures.  
 C. Presumption rule for the quantity of infringing reproductions.  
 D. 增加自适应流量控制阀调节压降  
 D. Courts may order infringers to provide accounting records.

2227

2228

2229

2230

2231

2232

2233

2234

2235

2236

2237

2238

2239

2240

2241

2242

2243

2244

2245

2246

2247

2248

2249

2250

2251

2252

2253

2254

2255

2256

2257

2258

2259

2260

2261

2262

2263

2264

2265

2266

2267

	Chinese	Priority Error	English
2218			
2219	问题: 针对发明专利CN202588024680H记载的“再生冷却通道设计”, 其权利要求明确微通道水力直径 $1\pm0.1\text{mm}$ 、冷却剂流速 $8\pm0.5\text{m/s}$ 、热流密度 $\geq5\text{MW/m}^2$ , 且压降损失 $<10\%$ 。有效规避路径需包含?	<b>Question:</b> Compared to the 2010 "Copyright Law of the People's Republic of China," which new provision on evidentiary rules for infringement was added in the 2020 revised version?	<b>Question:</b> A space solar patent (US 36,789,012AA) claims "ultralight perovskite solar films $\leq0.5\text{ g/W}$ with $>28\%$ efficiency after $10^6$ proton irradiation cycles at 1 MeV equivalent fluence" which modification constitutes non-infringement?
2220			
2221			
2222			
2223			
2224			
2225			
2226			
2227			
2228			
2229			
2230			
2231			
2232			
2233			
2234			
2235			
2236			
2237			
2238			
2239			
2240			
2241			
2242			
2243			
2244			
2245			
2246			
2247			
2248			
2249			
2250			
2251			
2252			
2253			
2254			
2255			
2256			
2257			
2258			
2259			
2260			
2261			
2262			
2263			
2264			
2265			
2266			
2267			

## Chinese

## Priority Error

## English

	Chinese	Hallucination Error	English
2243			
2244			
2245			
2246	问题: 甲游戏公司NPC对话树具独创分支逻辑, 乙公司AI聊天机器人模仿该逻辑架构。根据《著作权法实施条例》第13条, 哪一项是对的?	<b>Question:</b> Game Company A's NPC dialogue tree features a unique branching logic. Company B's AI chatbot imitates this logical structure. According to Article 13 of the Implementing Regulations of the Copyright Law, which statement is correct?	<b>Question:</b> In the case of "Funk Brothers Seed Co. v. Kalo Inoculant Co.", Which of the following is a potential remedy for trademark infringement?
2247			
2248			
2249			
2250			
2251			
2252			
2253			
2254			
2255			
2256			
2257			
2258			
2259			
2260			
2261			
2262			
2263			
2264			
2265			
2266			
2267			

## Chinese

## Priority Error

## English

	Chinese	Hallucination Error	English
2243			
2244			
2245			
2246	问题: 甲游戏公司NPC对话树具独创分支逻辑, 乙公司AI聊天机器人模仿该逻辑架构。根据《著作权法实施条例》第13条, 哪一项是对的?	<b>Question:</b> Game Company A's NPC dialogue tree features a unique branching logic. Company B's AI chatbot imitates this logical structure. According to Article 13 of the Implementing Regulations of the Copyright Law, which statement is correct?	<b>Question:</b> In the case of "Funk Brothers Seed Co. v. Kalo Inoculant Co.", Which of the following is a potential remedy for trademark infringement?
2247			
2248			
2249			
2250			
2251			
2252			
2253			
2254			
2255			
2256			
2257			
2258			
2259			
2260			
2261			
2262			
2263			
2264			
2265			
2266			
2267			

Figure 10: Priority error case study.

2268	Chinese	Mathematical Error	English
2269			
2270			
2271	问题: 甲公司拥有一注册商标, 核定使用在服装商品上。乙公司未经许可, 在其生产的服装上使用了与甲公司注册商标近似的标识, 导致甲公司销量下降。甲公司因乙公司侵权行为少销售服装1000件, 每件服装的利润为50元。乙公司共销售侵权服装2000件, 每件服装的利润为30元。根据《商标法》, 甲公司应获得的赔偿数额为多少?	Question: Company A is the copyright owner of a professional drawing software, and Company B copies and sells the software without permission. As a result, the court found that Company A had lost 100,000 copies sold at a price of \$199 per copy, resulting in an actual loss of \$19.9 million. At the same time, Company A spent 120,000 yuan in notary fees for obtaining evidence. How much should Company B compensate?	
2272	Question: Due to B's infringement, Company A sold 1,000 fewer clothing items, with a profit of \$50 per item. Company B sold 2,000 infringing clothing items, earning a profit of \$30 per item. According to China's Trademark Law, what amount should Company A be awarded in damages?	A. \$19.9 million. B. \$20.02 million. C. \$20.12 million. D. \$30 million.	
2273	A. 50,000元 B. ¥50,000. C. 60,000元 D. ¥60,000. E. 100,000元 F. ¥100,000. G. 200,000元 H. ¥200,000.		
2274	gpt-4o-mini: ×	A. \$19.9 million.	
2275	B. 60,000元 C. ¥60,000.		
2276			
2277			
2278			
2279			
2280			
2281			
2282			
2283			
2284			
2285	<b>Error Category: Mathematical Error</b> <b>Error Reason:</b> Due to Company B's infringement, Company A sold 1,000 fewer clothing items, with a profit of **\$50 per item**. Therefore, Company A's actual losses from the infringement are $1,000 \text{ items} \times \$50/\text{item} = \$50,000$ . Company B sold 2,000 infringing clothing items, with a profit of **\$30 per item**. Thus, Company B's illicit profits from the infringement are $2,000 \text{ items} \times \$30/\text{item} = \$60,000$ . Under China's Trademark Law (Article 63), the damages awarded to the rights holder (Company A) are generally the higher of: The rights holder's actual losses (\$50,000). The infringer's illicit profits (\$60,000). The compensation amount is **\$60,000**.	<b>Error Category: Mathematical Error</b> <b>Error Reason:</b> Company A lost 100,000 copies sold at a price of \$199 per copy. Actual loss = $100,000 \text{ copies} \times \$199/\text{copy} = \$19,900,000$ (or \$19.9 million). Company A incurred 120,000 yuan in notary fees. To include this in the compensation, we need to convert yuan to dollars. However, the problem does not provide a conversion rate. For the sake of this calculation, let's assume that 120,000 yuan is approximately \$17,000 (this is a rough estimate, as the actual conversion rate can vary). Total compensation = Actual loss + Additional costs. Total compensation = $\$19.9 \text{ million} + \$17,000 = \$19,917,000$ .	
2286			
2287			
2288			
2289			
2290			
2291			
2292			
2293			
2294			
2295			
2296			
2297			
2298			
2299			
2300			
2301			
2302			
2303			
2304			
2305			
2306			
2307			
2308			
2309			
2310			
2311			
2312			
2313			
2314			
2315			
2316			
2317			
2318	<b>问题:</b> <b>Question:</b> 根据《中华人民共和国专利法》, 下列哪项属于"实用新型"的定义?	<b>Question:</b> What is "dilution" under the Trademark Dilution Revision Act (TDRA)?	
2319	According to the Patent Law of the People's Republic of China, which of the following items falls under the definition of a "utility model"?		
2320	<b>选项:</b> <b>Options:</b> A.对产品的形状、图案或其结合提出的新设计 A. A new design concerning the shape, pattern, or combination thereof of a product B.对产品的形状、构造或其结合提出的实用新技术方案 B. A practical new technical solution relating to the shape, structure, or combination thereof of a product C.对方法或其改进提出的新技术方案 C. A new technical solution proposed for a method or its improvement D.对科学发现提出的技术应用 D. A technical application based on a scientific discovery	<b>Options:</b> A. Weakening a famous mark's uniqueness through unauthorized use B. Creating consumer confusion about a mark's source C. Using a descriptive term without secondary meaning D. Failing to renew a registration on time	
2321	<b>答案: B</b> <b>Answer: B</b>	<b>Answer: A</b>	
2318	<b>知识产权类型: 专利</b> <b>Intellectual Property Type: Patent</b>	<b>Intellectual Property Type:</b> <b>Trademark</b>	
2319			
2320			
2321			

Figure 12: Mathematical error case study.

Task 1-1: Legal Concept Memory	
<p><b>问题:</b> <b>Question:</b> 根据《中华人民共和国专利法》, 下列哪项属于"实用新型"的定义?</p> <p>According to the Patent Law of the People's Republic of China, which of the following items falls under the definition of a "utility model"?</p> <p><b>选项:</b> <b>Options:</b></p> <p>A.对产品的形状、图案或其结合提出的新设计 A. A new design concerning the shape, pattern, or combination thereof of a product B.对产品的形状、构造或其结合提出的实用新技术方案 B. A practical new technical solution relating to the shape, structure, or combination thereof of a product C.对方法或其改进提出的新技术方案 C. A new technical solution proposed for a method or its improvement D.对科学发现提出的技术应用 D. A technical application based on a scientific discovery</p> <p><b>答案: B</b> <b>Answer: B</b></p> <p><b>知识产权类型: 专利</b> <b>Intellectual Property Type: Patent</b></p>	<p><b>Question:</b> What is "dilution" under the Trademark Dilution Revision Act (TDRA)?</p> <p><b>Options:</b></p> <p>A. Weakening a famous mark's uniqueness through unauthorized use B. Creating consumer confusion about a mark's source C. Using a descriptive term without secondary meaning D. Failing to renew a registration on time</p> <p><b>Answer: A</b></p> <p><b>Intellectual Property Type:</b> <b>Trademark</b></p>

Figure 13: Data example of task 1-1.

2322  
2323  
2324  
2325  
2326  
2327  
2328  
2329  
2330  
2331  
2332  
2333  
2334  
2335  
2336  
2337  
2338  
2339  
2340  
2341  
2342  
2343  
2344

Task 1-2: Legal Clause Memory	
<p><b>问题:</b>  <b>Question:</b>          根据《中华人民共和国商标法(2019年修正)》第10条,下列哪种标志可以作为商标使用?          According to Article 10 of the Trademark Law of the People's Republic of China (2019 Amendment), which of the following signs may be used as a trademark?</p> <p><b>选项:</b>  <b>Options:</b>          A.与外国军旗近似的图形          A.Signs that are similar to the military flags of foreign countries          B.带有民族歧视性的文字          B. Text or symbols with ethnic discrimination          C.县级以上行政区划地名但具有其他含义的          C. Names of administrative divisions at or above the county level that have other meanings          D.与"红十字"名称相同的标志          D. Signs identical to the name 'Red Cross'  <b>答案: C</b>  <b>Answer: C</b> </p>	<p><b>Question:</b>          Under 17 U.S.C. § 110(1), what is allowed in classroom teaching?</p>  <p><b>Options:</b>          A. Public performance of a full movie          B. Display of copyrighted materials during face-to-face instruction          C. Unlicensed distribution of textbooks          D. Broadcasting a play to the public  <b>Answer: B</b></p>
<p><b>知识产权类型:</b> 商标  <b>Intellectual Property Type:</b> Trademark</p>	<p><b>Intellectual Property Type:</b>          Copyright</p>

Figure 14: Data example of task 1-2.

2345  
2346  
2347  
2348  
2349  
2350  
2351  
2352  
2353  
2354  
2355  
2356  
2357  
2358  
2359  
2360  
2361  
2362  
2363  
2364  
2365  
2366  
2367  
2368  
2369  
2370  
2371  
2372  
2373  
2374  
2375

Task 1-3: Legal Evolution	
<p><b>问题:</b>  <b>Question:</b>          相较于2008年《中华人民共和国专利法》, 2020年修订版本新增了哪项关于药品专利期限补偿的制度?          Compared to the 2008 Patent Law of the People's Republic of China, what new pharmaceutical patent term extension system was introduced in the 2020 amended version?</p> <p><b>选项:</b>  <b>Options:</b>          A.允许专利期限延长至20年以上          A. Permits the extension of patent terms beyond 20 years          B.针对仿制药上市给予专利豁免期          B. Provides a patent exemption period for generic drug marketing approval          C.要求药品专利必须进行期限登记          C. Requires mandatory term registration for pharmaceutical patents          D.对创新药专利给予最长5年的期限补偿          D. Grants a maximum 5-year term extension for innovative drug patents  <b>答案: D</b>  <b>Answer: D</b> </p>	<p><b>Question:</b>          Compared to 2015, what adjustment was made to patent term extension under 35 U.S.C. § 154(b)(1)(B) in 2022?</p>  <p><b>Options:</b>          A. Added compensation for PTAB appeal delays          B. Excluded delays caused by PTAB reviews          C. Limited extensions to biotechnology patents          D. Reduced the maximum extension to 3 years  <b>Answer: A</b></p>
<p><b>知识产权类型:</b> 专利  <b>Intellectual Property Type:</b> Patent</p>	<p><b>Intellectual Property Type:</b>          Patent</p>

Figure 15: Data example of task 1-3.

2376  
2377  
2378  
2379  
2380  
2381  
2382  
2383  
2384  
2385  
2386  
2387  
2388  
2389  
2390  
2391  
2392  
2393  
2394  
2395  
2396  
2397  
2398  
2399  
2400  
2401  
2402  
2403  
2404  
2405  
2406  
2407  
2408  
2409  
2410  
2411  
2412  
2413  
2414  
2415  
2416  
2417  
2418  
2419  
2420  
2421  
2422  
2423  
2424  
2425  
2426  
2427  
2428  
2429

Task 1-4: Typical Case Memory	
<p><b>问题:</b>  <b>Question:</b>          “郑州某研究所与陈某侵害植物新品种权纠纷案”中，陈被认定侵权的行为是什么？          In the case of Zhengzhou Research Institute v. Chen regarding infringement of new plant variety rights dispute, What are the acts of infringement that Chen was found to have committed?</p> <p><b>选项:</b>  <b>Options:</b>          A. 未经许可繁育“天使红”石榴新品种的繁殖材料          A. Propagation of reproductive materials of the new pomegranate variety 'Angel Red' without authorization          B. 销售假冒“天使红”石榴的果实          B. Sale of counterfeit fruits of the 'Angel Red' pomegranate variety          C. 未经授权使用“天使红”商标          C. Unauthorized use of the 'Angel Red' trademark          D. 未支付品种许可费用          D. Non-payment of variety licensing fees  <b>答案: A</b>  <b>Answer: A</b> </p>	<p><b>Question:</b>          In the case of "Kewanee Oil Co. v. Bicron Corp.", what was the Supreme Court's decision regarding the preemption of state trade secret law by federal patent law?</p>  <p><b>Options:</b>          A. The Court did not address the issue of preemption          B. The Court ruled that preemption applied only in specific cases          C. State trade secret law was preempted by federal patent law          D. State trade secret law was not preempted by federal patent law  <b>Answer: D</b></p>
<p><b>知识产权类型:</b> 植物新品种  <b>Intellectual Property Type:</b> New Plant Variety</p>	<p><b>Intellectual Property Type:</b>          Trade Secret</p>

Figure 16: Data example of task 1-4.

Task 1-5-1: Patent IPC Classification	
<p><b>问题:</b>  <b>Question:</b>  <b>标题:</b> 基于物联网的数字孪生城市交通灯控制系统  <b>Title:</b> IoT-based Digital Twin Urban Traffic Light Control System  <b>摘要:</b> 本发明公开了基于物联网的数字孪生城市交通灯控制系统，涉及智能控制的技术领域，包括采集模块、分析模块和控制模块，计算拥挤度变化率，得到集中时间段，计算历史自然数据的权重，构造干扰函数，计算干扰时长和新的集中时间段，建立交通模型，将新的拥挤度变化率与拥挤度变化量阈值进行比较，执行第一操作，并建立第一映射关系。本发明通过实时监测和分析交通流量、拥挤度变化情况，以及考虑自然环境因素的影响，优化交通流的通行时间分配，动态调整交通信号灯的控制策略，通过历史交通数据和自然数据的分析，能够预测未来的交通情况，提前做好相应的交通信号调整和路线规划，物联网技术的应用使得交通灯控制系统能够实现远程监控和集中管理。  <b>Abstract:</b> The present invention discloses an IoT-based digital twin urban traffic signal control system... (Omit)  <b>答案: G08G1/08</b>  <b>Answer: G08G1/08</b> </p>	<p><b>Question:</b>  <b>Title:</b> System and method for migrating agents between mobile devices  <b>Abstract:</b> Mobile agents can be deployed to location aware mobile devices within specific regions of interest to achieve specific goals in respect of events occurring in the region of interest. In order to ensure that the agent can persist within the region of interest until the agent goals are achieved, the agent is configured to locate other devices within the region of interest and to propagate itself, by moving or copying itself, to those other devices. When a device hosting the agent exits the region of interest, the agent is terminated, thereby freeing device resources.  <b>Answer: A01D34/43</b></p>
<p><b>知识产权类型:</b> 专利  <b>Intellectual Property Type:</b> Patent</p>	<p><b>Intellectual Property Type:</b>          Patent</p>

Figure 17: Data example of task 1-5-1.

2430

2431

2432

2433

2434

2435

2436

2437

2438

2439

2440

2441

2442

2443

2444

2445

2446

2447

2448

2449

2450

2451

2452

2453

2454

2455

2456

2457

2458

2459

2460

2461

2462

2463

2464

2465

2466

2467

2468

2469

2470

2471

2472

2473

2474

2475

2476

2477

2478

2479

2480

2481

2482

2483

Task 1-5-2: Patent CPC Classification	
Title:	Multipurpose machine for cultivating trees
Abstract:	<p>A multipurpose machine for cultivating trees, comprising an inverted U-shape structure that enables the machine to pass over existing trees to carry out pruning, disinfection or fruit picking tasks, provided at the bottom with wheels, driven by at least one motor that autonomously facilitates the movement thereof, and respective upper frames that telescopically couple to each other, being driven by respective cylinders to move the portion of the structure on the right with respect to the one on the left in order to vary the width of the machine. Likewise, the machine has the ability to raise or lower the upper structure of the same to adapt it to the height of the trees to be cultivated.s.</p>
Answer:	A01D46/30
Intellectual Property Type:	Patent

Figure 18: Data example of task 1-5-2.

Task 1-6: IP Element Identification	
问题:	Question: Please select the correct answer from A, B, C, and D. A lab engineers synthetic DNA sequences encoding Shakespearean sonnets. Competitors replicate them. What governs?
选项:	<p>Question: Please select the correct answer from A, B, C, and D. A lab engineers synthetic DNA sequences encoding Shakespearean sonnets. Competitors replicate them. What governs?</p> <p>Options:</p> <p>A. 商标显著性获得时间 A. Time of obtaining distinctiveness of the trademark B. 声音商标的混淆可能性判定 B. Determination of Confusion Potential of Sound Trademarks C. 描述性使用正当性 C. Legitimacy of Descriptive Use D. 驰名商标跨类保护 D. Cross class protection of well-known trademarks</p> <p>答案: B Answer: B</p>
知识产权类型: 专利 Intellectual Property Type: Patent	Intellectual Property Type: Patent

Figure 19: Data example of task 1-6.

2484  
2485  
2486  
2487  
2488  
2489  
2490  
2491  
2492  
2493  
2494  
2495  
2496  
2497  
2498  
2499  
2500  
2501  
2502  
2503  
2504  
2505  
2506  
2507  
2508  
2509  
2510  
2511  
2512  
2513  
2514  
2515  
2516  
2517  
2518  
2519  
2520  
2521  
2522  
2523  
2524  
2525  
2526  
2527  
2528  
2529  
2530  
2531  
2532  
2533  
2534  
2535  
2536  
2537

Task 1-7: Process Guidance	
<p><b>问题:</b>  <b>Question:</b>  <b>商标注册申请人可以通过什么方式提交《中华人民共和国商标法实施条例》规定的申请?</b>  <b>How can trademark registration applicants submit applications as stipulated in the Implementing Regulations of the Trademark Law of the People's Republic of China?</b></p> <p><b>选项:</b>  <b>Options:</b>  <b>A. 只能通过纸质方式</b>  <b>A. Can only be done in paper form</b>  <b>B. 只能通过电子方式</b>  <b>B. Can only be done through electronic means</b>  <b>C. 可以通过纸质或电子方式</b>  <b>C. Can be done through paper or electronic means</b>  <b>D. 只能通过代理机构提交</b>  <b>D. Can only be submitted through an agency</b></p> <p><b>答案: C</b>  <b>Answer: C</b></p>	<p><b>Question:</b>  <b>Please select the correct answer from A, B, C, and D. Who makes the decision that the international application is considered withdrawn?</b></p>  <p><b>Options:</b>  <b>A. International Bureau</b>  <b>B. Receiving Office</b>  <b>C. International Searching Authority</b>  <b>D. Chinese</b>  <b>Answer: B</b></p>
<p><b>知识产权类型: 商标</b>  <b>Intellectual Property Type: Trademark</b></p>	<p><b>Intellectual Property Type:</b>  <b>Patent</b></p>

Figure 20: Data example of task 1-7.

Task 2-1: Patent Technology Forecasting	
<p><b>问题:</b>  <b>Question:</b>  <b>“横机设备”技术秘密许可案中，横机设备的创新方向之一是提升其适应性和多功能性。以下哪种创新方向最能满足不同用户的需求?</b>  <b>In the trade secret licensing case involving "flat knitting machines", one of the key innovation directions for such equipment is enhancing adaptability and multifunctionality. Among the following innovation approaches, which one would best meet the diverse needs of different users?</b></p> <p><b>选项:</b>  <b>Options:</b>  <b>A. 开发适用于极端环境的横机设备，拓展其在特殊场景中的应用</b>  <b>A. Develop flat knitting machines engineered for extreme environments to expand their applications in specialized scenarios.</b>  <b>B. 开发适用于不同织物类型的横机设备，结合智能适配技术实现快速部署</b>  <b>B. Develop flat knitting machines suitable for different fabric types, and combine intelligent adaptation technology to achieve rapid deployment</b></p> <p><b>答案: B</b>  <b>Answer: B</b></p>	<p><b>Question:</b>  <b>Please select the correct answer from A, B, C, and D. A synthetic biology patent (US20259246819) engineers nitrogen-fixing cereal crops via chloroplast-targeted nif gene clusters with 95% ammonia production efficiency under 21% O<sub>2</sub>. Which agricultural subsidy reform drives adoption?</b></p>  <p><b>Options:</b>  <b>A. USDA Precision Agronomy Grants</b>  <b>B. Farm Bill Conservation Reserve Expansion</b>  <b>C. EPA CAFO Emission Trading System</b>  <b>D. UNFCCC Climate-Smart Agriculture</b>  <b>Answer: A</b></p>
<p><b>知识产权类型: 专利</b>  <b>Intellectual Property Type: Patent</b></p>	<p><b>Intellectual Property Type:</b>  <b>Patent</b></p>

Figure 21: Data example of task 2-1.

2538  
2539  
2540  
2541  
2542  
2543  
2544  
2545  
2546  
2547  
2548  
2549  
2550  
2551  
2552  
2553  
2554  
2555  
2556  
2557  
2558  
2559  
2560  
2561  
2562  
2563  
2564  
2565  
2566  
2567  
2568  
2569  
2570  
2571  
2572  
2573  
2574  
2575  
2576  
2577  
2578  
2579  
2580  
2581  
2582  
2583  
2584  
2585  
2586  
2587  
2588  
2589  
2590  
2591

Task 2-2: Infringement Behavior Determination	
<p><b>问题:</b>  <b>Question:</b>  <b>甲画家作品被用于训练AI模型, 乙公司生成"梵高风格化"图片销售。根据《著作权法实施条例》第21条, 哪一项是对的?</b>  <b>Artist A's works are used to train AI models, while Company B generates "Van Gogh stylized" images for sale. According to Article 21 of Regulation for the Implementation of the Copyright Law, which one is correct?</b></p> <p><b>选项:</b>  <b>Options:</b></p> <p>A. 不侵权, 因风格不受保护  <b>A. Not infringing, style is not protected</b>  <b>B. 构成侵权, 若训练数据包含未授权作品</b>  <b>B. If the training data contains unauthorized works, it constitutes infringement</b>  <b>C. 需比对笔触相似度</b>  <b>C. The similarity of brushstrokes needs to be compared</b>  <b>D. 若声明"AI辅助创作"则不侵权</b>  <b>D. If it is declared as "AI-assisted creation", there is no infringement</b></p> <p><b>答案: B</b>  <b>Answer: B</b></p>	<p><b>Question:</b>  <b>A company uses a logo that has a similar color scheme and general shape to another company's logo but different lettering. Is this trademark infringement?</b></p>  <p><b>Options:</b></p> <p><b>A. Yes, if the overall similarity can lead to consumer confusion</b>  <b>B. No, because of the different lettering</b>  <b>C. Only if the other company's logo is very new</b>  <b>D. Only if the company uses the logo on a large number of products</b></p> <p><b>Answer: A</b></p>
<p><b>知识产权类型: 著作权</b>  <b>Intellectual Property Type: Copyright</b></p>	<p><b>Intellectual Property Type:</b>  <b>Trademark</b></p>

Figure 22: Data example of task 2-2.

Task 2-3: Compensation Calculation	
<p><b>问题:</b>  <b>Question:</b>  <b>某游戏公司侵权使用他人美术设计, 法院认定: 权利人许可费为每幅设计5万元; 侵权使用设计20幅; 侵权游戏月收入100万元 (利润率60%)。赔偿金额应为?</b>  <b>A game company infringed upon others' art designs. The court ruled that the licensing fee for each design was 50,000 yuan. Twenty infringing designs were used; The monthly revenue of infringing games is 1 million yuan (with a profit margin of 60%). What should be the amount of compensation?</b></p> <p><b>选项:</b>  <b>Options:</b></p> <p><b>A. <math>20 \times 5\text{万} = 100\text{万}</math></b>  <b>A. <math>20 \times 50,000 = 1 \text{ million}</math></b>  <b>B. <math>100\text{万} \times 60\% \times 12\text{月} = 720\text{万}</math></b>  <b>B. <math>1 \text{ million} \times 60\% \times 12 \text{ months} = 7.2 \text{ million}</math></b>  <b>C. <math>\max(100\text{万}, 720\text{万}) = 720\text{万}</math></b>  <b>C. <math>\max(1 \text{ million}, 7.2 \text{ million}) = 7.2 \text{ million}</math></b>  <b>D. 500万元以下 (如500万)</b>  <b>D. Less than 5 million yuan (e.g. 5 million)</b></p> <p><b>答案: A</b>  <b>Answer: A</b></p>	<p><b>Question:</b>  <b>The defendant's misappropriation of trade secrets resulted in a loss of \$500,000 in the plaintiff's income and a profit of \$300,000 for the defendant. The court awarded damages on the basis of "unjust enrichment". What is the maximum number of damages a plaintiff can receive under Section 3(b)(1) of the Uniform Trade Secrets Act (UTSA)?</b></p>  <p><b>Options:</b></p> <p><b>A. 300,000</b>  <b>B. 500,000</b>  <b>C. 800,000</b>  <b>D. 1,500,000</b></p> <p><b>Answer: B</b></p>
<p><b>知识产权类型: 著作权</b>  <b>Intellectual Property Type: Copyright</b></p>	<p><b>Intellectual Property Type:</b>  <b>Trade secret</b></p>

Figure 23: Data example of task 2-3.

2592

2593

2594

2595

2596

2597

2598

2599

2600

2601

2602

2603

2604

2605

2606

2607

2608

2609

2610

2611

2612

2613

2614

2615

2616

2617

2618

2619

2620

2621

2622

2623

2624

2625

2626

2627

2628

2629

2630

2631

2632

2633

2634

2635

2636

2637

2638

2639

2640

2641

2642

2643

2644

2645

Task 2-4: Patent Valuation	
<p><b>问题:</b>  <b>Question:</b>          某抗癌药化合物专利剩余保护期5年，年销售额达12亿元。若该药被纳入国家医保集采目录导致单价下降60%，其专利价值趋势是？          The remaining patent protection period of a certain anti-cancer drug compound is five years, with an annual sales volume reaching 1.2 billion yuan. If the drug is included in the national medical insurance centralized procurement list, resulting in a 60% drop-in unit price, what is the trend of its patent value?</p> <p><b>选项:</b>  <b>Options:</b>          A. 因销量暴增翻倍          A. Because the sales volume has doubled sharply          B. 因利润压缩大幅贬值          B. It depreciated significantly due to the compression of profits          C. 转为技术秘密后增值          C. It increases in value after being converted into a technical secret          D. 因政策保护维持不变          D. It remains unchanged due to policy protection  <b>答案: B</b>  <b>Answer: B</b> </p>	<p><b>Question:</b>          Tesla's electric vehicle charging patents (e.g., U.S. 20170171460) were licensed to Rivian under a fixed-fee model. Why might this structure reduce valuation risk for Tesla?</p>  <p><b>Options:</b>          A. Fixed fees guarantee upfront cash flow          B. Avoids dependency on fluctuating royalty rates          C. Eliminates litigation risks          D. Both A and B  <b>Answer: D</b></p>
<p><b>知识产权类型: 专利</b>  <b>Intellectual Property Type: Patent</b></p>	<p>Intellectual Property Type:          Patent</p>

Figure 24: Data example of task 2-4.

Task 2-5: Trade Secret Requirements	
<p><b>问题:</b>  <b>Question:</b>          根据《反不正当竞争法》，下列哪项不属于商业秘密的构成要件？          According to the Anti-Unfair Competition Law, which of the following is not a constituent element of trade secrets?</p> <p><b>选项:</b>  <b>Options:</b>          A. 不为公众所知悉          A. Unknown to the public          B. 具有商业价值          B. Have commercial value          C. 权利人已采取合理保密措施          C. The right holder has taken reasonable confidentiality measures          D. 已向行政机关登记备案          D. It has been registered and filed with the administrative authority  <b>答案: D</b>  <b>Answer: D</b> </p>	

Figure 25: Data example of task 2-5.

2646  
2647  
2648  
2649  
2650  
2651  
2652  
2653  
2654  
2655  
2656  
2657  
2658  
2659  
2660  
2661  
2662  
2663  
2664  
2665  
2666  
2667  
2668  
2669  
2670  
2671  
2672  
2673  
2674  
2675  
2676  
2677  
2678  
2679  
2680  
2681  
2682  
2683  
2684  
2685  
2686  
2687  
2688  
2689  
2690  
2691  
2692  
2693  
2694  
2695  
2696  
2697  
2698  
2699

Task 3-1: Patent Document Proofreading	
<p><b>问题:</b>  <b>Question:</b>  <b>优先权文本记载:</b> “支架的轴向弹性模量为10-20GPa”  <b>中国申请修改为..... (后略)</b>  The priority text states that "the axial elastic modulus of the bracket is 10-20GPa", and the Chinese application is modified to ... (Omit)</p> <p><b>选项:</b>  <b>Options:</b>  A.轴向模量与径向刚度的物理量纲不同  A. The physical dimensions of axial modulus and radial stiffness are different.  B.优先权文件未提供径向刚度测试数据  B. The priority document does not provide radial stiffness test data.  C.15GPa阈值与原始参数范围无直接对应关系  C. There is no direct correlation between... (Omit)  D.未说明测试方法的边界条件  D. The boundary conditions for the test method are not specified.  <b>答案: C</b>  <b>Answer: C</b></p>	<p><b>Question:</b>  Priority doc shows "battery capacity retention <math>\geq 80\%</math> after 500 cycles at 1C." Continuation claims: "<math>\geq 90\%</math> retention after 300 cycles at 2C." Spec only tests 1C cycling. Core deception?</p>  <p><b>Options:</b>  A.C-rate doubling accelerates degradation nonlinearly.  B. Cycle count reduction invalidates long-term performance.  C. Capacity retention depends on voltage cutoff.  D.No data supports 2C rate performance.  <b>Answer: D</b></p>
<p><b>知识产权类型:</b> 专利  <b>Intellectual Property Type:</b> Patent</p>	<p><b>Intellectual Property Type:</b>  Patent</p>

Figure 26: Data example of task 3-1.

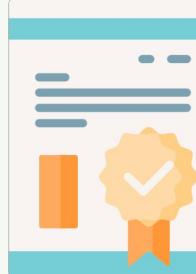
Task 3-2: Patent Validity Identification	
<p><b>问题:</b>  <b>Question:</b>  某骨科螺钉专利的合金比例, 与某冶炼厂15年前客户定制产品(未申请专利)的配方记录完全一致。该专利缺乏?  The alloy ratio of an orthopedic screw patent is exactly the same as the formula record of a customer's customized product (not patented) from a smelter 15 years ago. What is missing from this patent?</p> <p><b>选项:</b>  <b>Options:</b>  A.充分公开  A. Full disclosure  B.实用性  B. Practicability  C.创造性  C. Creativeness  D.新颖性  D. Novelty  <b>答案: D</b>  <b>Answer: D</b></p>	<p><b>Question:</b>  An AI startup patented a neural network architecture in 2022. Evidence shows the same architecture was fully implemented in an open-source GitHub repository with 500+ stars 14 months pre-filing. What invalidates the patent?</p>  <p><b>Options:</b>  A.Novelty  B. Non-obviousness  C. Utility  D. Enablement  <b>Answer: A</b></p>
<p><b>知识产权类型:</b> 专利  <b>Intellectual Property Type:</b> Patent</p>	<p><b>Intellectual Property Type:</b>  Patent</p>

Figure 27: Data example of task 3-2.

2700  
2701  
2702  
2703  
2704  
2705  
2706  
2707  
2708  
2709  
2710  
2711  
2712  
2713  
2714  
2715  
2716  
2717  
2718  
2719  
2720  
2721  
2722

Task 3-3: Patent Match	
<p><b>问题:</b>  <b>Question:</b>          请从A,B,C,D四个选项中选出与下述专利最相似的专利序号,该序号是?..... (后略)          Please select the patent number that is most similar to the following patent from the four options A, B, C and D.          What is the number? ... (Omit)</p> <p><b>选项:</b>  <b>Options:</b>          A.一种多功能洁面仪..... (后略)          A. A multi-functional facial cleanser... (Omit)          B.一种具有吸取毛发功能的宠物毛刷及其使用方法..... (后略)          B. A pet brush with hair absorption function and its use method... (Omit)          C.一种吸水刷头,包括..... (后略)          C. A water-absorbing brush head, including... (Omit)          D.本发明公开了一种多功能化妆刷..... (后略)          D. The invention discloses a multi-functional makeup brush... (Omit)  <b>答案: C</b>  <b>Answer: C</b> </p>	<p><b>Question:</b>          Provided is an adipose tissue preservation solution, comprising: polyethylene glycol 400, human albumin, norfloxacin, low molecular dextran, and Ac-DEVD-CHO... (Omit)</p> <p><b>Options:</b>          A. Provided in the present invention are a mesenchymal stem cell injection... (Omit)          B. The present invention provides a mesenchymal stem cell injection... (Omit)          C. Provided in the present invention is a cell freezing medium for clinical use... (Omit)          D. The present invention provides a mesenchymal stem cell injection... (Omit)  <b>Answer: D</b></p>
<b>知识产权类型: 专利</b> <b>Intellectual Property Type: Patent</b>	<b>Intellectual Property Type:</b> <b>Patent</b>

Figure 28: Data example of task 3-3.

2723  
2724  
2725  
2726  
2727  
2728  
2729

Task 3-4: Rights Attribution Analysis	
<p><b>问题:</b>  <b>Question:</b>          丙公司技术员张某,完成本职工作之余,利用业余时间,自行购买材料研发“新型焊接工具”..... (后略)          Question: Zhang, a technician of Company C, developed a "new welding tool" by himself using spare time and materials purchased by himself after completing his own work... (Omit)</p> <p><b>选项:</b>  <b>Options:</b>          A.丙公司,因张某是公司员工。          A. Company C, because... (Omit)          B.张某,因非职务发明创造。          B. Zhang, because it is not an invention made in the course of his duties.          C.双方共有,因张某员工身份关联。          C. The two parties share the same identity due to Zhang's employee status.          D.归当地政府,鼓励创新。          D. Return to the local government... (Omit)  <b>答案: B</b>  <b>Answer: B</b> </p>	<p><b>Question:</b>          A biotechnology researcher invents a novel gene-editing tool while employed at University X, using university lab equipment and government grant funds. The researcher's employment contract states, "All inventions arising from university-funded projects belong to the institution." Who owns the patent?</p> <p><b>Options:</b>          A. The researcher individually.  <b>B. University X.</b>          C. The government funding agency.          D. Shared between the researcher and University X.  <b>Answer: B</b></p>
<b>知识产权类型: 专利</b> <b>Intellectual Property Type: Patent</b>	<b>Intellectual Property Type:</b> <b>Patent</b>

Figure 29: Data example of task 3-4.

2750  
2751  
2752  
2753

2754  
2755  
2756  
2757  
2758  
2759  
2760  
2761  
2762  
2763  
2764  
2765  
2766  
2767  
2768  
2769  
2770  
2771  
2772  
2773  
2774  
2775  
2776  
2777  
2778  
2779  
2780  
2781  
2782  
2783

### Task3-5: Patent Application Examination

**Question:**

**METHOD FOR PREVENTING OVERLOAD IN  
MOBILE TELEPHONE NETWORKS BY USING  
'ALWAYS-ON' IN THE CASE OF A CALL FROM A  
MOBILE TELEPHONE**

**Abstract:**

The invention relates to a method for preventing overload in telecommunications networks with IMS by always-on for a call generated by a user, in which process said user makes a PDP Context request and a GGSN of said network provides a free IP address to him or her and the PDP Context becomes active; and wherein an S-CSCF of the network creates a record in which an association between said IP address and a characteristic identity of the IMS network is included... (Omit)

**Options:**

**A. Allowed**

**B. Refuse**

**Answer: A**



**Intellectual Property Type:**

**Patent**

2784  
2785

Figure 30: Data example of task 3-5.

2786  
2787  
2788  
2789  
2790  
2791  
2792  
2793  
2794  
2795  
2796  
2797  
2798  
2799  
2800  
2801  
2802  
2803  
2804  
2805  
2806  
2807

### Task4-1: Abstract Generation

**问题:**

**Question:**

**#权利要求**

**#Claims**

1. 一种水田搅浆平地机，包括牵引架(1)、挡泥罩(11)、平地刮板(12)、驱动装置(13)、搅拌轴(14)和搅浆刀(15)，所述牵引架(1)与行走机械连接，所述牵引架(1)用于对挡泥罩(11)进行固定..... (后略)

1. A paddy field mixing and leveling machine, comprising a traction frame (1), a mudguard cover (11), a leveling scraper (12), a driving device (13), a mixing shaft (14), and a mixing blade (15), wherein the traction frame (1) is connected to a walking machinery, and the traction frame (1) is used to fix the mudguard cover (11)... (Omitted later) application based on a scientific discovery... (Omit)

**答案: 本发明涉及一种水田搅浆平地机，主要包括牵引架、挡泥罩..... (后略)**

**Answer:** The present invention relates to a paddy field mixing and leveling machine, which mainly includes components such as a traction frame, a mud blocking cover... (Omit)

**知识产权类型: 专利**

**Intellectual Property Type: Patent**

**Question:**

**# Claims**

1. A method comprising: obtaining circulating white blood cells from a subject; treating the white blood cells with a DNA damaging agent; performing a flow cytometry based functional variant analysis (FVA)... (Omit)



**Answer:**

The present invention relates to a method for assessing the functional status of DNA double strand break (DSB) repair pathway genes in circulating white blood cells obtained from a subject... (Omit)

**Intellectual Property Type:**

**Trademark**

Figure 31: Data example of task 4-1.

2808

2809

2810

2811

2812

2813

2814

2815

2816

2817

2818

2819

2820

2821

2822

2823

2824

2825

2826

2827

2828

2829

2830

2831

2832

2833

2834

2835

2836

2837

2838

2839

2840

2841

2842

2843

2844

2845

2846

2847

2848

2849

2850

2851

2852

2853

2854

2855

2856

2857

2858

2859

2860

2861

Task 4-2: Dependent Claim Generation	
<p><b>问题:</b>  <b>Question:</b>  <b>#独立权利要求项</b>  <b>#Independent claims</b>  <b>1.自动麻将机, 其特征在于, 包括环形推牌装置, 所述环形推牌装置包括推牌槽, ..... (后略)</b>  <b>1. An automatic mahjong machine, characterized by comprising a circular card pushing device, wherein the circular card pushing device comprises a card pushing groove....(Omit)</b>  <b>请根据给定的独立权利要求项生成其对应的所有从属权利要求。</b>  <b>Please generate all dependent claims corresponding to the given independent claims.</b>  <b>答案: 根据权利要求1所述的自动麻将机, 其特征在于, 所述环形推牌装置的推牌槽的基座上设有驱动装置, 用于驱动推牌槽内的麻将牌沿推牌槽移动..... (后略)</b>  <b>Answer: The automatic mahjong machine according to claim 1, characterized in that a driving device is provided on the base of the pushing slot of the circular pushing device, for driving the mahjong tiles in the pushing slot to move along the pushing slot ... (Omit)</b></p>	<p><b>Question:</b>  <b># Independent Claim</b>  <b>1. An information processing device comprising: a processor; and a memory storing, movement information relating to movement of a vehicle that is not occupied by a user the movement of the vehicle including movement to change a parking position of the vehicle...</b>  <b>Answer:</b>  <b>Dependent Claims</b>  <b>The information processing device of claim 1, wherein the processor is further configured to receive the instruction from the external device via a wireless communication protocol.</b>  <b>3. The information processing device of claim 1, wherein the memory further stores historical movement information of the vehicle... (Omit)</b></p>
<p><b>知识产权类型: 专利</b>  <b>Intellectual Property Type: Patent</b></p>	<p><b>Intellectual Property Type:</b>  <b>Trademark</b></p>

Figure 32: Data example of task 4-2.

Task 4-3: Design-Around Solution Generation	
<p><b>问题:</b>  <b>Question:</b>  <b>请从A, B, C, D四个选项中选出题目对应的答案。</b>  <b>专利涉及散热器的蛇形迂回水道布局, 下列哪种改进最可能规避侵权?</b>  <b>In the case of Zhengzhou Research Institute v. Chen regarding infringement of new plant variety rights dispute, What are the acts of infringement that Chen was found to have committed?</b>  <b>选项:</b>  <b>Options:</b>  <b>A. 将水道壁厚减少1mm</b>  <b>A. Reduce the wall thickness of the waterway by 1mm</b>  <b>B. 设计树状分形分支水道</b>  <b>B. Design a tree like fractal branch waterway</b>  <b>C. 增加水道内部纳米涂层</b>  <b>C. Add nano coating inside the waterway</b>  <b>D. 加装水温LED指示灯</b>  <b>D. Install water temperature LED indicator light</b>  <b>答案: B</b>  <b>Answer: B</b></p>	<p><b>Question:</b>  <b>A synthetic biology patent (US 19,876,543) claims "CRISPRa activation system with dCas9-VPR fusion protein and modified sgRNA containing MS2 aptamers." Which redesign escapes infringement?</b></p>  <p><b>Options:</b>  <b>A. Replacing VPR with p300 activation domain</b>  <b>B. Using PP7 aptamers instead of MS2</b>  <b>C. Implementing zinc finger proteins instead of dCas9</b>  <b>D. Adding nuclear localization signals to sgRNA</b>  <b>Answer: C</b></p>
<p><b>知识产权类型: 专利</b>  <b>Intellectual Property Type: Patent</b></p>	<p><b>Intellectual Property Type:</b>  <b>Trademark</b></p>

Figure 33: Data example of task 4-3.

## 2862 N CASE STUDY

2863  
 2864 We provide extensive case studies for each task, including both correct and erroneous responses in  
 2865 both Chinese and English, as shown from Figure 34 to Figure 75. These case studies offer deeper  
 2866 insight into the scope of the model's capabilities in the field of intellectual property.  
 2867

2868  
 2869  Chinese Task 1-1: Correct Case English

2870  
 2871 问题:根据《中华人民共和国专利法》，以下哪项是“专利侵权纠纷中  
 2872 的现有技术抗辩”？  
 2873 Question: According to the "Patent Law of the People's Republic of  
 2874 China," which of the following refers to the "existing technology  
 2875 defense" in patent infringement disputes?  
 2876 A.被控侵权人证明实施的技术属于现有技术  
 2877 B.被控侵权人证明其技术具有创造性  
 2878 C.要求提供国际检索报告  
 2879 D.必须通过司法鉴定确认  
 2880 D.必须通过司法鉴定确认  
 2881  
 2882 gpt-4o-mini:  A.被控侵权人证明实施的技术属于现有技术  
 2883 B.被控侵权人证明其技术具有创造性  
 2884 According to the Patent Law of the People's Republic of China, the prior  
 2885 art defense in patent infringement disputes refers to the defense that the  
 2886 accused infringer can raise by proving that the technology implemented  
 2887 belongs to the prior art. This means that if the technology accused of  
 2888 infringement already existed before the patent application, the accused  
 2889 infringer can claim that it does not constitute infringement. Therefore,  
 2890 the correct option is: A  
 2891  
 2892 Question: Under the Lanham Act, what is the "Official Gazette"?

2893 A.A government database of abandoned marks.  
 2894 B.A guide for drafting trademark applications.  
 2895 C.A publication where proposed trademarks are listed for opposition.  
 2896 D.The number of each party and their representatives participating in  
 2897 the oral hearing of a reexamination case shall not exceed four.  
 2898  
 2899 gpt-4o-mini:  C. A publication where proposed trademarks are listed for opposition.  
 2900 To determine the correct answer regarding the "Official Gazette" under  
 2901 the Lanham Act, we need to analyze the options provided:  
 2902 A. A government database of abandoned marks - This option does not  
 2903 accurately describe the Official Gazette, as it is not specifically focused  
 2904 on abandoned marks.  
 2905 B. A guide for drafting trademark applications - This option is also  
 2906 incorrect, as the Official Gazette does not serve as a guide for drafting  
 2907 applications.  
 2908 C. A publication where proposed trademarks are listed for opposition -  
 2909 This option aligns with the purpose of the Official Gazette. The Official  
 2910 Gazette is indeed a publication where trademarks that have been  
 2911 applied for are published, allowing third parties to oppose the  
 2912 registration of those marks.  
 2913 D. A journal for patent filings - This option is incorrect because the  
 2914 Official Gazette is specifically related to trademarks, not patents.  
 2915

Figure 34: Correct case of task 1-1.

2891  Chinese Task 1-1: Error Case English

2892  
 2893 问题:《中华人民共和国刑法》中对于“销售侵权复制品罪”的定义是?  
 2894 Question: What is the definition of the "crime of selling infringing  
 2895 reproductions" in the Criminal Law of the People's Republic of China?  
 2896 A.以营利为目的，销售明知是侵犯著作权的复制品，数量较大的  
 2897 B.以营利为目的，销售明知是侵犯著作权的复制品，造成严重后果的  
 2898 C.以营利为目的，销售明知是侵犯著作权的复制品，情节严重的  
 2899 D.以营利为目的，销售明知是侵犯著作权的复制品，违法所得数额巨大的  
 2900 D.以营利为目的，销售明知是侵犯著作权的复制品，违法所得数额巨大的  
 2901 C.以营利为目的，销售明知是侵犯著作权的复制品，情节严重的  
 2902  
 2903 gpt-4o-mini:  C.以营利为目的，销售明知是侵犯著作权的复制品，情节严重的  
 2904 C. Selling, for profit, reproductions that one clearly knows infringe  
 2905 copyright, where the circumstances are serious.  
 2906  
 2907 Error Category: **Reasoning error**  
 2908 Error Reason: According to the Criminal Law of the People's Republic of  
 2909 China, the definition of the "crime of selling infringing reproductions"  
 2910 is "selling, for profit, reproductions that one clearly knows infringe  
 2911 copyright, where the circumstances are serious"\*\*\* Therefore, the correct  
 2912 answer is Option C.  
 2913  
 2914 Question: According to the United States Code, Title 35 - Patents, what  
 2915 is the definition of "patent"?

2916 A. A temporary monopoly granted by the government to an inventor.  
 2917 B. A trademark or service mark used in commerce.  
 2918 C. A legal document granting the inventor exclusive rights to their  
 2919 invention for a limited time  
 2920 D. A copyright protecting artistic works.  
 2921  
 2922 gpt-4o-mini:  D. A copyright protecting artistic works.  
 2923  
 2924 Error Category: **Consistency Error**  
 2925 Error Reason: A patent is indeed a legal document that grants the inventor  
 2926 exclusive rights to their invention for a limited time, typically 20 years from  
 2927 the filing date. This definition aligns with option C. Option A describes a  
 2928 patent as a temporary monopoly, which is somewhat accurate but not as  
 2929 precise as option C. Option B refers to trademarks, and option D pertains to  
 2930 copyrights, neither of which are relevant to the definition of a patent.  
 2931  
 2932  
 2933  
 2934  
 2935

Figure 35: Error case of task 1-1.

2916  
2917  
2918  
2919  
2920

Chinese      Task 1-2: Correct Case      English

问题:根据《中华人民共和国专利法(2020年修正)》第65条规定,专利侵权纠纷的解决途径不包括?  
Question: According to Article 65 of the Patent Law of the People's Republic of China (2020 Amendment), which of the following is NOT a resolution pathway for patent infringement disputes?

A. 当事人协商  
A. Negotiation between the parties.  
B. 向人民法院起诉  
B. Filing a lawsuit with a people's court.  
C. 请求市场监管部门罚款  
C. Requesting market regulatory authorities to impose fines.  
D. 请求管理专利工作的部门处理  
D. Requesting the patent administration department to handle the dispute.

gpt-4o-mini:  C. 请求市场监管部门罚款  
C. Requesting market regulatory authorities to impose fines.

According to Article 65 of the Patent Law of the People's Republic of China (Revised in 2020), the main ways to resolve patent infringement disputes include negotiation between the parties, filing a lawsuit with the people's court, and requesting the department responsible for managing patent work to handle them. However, requesting fines from market regulatory authorities is not one of the ways to resolve patent infringement disputes. Therefore, the correct answer is: Answer: C

Question: How does 35 U.S.C. § 112(d) govern multiple dependent claims?  
A. Prohibits combining multiple dependencies  
B. Allows referencing preceding claims in alternative form.  
C. Requires separate filing fees.  
D. Limits to three embodiments.

gpt-4o-mini:  B. Allows referencing preceding claims in alternative form.  
To understand how 35 U.S.C. § 112(d) governs multiple dependent claims, we need to analyze what multiple dependent claims are and how they are treated under U.S. patent law.  
1. Multiple Dependent Claims: These are claims that depend on more than one preceding claim. For example, a claim might state that it depends on both Claim 1 and Claim 2.  
2. Prohibition on Combining Dependencies: Option A states that it prohibits combining multiple dependencies. This is not accurate because § 112(d) allows for multiple dependent claims, but it does impose certain restrictions.  
3. Referencing Preceding Claims: Option B suggests that it allows referencing preceding claims in alternative form. This is indeed correct. Multiple dependent claims can refer back to more than one claim, but they must do so in a way that is clear and does not create ambiguity... (omit)

2929

2930

2931

2932

2933

2934

2935

2936

2937

2938

2939

2940

2941

2942

2943

2944

2945

2946

2947

2948

2949

2950

2951

2952

2953

2954

2955

2956

2957

2958

2959

2960

2961

2962

2963

2964

2965

2966

2967

2968

2969

Figure 36: Correct case of task 1-2.

Chinese      Task 1-2: Error Case      English

问题:根据《中华人民共和国专利法(2020年修正)》第25条,不能授予专利权的是?  
Question: According to Article 25 of the Patent Law of the People's Republic of China (2020 Amendment), which of the following is not patentable?

A. 新药制备方法  
A. A method for preparing a new drug.  
B. 疾病诊断仪器  
B. Diagnostic instruments for diseases.  
C. 植物品种的生产方法  
C. Methods for producing plant varieties.  
D. 计算机程序规则  
D. Rules of computer programs.

gpt-4o-mini:  C. 植物品种的生产方法  
C. Methods for producing plant varieties.

Error Category: **Staleness Error**  
Error Reason: According to Article 25 of the Patent Law of the People's Republic of China (2020 Amendment), the following are excluded from patentability: methods for producing plant varieties and rules of computer programs. Therefore, the correct answer is Option C.

gpt-4o-mini:  C. The inventor publicly uses the invention in the U.S. 11 months before filing.

Question: Under 35 U.S.C. § 102(b), which scenario would NOT bar patentability?  
A. The invention was on sale in Canada 13 months before filing.  
B. A U.S. journal article describes the invention 14 months before filing.  
C. The inventor publicly uses the invention in the U.S. 11 months before filing.  
D. A third party files a foreign patent application 16 months before the U.S. filing.

Error Category: **Consistency Error**  
Error Reason: A. This scenario would bar patentability because the sale occurred more than one year before the U.S. filing date, and it happened outside the U.S. B. This scenario would also bar patentability because the publication occurred more than one year before the U.S. filing date. C. This scenario would NOT bar patentability because the public use occurred within the one-year grace period before the filing date. D. This scenario would bar patentability because the foreign filing occurred more than one year before the U.S. filing date.

Figure 37: Error case of task 1-2.

2970  
2971  
2972  
2973  
2974  
2975  
2976  
2977  
2978  
2979  
2980  
2981  
2982  
2983  
2984  
2985  
2986  
2987  
2988  
2989  
2990  
2991  
2992  
2993  
2994  
2995  
2996  
2997  
2998  
2999  
3000  
3001  
3002  
3003  
3004  
3005  
3006  
3007  
3008  
3009  
3010  
3011  
3012  
3013  
3014  
3015  
3016  
3017  
3018  
3019  
3020  
3021  
3022  
3023

✓
Chinese
Task 1-3: Correct Case
English

问题: 相较于2000年《中华人民共和国专利法》，2008年修订版本新增了哪项关于现有技术抗辩的明确规定？

Question: Compared to the 2000 Patent Law of the People's Republic of China, which of the following explicit provisions regarding the "prior art defense" was newly added in the 2008 Amendment?

A. 允许被诉侵权人援引现有技术主张不侵权  
A. Allowing the alleged infringer to invoke prior art to claim non-infringement.  
B. 要求现有技术必须为国内外公开出版物  
B. Requiring prior art to be limited to publicly available publications domestically or abroad.  
C. 规定现有技术抗辩需经专利局确认  
C. Stating that prior art defenses must be confirmed by the patent office.  
D. 限定现有技术抗辩仅适用于发明专利  
D. Limiting prior art defenses only to invention patents.

gpt-4o-mini: ✓

A. 允许被诉侵权人援引现有技术主张不侵权  
A. Allowing the alleged infringer to invoke prior art to claim non-infringement.  
According to the revised Patent Law of the People's Republic of China in 2008, a clear provision on the defense of prior art has been added, allowing the accused infringer to claim non-infringement by citing prior art. Therefore, the correct answer is: Answer: A

Question: What change was made to the citation format of Public Law 112-29 in the April 2013 revision of 35 U.S.C. §5 compared to the January 2013 version?

A. The statute number was changed from 112-29 to 112-30.  
B. The effective date was removed entirely.  
C. A footnote explaining the effective date was added.  
D. The effective date was moved into parentheses following the statute number.

gpt-4o-mini: ✓

D. The effective date was moved into parentheses following the statute number.  
To determine the correct answer regarding the change made to the citation format of Public Law 112-29 in the April 2013 revision of 35 U.S.C. §5 compared to the January 2013 version, we need to analyze the options provided:  
A. The statute number was changed from 112-29 to 112-30 - This option suggests a change in the statute number, which is not typically the case in revisions unless there is a new law enacted.  
B. The effective date was removed entirely - This option implies that the effective date was completely omitted, which is unlikely as effective dates are important in legal citations.  
C. A footnote explaining the effective date was ...(omit)

Figure 38: Correct case of task 1-3.

3001  
3002  
3003  
3004  
3005  
3006  
3007  
3008  
3009  
3010  
3011  
3012  
3013  
3014  
3015  
3016  
3017  
3018  
3019  
3020  
3021  
3022  
3023

✗
Chinese
Task 1-3: Error Case
English

问题: 相较于2000年《中华人民共和国专利法》，2008年修订版本新增了哪项关于外观设计专利申请的实质性条件？

Question: Compared to the 2000 Patent Law of the People's Republic of China, which of the following substantive requirements for design patent applications was newly added in the 2008 Amendment?

A. 要求外观设计必须具有明显区别性  
A. The design must be clearly distinguishable.  
B. 增加“富有美感”的创造性标准  
B. The introduction of a "aesthetic appeal" creative standard.  
C. 规定不得与他人在先权利冲突  
C. A requirement that the design must not conflict with prior legal rights of others.  
D. 要求提交产品六面视图  
D. Submission of six-side orthographic views of the product.

gpt-4o-mini: ✗

B. 增加“富有美感”的创造性标准  
B. The introduction of a "aesthetic appeal" creative standard.

Question: Compared to the 2015 U.S. Patent Act (35 U.S.C.), what key change was made to disclosure requirements in the 2022 revision?

A. Removed specific format requirements for "written description" under Section 112.  
B. Added mandatory disclosure obligations for genetic resource origins under Section 114.  
C. Amended Section 101 to exclude patentability of software algorithms.  
D. Extended electronic filing deadlines for patent applications under Section 111.

gpt-4o-mini: ✗

C. The inventor publicly uses the invention in the U.S. 11 months before filing.

Error Category: **Obsolescence Error**  
Error Reason: According to the 2008 Amended Patent Law of the People's Republic of China, a new substantive condition was added for design patent applications: the design must meet a creative standard of "aesthetic appeal". Therefore, the correct answer is Option B.

Error Category: **Consistency Error**  
Error Reason: A. This option suggests a change in the requirements for how a written description must be formatted. B. This option implies a new requirement for disclosing the origins of genetic resources. C. This option suggests a fundamental change in what can be patented, specifically targeting software algorithms, which is a major shift but not directly related to disclosure requirements. D. This option pertains to filing deadlines rather than disclosure requirements. After evaluating these options, the most relevant change to disclosure requirements in the 2022 revision appears to be related to the written description under Section 112.

Figure 39: Error case of task 1-3.

3024  
3025  
3026  
3027  
3028  
3029  
3030  
3031  
3032  
3033  
3034  
3035  
3036  
3037  
3038  
3039  
3040  
3041  
3042  
3043  
3044  
3045  
3046  
3047  
3048  
3049  
3050  
3051  
3052  
3053  
3054  
3055  
3056  
3057  
3058  
3059  
3060  
3061  
3062  
3063  
3064  
3065  
3066  
3067  
3068  
3069  
3070  
3071  
3072  
3073  
3074  
3075  
3076  
3077

Chinese Task 1-4: Correct Case English

问题：“丹玉405号”玉米植物新品种侵权案中，凌海市农某种业科技公司被指控的侵权行为包括以下哪些？

Question: In the "Danyu 405" Corn New Plant Variety Infringement Case, what alleged infringements were committed by Linghai City Nong Mou Seed Technology Co., Ltd.?

A. 未经授权生产“丹玉405号”玉米种子  
A. Unauthorized production of "Danyu 405" corn seeds.  
B. 以其他品种名称套牌销售“丹玉405号”玉米种子  
B. By other品种名称套牌销售“丹玉405号”玉米种子  
C. Repackaging and selling "Danyu 405" corn seeds under another variety name.  
C. Re却aging and selling "Danyu 405" corn seeds under another variety name.  
D. 多次实施套牌侵权、重复侵权  
D. Re却aging and selling "Danyu 405" corn seeds under another variety name.  
D. 以上都是  
D. All of the above.

gpt-40-mini:   
D. 以上都是  
In the infringement case of the "Danyu 405" corn plant new variety, Linghai Agricultural Technology Company was accused of infringement, including: A. Unauthorized production of "Danyu 405" corn seeds - this is a clear infringement, as unauthorized production constitutes an infringement of intellectual property rights. B. Selling "Danyu 405" corn seeds under other variety names is also an infringement... (omit)

Question: In the case of "Mattel Inc. v. MGA Entertainment Inc.", What was the primary reason for the lawsuit filed by Mattel Inc. against MGA Entertainment Inc.?

A. MGA's production of Bratz dolls allegedly infringed on Mattel's copyrights.  
B. MGA's advertising campaigns were misleading consumers.  
C. MGA failed to pay licensing fees to Mattel.  
D. MGA used Mattel's trade secrets to develop its products.

gpt-40-mini:   
A. MGA's production of Bratz dolls allegedly infringed on Mattel's copyrights.  
In the case of "Mattel Inc. v. MGA Entertainment Inc.", the primary reason for the lawsuit filed by Mattel Inc. against MGA Entertainment Inc. was that MGA's production of Bratz dolls allegedly infringed on Mattel's copyrights. Mattel claimed that the Bratz dolls were developed by a former Mattel employee who had used Mattel's intellectual property in the process.  
Answer: A

Figure 40: Correct case of task 1-4.

3054  
3055  
3056  
3057  
3058  
3059  
3060  
3061  
3062  
3063  
3064  
3065  
3066  
3067  
3068  
3069  
3070  
3071  
3072  
3073  
3074  
3075  
3076  
3077

Chinese Task 1-4: Error Case English

问题：“西某股份公司、西某（中国）有限公司与宁波奇某电器有限公司等侵害商标权及不正当竞争纠纷案”中，西某股份公司及西某（中国）有限公司提起诉讼的主要原因是？

Question: In the case, what were the main reasons for Xi Mou Co., Ltd. and Xi Mou (China) Co., Ltd. to file the lawsuit?

A. 宁波奇某电器有限公司未经授权使用“西门子”商标  
A. Ningbo Qi Mou Electric Appliance Co., Ltd. used the Siemens trademark without authorization.  
B. 宁波奇某电器有限公司使用“上海西门子电器有限公司”作为企业名称  
B. Ningbo Qi Mou Electric Appliance Co., Ltd. used Shanghai Siemens Electric Appliance Co., Ltd. as its corporate name.  
C. 宁波奇某电器有限公司未经授权销售“西门子”产品  
C. Ningbo Qi Mou Electric Appliance Co., Ltd. sold Siemens products without authorization.  
D. 宁波奇某电器有限公司在产品外包装及宣传活动中使用“上海西门子电器有限公司”名称  
D. Ningbo Qi Mou Electric Appliance Co., Ltd. used the name Shanghai Siemens Electric Appliance Co., Ltd. on product packaging.

gpt-40-mini:   
A. Ningbo Qi Mou Electric Appliance Co., Ltd. used the Siemens trademark without authorization

Question: In the case of "S. Victor Whitmill v. Warner Bros. Entertainment Inc.", What was the primary defense argument presented by Warner Bros. Entertainment Inc. in response to the copyright infringement claim?

A. The tattoo design was not original.  
B. The use of the tattoo design was fair use.  
C. The tattoo design was in the public domain.  
D. The use of the tattoo design was authorized by Mike Tyson.

gpt-40-mini:   
B. The use of the tattoo design was fair use.

Error Category: Priority Error  
Error Reason: A. This constitutes trademark infringement. B. This may involve unfair competition but does not directly relate to trademark rights. C. This also constitutes trademark infringement. D. This similarly involves unfair competition. Both Options A and C involve trademark infringement. Option A, however, more directly addresses the core issue of unauthorized trademark use, making it the primary reason for filing the lawsuit.

Error Category: Consistency Error  
Error Reason: In the case of "S. Victor Whitmill v. Warner Bros. Entertainment Inc.", Warner Bros. Entertainment Inc. primarily argued that the use of the tattoo design was fair use. They contended that the depiction of the tattoo on the character in the film was transformative and did not infringe on Whitmill's copyright.

Figure 41: Error case of task 1-4.

3078

3079

3080

3081

3082

3083

3084

3085

3086

3087

3088

3089

3090

3091

3092

3093

3094

3095

3096

3097

3098

3099

3100

3101

3102

3103

3104

3105

3106

3107

3108

3109

3110

3111

3112

3113

3114

3115

3116

3117

3118

3119

3120

3121

3122

3123

3124

3125

3126

3127

3128

3129

3130

3131

Chinese      Task 1-5-1: Correct Case      English

问题: 请根据专利的标题和摘要, 给出其对应的国际专利分类号。  
**Question:** Please provide the corresponding International Patent Classification (IPC) code based on the patent's title and abstract.

标题: 一种铁氧体磁芯双面磨床  
**Title:** A Double-Sided Grinding Machine for Ferrite Cores

摘要: 本实用新型公开了一种铁氧体磁芯双面磨床, 涉及双面磨削加工技术领域, 包括框架, 框架的安装板一端上方固定安装有对接板, 在安装板和对接板上转动安装有低速转盘, 安装板的另一端转动连接有打磨盘, 低速转盘的一侧插入上磨盘和下磨盘之间, 在安装板的下方设置有低速电机; 还包括设置在低速转盘上的限位机构。  
**Abstract:** The utility model discloses a double-sided grinding machine for ferrite cores, which pertains to the technical field of double-sided grinding processing. The machine comprises a frame. A docking plate is fixed above one end of the mounting plate of the frame, and a low-speed turntable is rotatably installed on the mounting plate and the docking plate. The other end of the mounting plate is rotatably connected to a grinding disc. One side of the low-speed turntable is inserted between an upper grinding disc and a lower grinding disc.

**gpt-4o-mini:**   
 Based on the title and abstract of the patent, this invention pertains to a device for double-sided grinding processing, especially for a grinding machine for ferrite magnetic cores. According to its functions and technical field, the corresponding International Patent Classification Number (IPC) can be inferred. In the International Patent Classification, technologies related to grinding machines and grinding processing are usually classified under B24 (grinding, polishing, buffing or similar processing), and for specific technologies related to double-sided grinding machines, they may be further subdivided. Therefore, considering the content of this patent, the most appropriate International Patent Classification Number is B24B (double-sided grinding).  
**Answer:** \*\*B24B7/17\*\*

Question:Please provide the corresponding International Patent Classification (IPC) code based on the patent's title and abstract.

**Title:** Device for stimulating the clitoris using a variable pressure field and method for generating a variable pressure field

**Abstract:** An example device for stimulating a clitoris includes a housing; a drive device configured to provide a drive movement; a pressure chamber at least partially surrounded by a chamber wall; a chamber wall portion forming a portion of the chamber wall, the drive device to cause the chamber wall portion to repeatedly move between different wall positions to cause a chamber volume of the pressure chamber to repeatedly increase and decrease to generate a variable pressure field; a housing opening fluidly connected to the pressure chamber, the variable pressure field to be output via the housing opening in the form of negative and positive pressures.

Figure 42: Correct case of task 1-5-1.

Chinese      Task 1-5-1: Error Case      English

问题: 请根据专利的标题和摘要, 给出其对应的国际专利分类号。  
**Question:** Please provide the corresponding International Patent Classification (IPC) code based on the patent's title and abstract.

标题: 一种港口航道淤泥分离离去水装置  
**Title:** A Port Channel Sludge Separation and Dewatering Device

摘要: 本实用新型涉及一种港口航道淤泥分离离去水装置, 属于淤泥处理技术领域, 包括机架、封闭盖板、外筒体、内滤筒、转轴、轴承套、进泥管, 转轴内孔中设有跟随其转动且能上下升降的控制轴, 控制轴的底部固定有封闭内孔底部的封闭锥板, 转轴的上部连接有驱动机构, 控制轴的顶部连接有升降控制机构, 外筒体的中部内壁上固定有环形的接水槽, 接水槽的底部固定有排水管。  
**Abstract:** The utility model relates to a port channel sludge separation and dewatering device, belonging to the technical field of sludge treatment. The device comprises a frame, a sealed cover plate, an outer cylinder, an inner filter cylinder, a rotating shaft, a bearing sleeve, and a sludge inlet pipe. A control shaft is arranged in the inner hole of the rotating shaft, which rotates with the shaft and can move vertically.

**gpt-4o-mini:**   
 B01D

**Error Category:** Reasoning Error  
**Error Reason:** Based on the patent's title and abstract, the corresponding International Patent Classification (IPC) code is likely related to Water treatment, sludge treatment, or separation technology. According to the IPC system, potential classification codes include: B01D: Separation B03B: Apparatus for separating solids from solids using wet techniques; C02F: Water or wastewater treatment. Given the device's specific function and application, the most appropriate classification is B01D (Separation).

Question:Please provide the corresponding International Patent Classification (IPC) based on the patent's title and abstract.

**Title:** Video monitoring system

**Abstract:** An asset tracking system includes a camera adapted to capture images and output signals representative of the images. The camera may include one or more depth sensors that detect distances between the depth sensor and objects positioned within the field of view of the one or more cameras. A computer device processes the image signals and/or depth signals from cameras and determines any one or more of the following: (a) whether a patient care protocol has been properly followed; (b) what condition a patient is in; (c) whether an infection control protocol has been properly followed; and (d) whether steps have been taken to reduce the risk of a patient from falling. Alerts may be issued if any conditions of importance are detected.

**gpt-4o-mini:**   
 A61B

**Error Category:** Reasoning Error  
**Error Reason:** Based on this information, the relevant IPC classes can be identified: G06T - Image data processing or generation, which includes image analysis and processing techniques; A61B - Diagnosis, surgery, identification of persons; devices for transferring or applying fluids to the body, which is relevant due to the healthcare context of monitoring patient conditions; G08B - Signaling or calling systems, which could relate to the alerting aspect of the system. Given the focus on video monitoring in a healthcare context, the most relevant IPC classification would likely fall under A61B, as it pertains to medical technology and monitoring systems.

Figure 43: Error case of task 1-5-1.

3132  
 3133  
 3134  
 3135  
 3136  
 3137  
 3138  
 3139  
 3140  
 3141  
 3142  
 3143  
 3144  
 3145  
 3146  
 3147  
 3148  
 3149  
 3150  
 3151  
 3152  
 3153  
 3154  
 3155  
 3156  
 3157  
 3158  
 3159  
 3160  
 3161  
 3162  
 3163  
 3164  
 3165  
 3166  
 3167  
 3168  
 3169  
 3170  
 3171  
 3172  
 3173  
 3174  
 3175  
 3176  
 3177  
 3178  
 3179  
 3180  
 3181  
 3182  
 3183  
 3184  
 3185

✓ Task 1-5-2: Correct Case

Please provide the corresponding Cooperative Patent Classification (CPC) based on the patent's title and abstract.

**Title:** Fishing bait, a mouldable composition for manufacturing the fishing bait, a method for manufacturing the fishing bait and a use of the mouldable composition

**Abstract:** The present invention relates to a fishing bait, a mouldable composition for manufacturing the fishing bait, a method for manufacturing the fishing bait and a use of the mouldable composition. The mouldable composition comprises a basic solution comprising water, a gelling agent and a plasticizer. The mouldable composition further comprises at least one cross-linking agent in the volume of the basic solution.

**gpt-4o-mini:**  To determine the appropriate Cooperative Patent Classification (CPC) for the provided patent title and abstract, we need to analyze the key components of the invention.

1. Fishing Bait: This indicates that the invention is related to products used in fishing, specifically bait.
2. Mouldable Composition: This suggests a focus on materials science, particularly formulations that can be shaped or formed.
3. Manufacturing Method: This implies that the invention includes a process for creating the fishing bait.
4. Composition Components: The mention of a basic solution, gelling agent, plasticizer, and cross-linking agent indicates a chemical composition aspect.

Given these points, the CPC classification would likely fall under categories related to fishing and bait, as well as materials and compositions used in their manufacture. The relevant CPC classes for fishing bait and related compositions typically include: A01K: This class covers "Animal husbandry" and includes fishing bait. \*\*A01K 85/00: This subclass specifically relates to "Baits for fishing". Considering the focus on the composition and method of manufacturing, the classification may also touch on aspects of chemical compositions, but the primary focus remains on fishing bait.

Thus, the most appropriate CPC classification for this patent would be: Answer: A01K 85/00

Figure 44: Correct case of task 1-5-2.

Task 1-5-2: Error Case

**Question:** Please provide the corresponding Cooperative Patent Classification (CPC) based on the patent's title and abstract.

**Title:** Multipurpose machine for cultivating trees.

**Abstract:** A multipurpose machine for cultivating trees, comprising an inverted U-shape structure that enables the machine to pass over existing trees to carry out pruning, disinfection or fruit picking tasks, provided at the bottom with wheels, driven by at least one motor that autonomously facilitates the movement thereof, and respective upper frames that telescopically couple to each other, being driven by respective cylinders to move the portion of the structure on the right with respect to the one on the left in order to vary the width of the machine. Likewise, the machine has the ability to raise or lower the upper structure of the same to adapt it to the height of the trees to be cultivated.

**gpt-4o-mini:**  A01B 79/00

**Error Category:** Reasoning Error

**Error Reason:** It features a unique structure that allows it to navigate around existing trees and adjust its width and height. The relevant CPC classifications for this type of agricultural machinery typically fall under the following sections: 1. \*\*A01B\*\* - This section covers "Soil working in agriculture or forestry; Cultivating; Planting" which includes machines and implements for cultivating trees. 2. \*\*A01D\*\* - This section pertains to "Harvesting" and includes machines for picking fruits and other agricultural products. Given the multifunctional nature of the machine, particularly its capabilities for both cultivation and harvesting, the most appropriate CPC classification would likely be: \*\*A01B 79/00\*\*. This classification specifically covers "Machines or apparatus for cultivating or working the soil, specially adapted for trees or shrubs".

Figure 45: Error case of task 1-5-2.

Chinese Task 1-6: Correct Case English

请从 A, B, C, D 四个选项中选出题目对应的答案。

工程师在专利公开日至授权日期间改进技术方案并实施，专利权人主张侵权。抗辩成立的关键是什么？

Please refer to A, B. Choose the answer corresponding to the question from the four options C and D. The engineer improves the technical solution and implements it during the period from the patent disclosure date to the authorization date, and the patent owner claims infringement. What is the key to establishing a defense?

A. 先用权成立要件  
 B. 临时保护期效力  
 C. 实验性使用例外  
 D. 实验性使用例外

**gpt-4o-mini:**  In patent law, the improvement and implementation of technical solutions by the patentee between the date of patent publication and the date of authorization involve the effectiveness of the temporary protection period. According to relevant laws and regulations, after a patent application is published, the applicant enjoys a certain temporary protection period, and others are not allowed to implement the technical solution during this period. Therefore, the key to the establishment of the defense lies in the effectiveness of the temporary protection period.

Answer: B

Please select the correct answer from A, B, C, and D.

An AI company trains a neural network to design nanobots but only publishes input output examples in its patent. Competitors decode the model. What invalidates the patent?

A. Lack of enablement per In re AI Models Corp.  
 B. Algorithmic abstraction as prior art  
 C. Trade secret protection for training data  
 D. Fair use of published research

**gpt-4o-mini:**  To determine which option invalidates the patent, we need to analyze each choice in the context of patent law and the situation described.

A. Lack of enablement per In re AI Models Corp. - This option suggests that the patent does not provide sufficient detail for someone skilled in the art to replicate the invention. If the patent only includes input output examples without a clear methodology or sufficient detail on how to create the nanobots, it could indeed be considered lacking in enablement.

B. Algorithmic abstraction as prior art - This option implies that the methods or algorithms used in the neural network are already known and thus cannot be patented. However, the scenario does not indicate that the competitors are using prior art; they are decoding the model based on the examples provided.

C. Trade secret protection for training data - This option suggests that the training data used to develop the neural network is protected as a trade secret. However, the question focuses on the patent's validity rather than the protection of trade secrets.

D. Fair use of published research - This option refers to the ability of competitors to use published research without infringing on the patent. However, the competitors are decoding the model based on the patent's input output examples, which may not fall under fair use.

Given these analyses, the most relevant option that directly addresses the potential invalidation ... (omit)

Figure 46: Correct case of task 1-6.

3186  
 3187  
 3188  
 3189  
 3190 问题:运动品牌商标“QUICKFIT”因长期未使用被第三方在健身APP领域  
 3191 注册。争议焦点涉及什么?  
 3192 **Question:** The sports brand trademark "QUICKFIT" was registered by  
 3193 a third party in the fitness app domain due to prolonged non-use. What  
 3194 does the dispute focus on?  
 3195 A. 商标通用化认定  
 3196 A. Determination of trademark genericide.  
 3197 B. 跨类保护范围  
 3198 B. Scope of cross-class protection.  
 3199 C. 驰名商标认定  
 3200 C. Determination of a well-known trademark.  
 3201 D. 商标共存协议  
 3202 D. Trademark coexistence agreement.  
 3203  
 3204  
 3205  
 3206  
 3207  
 3208  
 3209  
 3210  
 3211  
 3212  
 3213  
 3214  
 3215  
 3216  
 3217  
 3218  
 3219  
 3220  
 3221  
 3222  
 3223  
 3224  
 3225  
 3226  
 3227  
 3228  
 3229  
 3230  
 3231  
 3232  
 3233  
 3234  
 3235  
 3236  
 3237  
 3238  
 3239

 Chinese Task 1-6: Error Case English

问题:运动品牌商标“QUICKFIT”因长期未使用被第三方在健身APP领域  
 注册。争议焦点涉及什么?  
**Question:** The sports brand trademark "QUICKFIT" was registered by  
 a third party in the fitness app domain due to prolonged non-use. What  
 does the dispute focus on?  
 A. Nominative fair use under New Kids.  
 B. Parody defense under Rogers v. Koons.  
 C. First Amendment artistic expression.  
 D. Trademark tarnishment.

gpt-4o-mini:  B. 跨类保护范围  
 B. Scope of cross-class protection.

**Error Category:** **Consistency Error**  
**Error Reason:** Among the options: A. Determination of trademark genericide: Refers to a trademark becoming a generic term due to widespread use, which likely does not apply here. B. Scope of cross-class protection: Relates to trademark protection across different categories, which is likely relevant. C. Determination of a well-known trademark: Generally involves a trademark's fame and broader protection, but here, since the trademark was not actively used, this is likely not applicable. D. Trademark coexistence agreement: Usually involves agreements between trademark holders, which is likely irrelevant here. The most relevant option is B, as the dispute centers on the use and protection of the trademark in a different category (e.g., fitness apps). Cross-class protection would determine whether the original trademark owner can block third-party registration in unrelated fields, even if the trademark was not actively used.

gpt-4o-mini:  B. Parody defense under Rogers v. Koons.

**Error Category:** **Consistency Error**  
**Error Reason:** A. This doctrine allows for the use of a trademark to refer to the trademarked goods or services, but it typically applies when the use is necessary to identify the product and does not suggest sponsorship or endorsement. B. This doctrine allows for the use of a trademark in a way that is transformative and serves as a parody, which could be relevant if the designs are intended to mock or critique the luxury brands. C. This could apply if the designs are considered a form of artistic expression, but it may not provide a strong defense if the use of the logos is deemed to cause confusion. D. This doctrine applies when a trademark is used in a way that harms its reputation, which could be a concern for luxury brands if the anti-brand designs are seen as derogatory.

Figure 47: Error case of task 1-6.

3215  
 3216  
 3217  
 3218  
 3219  
 3220  
 3221  
 3222  
 3223  
 3224  
 3225  
 3226  
 3227  
 3228  
 3229  
 3230  
 3231  
 3232  
 3233  
 3234  
 3235  
 3236  
 3237  
 3238  
 3239

 Chinese Task 1-7: Correct Case English

请从A, B, C, D四个选项中选出题目对应的答案。  
 下列哪项不属于职务发明创造的情形?  
 Please refer to A, B. Choose the answer corresponding to the question from the four options C and D. Which of the following does not fall under the category of a service invention or creation?  
 A. 退休后10个月内完成与原工作相关的发明  
 A. Complete inventions related to the original job within 10 months after retirement  
 B. 利用个人设备完成的发明  
 B. Inventions made using personal devices  
 C. 履行单位临时指派任务完成的发明  
 C. Inventions completed by fulfilling temporary tasks assigned by the unit  
 D. 调离原单位后1年内作出的与原职位相关的发明  
 D. Inventions related to the original position made within one year after being transferred from the original unit

gpt-4o-mini:  Service inventions and creations refer to inventions and creations completed within the scope of duties. According to relevant laws and regulations, service inventions usually include inventions completed during work, utilizing unit resources, or under tasks assigned by the unit. Analysis options:  
 A. Inventions related to the original job completed within 10 months after retirement - this falls under the category of service inventions, as they are related to the original job even after retirement.  
 B. Inventions made using personal devices - this does not belong to service inventions, as it is made using personal devices rather than organizational resources. Invention completed by fulfilling temporary tasks assigned by the unit - This is a service invention because it was completed under the unit's assignment.  
 D. Inventions made within one year after being transferred from the original unit that are related to the original position may also be considered as service inventions, but are usually no longer considered as service inventions after the transfer. In summary, option B is the least suitable situation for service inventions and creations. Answer: \*\* B \*\*

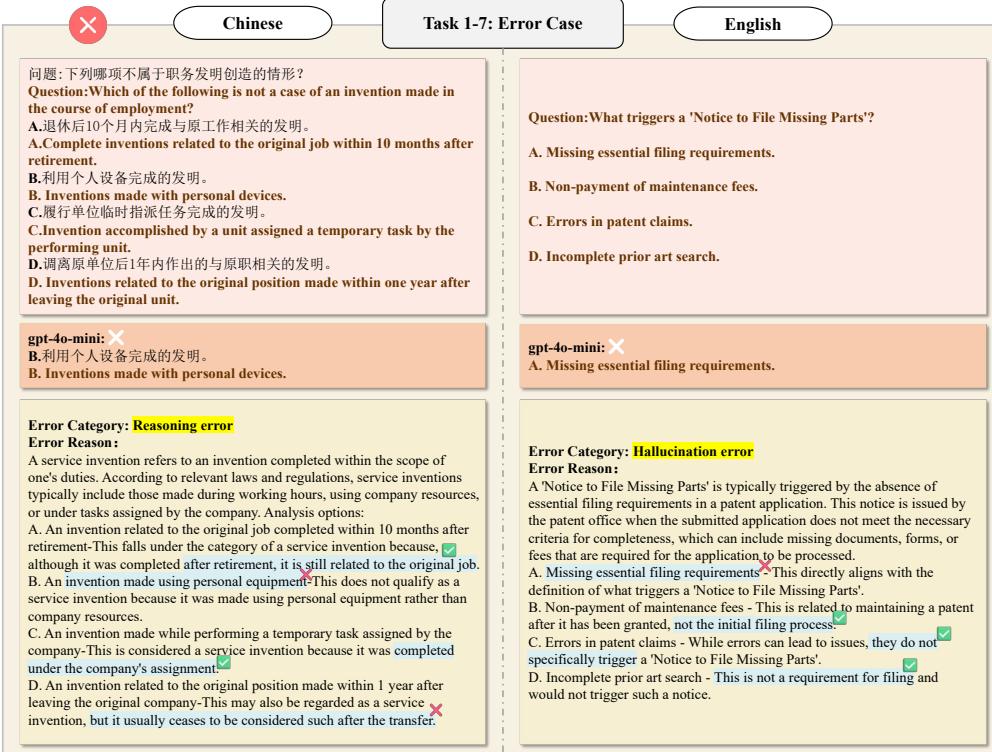
Please select the correct answer from A, B, C, and D.

Question: Which entities can be designated as International Searching Authorities?  
 A. Any national office  
 B. Only intergovernmental organizations  
 C. National offices or intergovernmental organizations  
 D. Non-governmental organizations

gpt-4o-mini:  International Searching Authorities (ISAs) are designated entities that conduct international searches for patent applications under the Patent Cooperation Treaty (PCT). According to the PCT regulations, ISAs can be either national patent offices or intergovernmental organizations that have been authorized to perform this function.  
 Given this information, the correct answer is: C. National offices or intergovernmental organizations.  
 Answer: C

Figure 48: Correct case of task 1-7.

3240  
3241  
3242  
3243  
3244  
3245  
3246  
3247  
3248  
3249  
3250  
3251  
3252  
3253  
3254  
3255  
3256  
3257  
3258  
3259  
3260  
3261  
3262  
3263  
3264  
3265  
3266  
3267  
3268  
3269  
3270  
3271  
3272  
3273  
3274  
3275  
3276  
3277  
3278  
3279  
3280  
3281  
3282  
3283  
3284  
3285  
3286  
3287  
3288  
3289  
3290  
3291  
3292  
3293



**Chinese**      **Task 1-7: Error Case**      **English**

问题:下列哪项不属于职务发明创造的情形?  
Question:Which of the following is not a case of an invention made in the course of employment?

A.退休后10个月内完成与原工作相关的发明。  
A. Complete inventions related to the original job within 10 months after retirement.

B.利用个人设备完成的发明。  
B. Inventions made with personal devices.

C.履行单位临时指派任务完成的发明。  
C. Invention accomplished by a unit assigned a temporary task by the performing unit.

D.调离原单位后1年内作出的与原职相关的发明。  
D. Inventions related to the original position made within one year after leaving the original unit.

**gpt-4o-mini:**  B.利用个人设备完成的发明。  
B. Inventions made with personal devices.

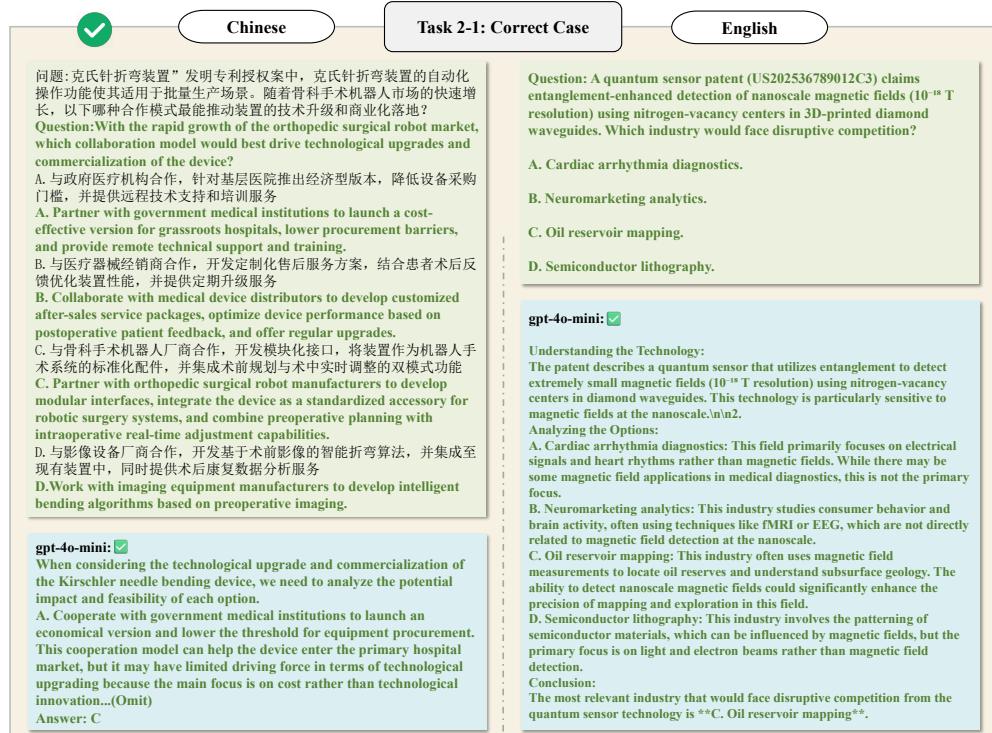
**Error Category:** Reasoning error  
**Error Reason:**  
A service invention refers to an invention completed within the scope of one's duties. According to relevant laws and regulations, service inventions typically include those made during working hours, using company resources, or under tasks assigned by the company. Analysis options:  
A. An invention related to the original job completed within 10 months after retirement. This falls under the category of a service invention because, , although it was completed after retirement, it is still related to the original job.  
B. An invention made using personal equipment. This does not qualify as a service invention because it was made using personal equipment rather than company resources.  
C. An invention made while performing a temporary task assigned by the company. This is considered a service invention because it was completed under the company's assignment.  
D. An invention related to the original position made within 1 year after leaving the original company. This may also be regarded as a service  invention, but it usually ceases to be considered such after the transfer.

**gpt-4o-mini:**  A. Missing essential filing requirements.

**Error Category:** Hallucination error  
**Error Reason:**  
A 'Notice to File Missing Parts' is typically triggered by the absence of essential filing requirements in a patent application. This notice is issued by the patent office when the submitted application does not meet the necessary criteria for completeness, which can include missing documents, forms, or fees that are required for the application to be processed.  
A. Missing essential filing requirements.  This directly aligns with the definition of what triggers a 'Notice to File Missing Parts'.  
B. Non-payment of maintenance fees - This is related to maintaining a patent after it has been granted, not the initial filing process.  
C. Errors in patent claims - While errors can lead to issues, they do not specifically trigger a 'Notice to File Missing Parts'.  
D. Incomplete prior art search - This is not a requirement for filing and would not trigger such a notice.

Figure 49: Error case of task 1-7.

3269  
3270  
3271  
3272  
3273  
3274  
3275  
3276  
3277  
3278  
3279  
3280  
3281  
3282  
3283  
3284  
3285  
3286  
3287  
3288  
3289  
3290  
3291  
3292  
3293



**Chinese**      **Task 2-1: Correct Case**      **English**

问题:“克氏针折弯装置”发明专利授权案中,克氏针折弯装置的自动化操作功能使其适用于批量生产场景。随着骨科手术机器人市场的快速增长,以下哪种合作模式最能推动装置的技术升级和商业化落地?  
Question:With the rapid growth of the orthopedic surgical robot market, which collaboration model would best drive technological upgrades and commercialization of the device?

A.与政府医疗机构合作,针对基层医院推出经济型版本,降低设备采购门槛,并提供远程技术支持和培训服务  
A. Partner with government medical institutions to launch a cost-effective version for grassroots hospitals, lower procurement barriers, and provide remote technical support and training.

B.与医疗器械经销商合作,开发定制化售后服务方案,结合患者术后反馈优化装置性能,并提供定期升级服务  
B. Collaborate with medical device distributors to develop customized after-sales service packages, optimize device performance based on postoperative patient feedback, and offer regular upgrades.

C.与骨科手术机器人厂商合作,开发模块化接口,将装置作为机器人手术系统的标准化配件,并集成术前规划与术中实时调整的双模式功能  
C. Partner with orthopedic surgical robot manufacturers to develop modular interfaces, integrate the device as a standardized accessory for robotic surgery systems, and combine preoperative planning with intraoperative real-time adjustment capabilities.

D.与影像设备厂商合作,开发基于术前影像的智能折弯算法,并集成至现有装置中,同时提供术后康复数据分析服务  
D. Work with imaging equipment manufacturers to develop intelligent bending algorithms based on preoperative imaging.

**gpt-4o-mini:**  When considering the technological upgrade and commercialization of the Kirschner needle bending device, we need to analyze the potential impact and feasibility of each option.  
A. Cooperate with government medical institutions to launch an economical version and lower the threshold for equipment procurement. This cooperation model can help the device enter the primary hospital market, but it may have limited driving force in terms of technological upgrading because the main focus is on cost rather than technological innovation... (Omit)  
Answer: C

**gpt-4o-mini:**  Understanding the Technology:  
The patent describes a quantum sensor that utilizes entanglement to detect extremely small magnetic fields ( $10^{-19}$  T resolution) using nitrogen-vacancy centers in 3D-printed diamond waveguides. This technology is particularly sensitive to magnetic fields at the nanoscale.  
Analyzing the Options:  
A. Cardiac arrhythmia diagnostics: This field primarily focuses on electrical signals and heart rhythms rather than magnetic fields. While there may be some magnetic field applications in medical diagnostics, this is not the primary focus.  
B. Neuromarketing analytics: This industry studies consumer behavior and brain activity, often using techniques like fMRI or EEG, which are not directly related to magnetic field detection at the nanoscale.  
C. Oil reservoir mapping: This industry often uses magnetic field measurements to locate oil reserves and understand subsurface geology. The ability to detect nanoscale magnetic fields could significantly enhance the precision of mapping and exploration in this field.  
D. Semiconductor lithography: This industry involves the patterning of semiconductor materials, which can be influenced by magnetic fields, but the primary focus is on light and electron beams rather than magnetic field detection.  
Conclusion:  
The most relevant industry that would face disruptive competition from the quantum sensor technology is **C. Oil reservoir mapping**.

Figure 50: Correct case of task 2-1.

3294

3295

3296

3297

3298

3299

3300

3301

3302

3303

3304

3305

3306

3307

3308

3309

3310

3311

3312

3313

3314

3315

3316

3317

3318

3319

3320

3321

3322

3323

3324

3325

3326

3327

3328

3329

3330

3331

3332

3333

3334

3335

3336

3337

3338

3339

3340

3341

3342

3343

3344

3345

3346

3347

Chinese: Task 2-1: Error Case

English: Task 2-1: Error Case

Chinese: 请从A, B, C, D四个选项中选出题目对应的答案。“耐克”用于运动装备，“奈克”用于运动装，都有一定市场份额。是否构成商标侵权？

English: Please select the correct answer from A, B, C, and D. A toy manufacturer makes a toy that has a similar shape to a patented toy and also uses a character from a copyrighted cartoon on it. What infringements occur?

Chinese: A. 需考虑商标的知名度和影响力  
B. 是，因标识近似且商品相同  
C. 否，因“奈克”有自身特色  
D. 是，但受指示性使用免责

English: A. Only patent infringement.  
B. Only copyright infringement.  
C. Both patent and copyright infringement.  
D. It's okay as long as the toys are sold at a low price.

gpt-4o-mini: ✗  
C. Both patent and copyright infringement.

Error Category: Reasoning Error  
Error Reason: When analyzing the situation of the "Nike" and "Nike" trademarks, we need to consider several key factors. Firstly, the popularity and influence of a trademark are important criteria for determining trademark infringement. If Nike is a well-known brand, its trademark protection will be stronger. Secondly, the similarity of identification and the similarity of goods are also important criteria for determining infringement. If "Nike" and "Nike" are visually or phonetically similar and belong to the same product category, the likelihood of infringement will increase. Option A mentions the need to consider the brand's visibility and influence, which is an important legal factor to consider. Option B directly states that infringement is caused by similar labeling and identical products, which is also a reasonable judgment. Option C believes that "Nike" has its own characteristics and therefore does not constitute infringement, which may be valid in certain circumstances, but if "Nike" is too similar to "Nike", it may still constitute infringement. Option D mentions the indicative use exemption, which usually applies in cases of reasonable use, but whether it applies in this context requires specific analysis. Taking all these factors into consideration, the most reasonable answer is A, because the popularity and influence of a trademark are the basis for determining whether it constitutes infringement.

Figure 51: Error case of task 2-1.

Chinese: Task 2-2: Correct Case

English: Task 2-2: Correct Case

Chinese: 问题：一首歌曲《爱的旋律》，有独特的旋律和歌词，被诉作品：《情的乐章》，部分旋律相似但歌词不同。有评论认为旋律有借鉴是否构成著作权侵权？

English: Question: A song titled "The Melody of Love" has a unique melody and lyrics, while the accused work, "The Symphony of Love," has a similar melody but different lyrics. Some critics believe the melody is borrowed. Does this constitute copyright infringement?

Chinese: A. 是，因部分旋律相似  
B. 否，因歌词不同  
C. 否，但受教使用免责  
C. 是，但受教使用免责  
C. Yes, but disclaimer for educational use.  
D. 需确定相似部分是否构成实质性相似  
D. It is necessary to determine whether the similar parts constitute substantial similarity.

gpt-4o-mini: ✗  
To determine whether there is copyright infringement, the key lies in the existence of substantial similarity. Although "The Melody of Love" and "The Symphony of Love" have some similarities in melody, the different lyrics may affect the judgment of infringement. Just having some melody similarity doesn't necessarily mean infringement, so option A is not conclusive. And the difference in lyrics alone doesn't rule out the possibility of melody-related copyright infringement, so option B is incorrect. Option C's mention of educational use disclaimer is not relevant to the core issue here, which is mainly about determining similarity. Therefore, a detailed analysis of the similar parts is required to determine whether they constitute substantial similarity.

Answer: D. It is necessary to determine whether the similar parts constitute substantial similarity.

English: Question: An intern at Company H takes a draft of a new product design home without permission. Is this an infringement of trade secrets?

Chinese: A. No, as long as the intern doesn't show it to anyone.  
B. Only if the intern intends to use it for personal gain.  
C. Yes, because they took the confidential draft without permission.  
D. Only if the new product is a revolutionary one.

gpt-4o-mini: ✗  
Analysis:  
1. A. This option is incorrect. The act of taking the draft out of the workplace without permission already violates the confidentiality of the trade secret.  
2. B. This is also wrong. The company has the right to keep its product - design information private, and the intern's action of taking it home without consent breaches that right, regardless of the motive for personal gain.  
3. C. This is the correct answer. By taking the confidential draft of the new product design home without permission, the intern has directly violated the company's protection of its trade secrets. Unauthorized access and removal of such information are key elements in determining trade - secret infringement.  
4. D. This option is incorrect. The revolutionary nature of the product has no bearing on whether the act of taking the draft is an infringement. Trade - secret protection applies to any confidential business information that gives the company a competitive edge, not just to revolutionary products.

Answer: C. Yes, because they took the confidential draft without permission.

Figure 52: Correct case of task 2-2.

3348	
3349	
3350	
3351	
3352	问题:美国某公司拥有一项关于医疗设备的发明专利。中国丙企业生产的医疗设备在工作原理上与美国公司专利产品相同,但在外观设计上有较大差异。依据美国专利法及相关案例,判定丙企业是否侵权的关键在于:
3353	<b>Question: A certain American company holds a patent for an invention related to medical devices. The medical devices produced by Chinese company C are identical in terms of working principles to the American company's patented product, but they differ significantly in terms of appearance design. According to U.S. patent law and relevant case law, the key factor in determining whether company C has infringed is:</b>
3354	A.工作原理相同就构成侵权,不论外观差异
3355	B.若工作原理相同,外观设计不同则构成侵权
3356	C.外观设计是否导致产品的实质性不同
3357	D.外观设计是否导致产品的实质性不同
3358	<b>B</b> Whether the differences in appearance lead to substantial differences in the products.
3359	C.丙企业产品的生产成本是否低于美国公司产品
3360	C. Is the production cost of the products of Enterprise C lower than that of the products of the American company.
3361	D.美国公司专利的申请时间
3362	D. The application time for patents of American companies.

3363 **gpt-4o-mini: X**  
3364 A. 工作原理相同就构成侵权, 不论外观差异  
3365 A. If the working principles are the same, it constitutes infringement, regardless of the differences in appearance.

**Error Category: Consistency error and Reasoning error**

**Error Reason:** According to U.S. patent law, the key factor in determining infringement is the claims of the patent and the substantial differences between the products. Although identical working principles may raise concerns about infringement, the more important consideration is whether the differences in appearance design result in substantial differences in the function or effect of the products. Therefore, the correct answer is: Answer: "B".

**Question:** A toy manufacturer makes a toy that has a similar shape to a patented toy and also uses a character from a copyrighted cartoon on it. What infringements occur?

- A. Only patent infringement.
- B. Only copyright infringement.
- C. Both patent and copyright infringement.
- D. It's okay as long as the toys are sold at a low price.

**gpt-4o-mini: X**  
D. It's okay as long as the toys are sold at a low price.

**Error Category: Consistency error**   
**Error Reason:** The explanation clearly states that both patent infringement and copyright infringement are occurring due to the actions of the toy manufacturer. However, the selected answer D claims that it's okay as long as the toys are sold at a low price, which is inconsistent with the reasoning in the explanation. There is a contradiction between the answer chosen and the reasoning provided to support it.

Figure 53: Error case of task 2-2.

3378		Chinese	Task 2-3
3379			
3380	问题:某集体商标侵权案,法院认定:权利人认证费用为每件产品50元;侵权人销售未认证产品30万件。赔偿金额应为?		
3381	<b>Question: In a case of collective trademark infringement, the court</b>		
3382	<b>ruled that the certification fee for the right holder was 50 yuan per product. The infringer sold 20,000 uncertified products. What should be the amount of compensation?</b>		
3383	A. $20,000 \times 50 \times 3 = 300$ 万		
3384	B. $20,000 \times 50 \times 3 \times 3 = 3$ million.		
3385	C. $20,000 \times 50 = 100$ 万		
3386	D. $20,000 \times 50 \times 2 = 1$ million.		
	E. $20,000 \times 50 \times 2 \times 2 = 200$ 万		
	F. $20,000 \times 50 \times 2 \times 2 \times 2 = 2$ million.		
	G. 法院酌定50万元		
	H. The court determined 500,000 yuan at its discretion.		

Question: Company B sued Company A for patent infringement, and the court ordered Company A to compensate Company B for the actual losses of US\$600,000 and bear the reasonable attorney's fees of US\$150,000 for Company B. How much compensation does Company A have to pay in total under U.S. law?

- A. \$600,000.
- B. \$750,000.
- C. \$1,200,000.
- D. \$1,800,000.

**gpt-4o-mini:**

Information Analysis:  
Certification fee per product for the right - holder: 50 yuan. Number of uncertified products sold by the infringer: 20,000.  
Calculation Method  
The total certification fee for uncertified products is calculated as follows: Number of products × Certification fee per product = 20,000 × 50 yuan = 1 million yuan.  
Analysis:  
A. This option applies a 3 - fold compensation. However, there is no indication in the question that a 3 - fold compensation should be used.  
B. This option calculates the compensation based on the actual loss.

**gpt-4o-mini:**  **Information Extraction:**  
Actual losses that Company A needs to compensate Company B for: US\$600,000. Reasonable attorney's fees that Company A needs to bear for Company B: US\$150,000.

**Calculation Process:**  
Substituting the given values: Total compensation = US\$600,000 + US\$150,000 = US\$750,000.

**Analysis:**

- A. This option only includes the actual losses and ignores the attorney's fees, so it is incorrect.
- B. This option correctly sums up the actual losses and the attorney's fees, which is in line with the legal requirements for compensation in this case.
- C. This amount has no basis in the calculation according to the given information and is incorrect.
- D. This amount is also not related to the correct calculation of the sum of actual losses and attorney's fees, so it is incorrect.

Answer: B. \$750,000

Figure 54: Correct case of task 2-3.

3402

3403  Chinese Task 2-3: Error Case English

3404

3405

3406 问题: 某茶商冒用“西湖龙井”地理标志, 销售金额300万元, 获利50万元。赔偿计算依据应为?

3407 Question: A tea merchant falsely used the "West Lake Longjing" geographical indication, with sales amounting to 3 million yuan and a profit of 500,000 yuan. What should be the basis for calculating the compensation?

3408 A. 权利人实际损失 (需举证) B. 侵权获利50万元 C. 销售金额300万元 D. 法定赔偿500万元

3409 A. Actual loss of the rights holder (evidence required). B. Infringer's profit of 500,000 yuan. C. Sales amount of 3 million yuan. D. Statutory compensation of 5 million yuan.

3410

3411 **gpt-4o-mini: ✗**  
D. 参加复审案件口头审理的每方当事人及其代理人的人数不得超过四人  
D. The number of each party and their representatives participating in the oral hearing of a reexamination case shall not exceed four.

3412

3413

3414

3415

3416

3417

3418

3419

3420

3421

3422

3423

3424

3425

3426

3427

3428

3429

3430

3431

3432

3433 问题: 某基因编辑CRISPR专利在欧美被认定为无效, 但在中国维持有效, 其全球估值应如何调整?

3434 Question: A CRISPR patent for gene editing has been deemed invalid in Europe and the United States, but remains valid in China. How should its global valuation be adjusted?

3435 A. 按中国市场规模计算

3436 A. Calculated by the scale of the Chinese market.

3437 B. 完全失去商业价值

3438 B. Completely lose commercial value.

3439 C. 依赖专利池交叉许可

3440 C. Rely on patent pool cross-licensing.

3441 D. 需重新提交PCT申请

3442 D. The PCT application needs to be resubmitted.

3443

3444

3445

3446

3447

3448

3449

3450

3451

3452

3453

3454

3455

3456

3457

3458

3459

3460

3461

3462

3463

3464

3465

3466

3467

3468

3469

3470

3471

3472

3473

3474

3475

3476

3477

3478

3479

3480

3481

3482

3483

3484

3485

3486

3487

3488

3489

3490

3491

3492

3493

3494

3495

3496

3497

3498

3499

3500

3501

3502

3503

3504

3505

3506

3507

3508

3509

3510

3511

3512

3513

3514

3515

3516

3517

3518

3519

3520

3521

3522

3523

3524

3525

3526

3527

3528

3529

3530

3531

3532

3533

3534

3535

3536

3537

3538

3539

3540

3541

3542

3543

3544

3545

3546

3547

3548

3549

3550

3551

3552

3553

3554

3555

3556

3557

3558

3559

3560

3561

3562

3563

3564

3565

3566

3567

3568

3569

3570

3571

3572

3573

3574

3575

3576

3577

3578

3579

3580

3581

3582

3583

3584

3585

3586

3587

3588

3589

3590

3591

3592

3593

3594

3595

3596

3597

3598

3599

3600

3601

3602

3603

3604

3605

3606

3607

3608

3609

3610

3611

3612

3613

3614

3615

3616

3617

3618

3619

3620

3621

3622

3623

3624

3625

3626

3627

3628

3629

3630

3631

3632

3633

3634

3635

3636

3637

3638

3639

3640

3641

3642

3643

3644

3645

3646

3647

3648

3649

3650

3651

3652

3653

3654

3655

3656

3657

3658

3659

3660

3661

3662

3663

3664

3665

3666

3667

3668

3669

3670

3671

3672

3673

3674

3675

3676

3677

3678

3679

3680

3681

3682

3683

3684

3685

3686

3687

3688

3689

3690

3691

3692

3693

3694

3695

3696

3697

3698

3699

3700

3701

3702

3703

3704

3705

3706

3707

3708

3709

3710

3711

3712

3713

3714

3715

3716

3717

3718

3719

3720

3721

3722

3723

3724

3725

3726

3727

3728

3729

3730

3731

3732

3733

3734

3735

3736

3737

3738

3739

3740

3741

3742

3743

3744

3745

3746

3747

3748

3749

3750

3751

3752

3753

3754

3755

3756

3757

3758

3759

3760

3761

3762

3763

3764

3765

3766

3767

3768

3769

3770

3771

3772

3773

3774

3775

3776

3777

3778

3779

3780

3781

3782

3783

3784

3785

3786

3787

3788

3789

3790

3791

3792

3793

3794

3795

3796

3797

3798

3799

3800

3801

3802

3803

3804

3805

3806

3807

3808

3809

3810

3811

3812

3813

3814

3815

3816

3817

3818

3819

3820

3821

3822

3823

3824

3825

3826

3827

3828

3829

3830

3831

3832

3833

3834

3835

3836

3837

3838

3839

3840

3841

3842

3843

3844

3845

3846

3847

3848

3849

3850

3851

3852

3853

3854

3855

3856

3857

3858

3859

3860

3861

3862

3863

3864

3865

3866

3867

3868

3869

3870

3871

3872

3873

3874

3875

3876

3877

3878

3879

3880

3881

3882

3883

3884

3885

3886

3887

3888

3889

3890

3891

3892

3893

3894

3895

3896

3897

3898

3899

3900

3901

3902

3903

3904

3905

3906

3907

3908

3909

3910

3911

3912

3913

3914

3915

3916

3917

3918

3919

3920

3921

3922

3923

3924

3925

3926

3927

3928

3929

3930

3931

3932

3933

3934

3935

3936

3937

3938

3939

3940

3941

3942

3943

3944

3945

3946

3947

3948

3949

3950

3951

3952

3953

3954

3955

3956

3957

3958

3959

3960

3961

3962

3963

3964

3965

3966

3967

3968

3969

3970

3971

3972

3973

3974

3975

3976

3977

3978

3979

3980

3981

3982

3983

3984

3985

3986

3987

3988

3989

3990

3991

3992

3993

3994

3995

3996

3997

3998

3999

4000

4001

4002

4003

4004

4005

4006

4007

4008

4009

4010

4011

4012

4013

4014

4015

4016

4017

4018

4019

4020

4021

4022

4023

4024

4025

4026

4027

4028

4029

4030

4031

4032

4033

4034

4035

4036

4037

4038

4039

4040

4041

4042

4043

4044

4045

4046

4047

4048

4049

4050

4051

4052

4053

4054

4055

4056

4057

4058

4059

4060

4061

4062

4063

4064

4065

4066

4067

4068

4069

4070

4071

4072

4073

4074

4075

4076

4077

4078

4079

4080

4081

4082

4083

4084

4085

4086

4087

4088

4089

4090

4091

4092

4093

4094

4095

4096

4097

4098

4099

4100

4101

4102

4103

4104

4105

4106

4107

4108

4109

4110

4111

4112

4113

4114

4115

4116

4117

4118

4119

4120

4121

4122

4123

4124

4125

4126

4127

4128

4129

4130

4131

4132

4133

4134

4135

4136

4137

4138

4139

4140

4141

4142

4143

4144

4145

4146

4147

4148

4149

4150

4151

4152

4153

4154

4155

4156

4157

4158

4159

4160

4161

4162

4163

4164

4165

4166

4167

4168

4169

4170

4171

4172

4173

4174

4175

4176

4177

4178

4179

4180

4181

4182

4183

4184

4185

4186

4187

4188

4189

4190

4191

4192

4193

4194

4195

4196

4197

4198

4199

4200

4201

4202

4203

4204

4205

4206

4207

4208

4209

4210

4211

4212

4213

4214

4215

4216

4217

4218

4219

4220

4221

4222

4223

4224

4225

4226

4227

4228

4229

4230

4231

4232

4233

4234

4235

4236

4237

4238

4239

4240

4241

4242

4243

4244

4245

4246

4247

4248

4249

4250

4251

4252

4253

4254

4255

4256

4257

4258

4259

4260

4261

4262

4263

4264

4265

4266

4267

4268

4269

4270

4271

4272

4273

4274

4275

4276

4277

4278

4279

4280

4281

4282

4283

4284

4285

4286

4287

4288

4289

4290

4291

4292

4293

4294

4295

4296

4297

4298

4299

4300

4301

4302

4303

4304

4305

4306

4307

4308

4309

4310

4311

4312

4313

4314

4315

4316

4317

4318

4319

4320

4321

4322

4323

4324

4325

4326

4327

4328

4329

4330

4331

4332

4333

4334

4335

4336

4337

4338

4339

4340

4341

4342

4343

4344

4345

4346

4347

4348

4349

4350

4351

4352

4353

4354

4355

4356

4357

4358

4359

4360

4361

4362

4363

4364

4365

4366

4367

4368

4369

4370

4371

4372

4373

4374

4375

4376

4377

4378

4379

4380

4381

4382

4383

4384

4385

4386

4387

4388

4389

4390

4391

4392

4393

4394

4395

4396

4397

4398

4399

4400

4401

4402

4403

4404

4405

4406

4407

4408

4409

4410

4411

4412

4413

4414

4415

4416

4417

4418

4419

4420

4421

4422

4423

4424

4425

4426

4427

4428

4429

4430

4431

4432

4433

4434

4435

4436

4437

4438

4439

4440

4441

4442

4443

4444

4445

4446

4447

4448

4449

4450

4451

4452

4453

4454

4455

4456

4457

4458

4459

4460

4461

4462

4463

4464

4465

4466

4467

4468

4469

4470

4471

4472

4473

4474

4475

4476

4477

4478

4479

4480

4481

4482

4483

4484

4485

4486

4487

4488

4489

4490

4491

4492

4493

4494

4495

4496

4497

4498

4499

4500

4501

4502

4503

4504

4505

4506

4507

4508

4509

4510

4511

4512

4513

4514

4515

4516

4517

4518

4519

4520

4521

4522

4523

4524

4525

4526

4527

4528

4529

4530

4531

4532

4533

4534

4535

4536

4537

4538

4539

4540

4541

4542

4543

4544

4545

4546

4547

4548

4549

4550

4551

4552

4553

4554

4555

4556

4557

4558

4559

4560

4561

4562

4563

4564

4565

4566

4567

4568

4569

4570

4571

4572

4573

4574

4575

4576

4577

4578

4579

4580

4581

4582

4583

4584

4585

4586

4587

4588

4589

4590

4591

4592

4593

4594

4595

4596

4597

4598

4599

4600

4601

4602

4603

4604

4605

4606

4607

4608

4609

4610

4611

4612</p

3456

3457

3458

3459

3460

3461

3462

3463

3464

3465

3466

3467

3468

3469

3470

3471

3472

3473

3474

3475

3476

3477

3478

3479

3480

3481

3482

3483

3484

3485

3486

3487

3488

3489

3490

3491

3492

3493

3494

3495

3496

3497

3498

3499

3500

3501

3502

3503

3504

3505

3506

3507

3508

3509

✓ Task 2-5: Correct Case

问题: 下列哪种情形构成“违反保密义务”?

Question: Which of the following circumstances constitutes "violation of the duty of confidentiality"?

A. 员工离职后使用自身记忆中的技术诀窍  
A. Employees use the technical know-how in their memory after leaving the company.

B. 合作方将约定保密的信息在学术论文中发表  
B. The cooperating party will publish the confidential information as agreed in the academic paper.

C. 根据法院要求披露商业秘密  
C. Disclose trade secrets as required by the court.

D. 向税务机关提交包含商业秘密的审计报告  
D. Submit an audit report containing trade secrets to the tax authority.

**gpt-4o-mini:** ✓

In this question, we need to determine which of the given circumstances constitutes a "violation of the duty of confidentiality".

Analysis:

A. Memories in an employee's mind are generally not subject to legal protection in the context of confidentiality obligations. There is no improper disclosure or use of protected confidential information in a way that breaches an agreement or legal duty.

B. This clearly constitutes a violation of the duty of confidentiality. The cooperating party has an obligation under the confidentiality agreement not to disclose the confidential information. Publishing it in an academic paper goes against this agreement, which is a direct breach of the confidentiality duty.

C. When the court requires the disclosure of trade secrets, it is a legally-mandated act. The obligation to comply with court orders supersedes the normal confidentiality obligations in such cases.

D. Submitting the audit report to the tax authority is a compliant action required by law. It is not an unauthorized or improper disclosure of confidential information.

Answer: B. The cooperating party will publish the confidential information as agreed in the academic paper.

Figure 57: Correct case of task 2-5.

✗ Chinese ✓ Task 2-4: Error Case ✗ English

问题: 某机械专利被无效宣告, 但无效决定显示被诉侵权方实际使用了专利的改进技术, 此时原专利权价值如何变化?

Question: A certain mechanical patent was declared invalid, but the invalidation decision indicated that the accused infringer actually used the improved technology of the patent. How would the value of the original patent right change at this time?

A. 完全归零  
A. Completely reset to zero.

B. 转化为商业秘密增值  
B. Transform into trade secrets to increase value.

C. 可通过反垄断诉讼获得部分赔偿  
C. Partial compensation can be obtained through anti-monopoly litigation.

D. 需重新缴纳年费恢复  
D. The annual fee needs to be paid again for restoration.

**gpt-4o-mini:** ✗

C. 可通过反垄断诉讼获得部分赔偿  
C. Partial compensation can be obtained through anti-monopoly litigation.

**Error Category: Consistency Error and Reasoning error**

**Error Reason:** In this case, although the original patent right was invalidated, the value of the original patent does not completely drop to zero because the accused infringer is actually using an improved version of the patented technology. On the contrary, the original patent right may be transformed into a trade secret, thus retaining some value to a certain extent. Therefore, the most appropriate option is B. ✗

**Question:** A pharmaceutical patent for a blockbuster drug expires in 6 months, with generic competitors already filing ANDA applications. How will its value likely change?

A. Spike due to final monopoly pricing.  
B. Drop to near-zero immediately.  
C. Retain residual value from manufacturing know-how.  
D. Increase due to compulsory licensing.

**gpt-4o-mini:** ✗

B. Drop to near-zero immediately.

**Error Category: Consistency Error and Reasoning error**

**Error Reason:** When a pharmaceutical patent for a blockbuster drug is about to expire, the value of the patent typically decreases as generic competitors are poised to enter the market. The expiration of the patent means that the company will lose its exclusive rights to sell the drug, allowing generic manufacturers to produce and sell their versions, which usually leads to a significant drop in the original drug's market price and revenue. Let's analyze the options: A. Spike due to final monopoly pricing - This is unlikely because the market anticipates the entry of generics, which would drive prices down rather than up. B. Drop to near-zero immediately - While the value will decrease significantly, it may not drop to near-zero immediately, as there could still be some residual value in the brand or remaining market share. C. Retain residual value from manufacturing know-how - This is a possibility, as the original company may still have valuable expertise and processes that could be monetized, but it does not directly address the patent's value. D. Increase due to compulsory licensing - Compulsory licensing typically occurs in specific circumstances, such as public health emergencies, and is not a common outcome for expiring patents. Considering these points, the most accurate answer is that the value of the patent will likely drop significantly as generics enter the market, but it may not drop to zero immediately. Therefore, the best option that reflects the situation is: Answer: C ✓

Figure 58: Error case of task 2-4.

3510  
3511  
3512  
3513  
3514  
3515  
3516  
3517  
3518  
3519  
3520  
3521  
3522  
3523  
3524  
3525  
3526  
3527  
3528  
3529  
3530  
3531  
3532  
3533  
3534  
3535  
3536  
3537  
3538  
3539  
3540  
3541  
3542  
3543  
3544  
3545  
3546  
3547  
3548  
3549  
3550  
3551  
3552  
3553  
3554  
3555  
3556  
3557  
3558  
3559  
3560  
3561  
3562  
3563

X **Task 2-5: Error Case**

问题:1999年, AB齿轮公司(私营)为获取国有CD齿轮集团的商业秘密,通过退休职工马某介绍,贿赂CD集团工程师张某、技术科科长姚某,多次窃取已采取保密措施的齿轮箱图纸(型号包括E750、F300等),用于生产并造成CD集团经济损失377.6万元。经鉴定,涉案图纸重置成本价值51.7万元。AB公司被控单位行贿罪和侵犯商业秘密罪,法定代表人张一奇自首,马某、张某、姚某分别被控介绍贿赂罪和受贿罪。本案商业秘密保护的核心漏洞是?

Question: In 1999, AB Gear Company (a private company) obtained commercial secrets from the state-owned CD Gear Group through retired employee Mr. Ma, who introduced bribery to CD Group engineers Mr. Zhang and Mr. Yao, the head of the technical department. They repeatedly stole confidential gearbox blueprints (including models such as E750, F300, etc.) for production, causing CD Group an economic loss of 3.776 million yuan. After evaluation, the reset cost value of the involved blueprints was 517,000 yuan. AB Company was charged with bribery and infringement of commercial secrets, and the legal representative, Zhang Yiqi, turned himself in. Mr. Ma, Mr. Zhang, and Mr. Yao were respectively charged with introducing bribery and accepting bribes. What is the core loophole in the protection of commercial secrets in this case?

A.图纸具有经济价值  
B.图纸未被公开  
C.通过贿赂手段获取  
D.造成重大损失  
D. Cause significant losses.

**gpt-4o-mini:** X  
D.参加复审案件口头审理的每方当事人及其代理人的人数不得超过四人  
D.The number of each party and their representatives participating in the oral hearing of a reexamination case shall not exceed four.

**Error Category:** Halucination Error and Priority error ✓

**Error Reason:** In this case, AB Gear Company obtained the commercial secrets of CD Gear Group through bribery, and these blueprints were protected by confidentiality measures and not made public. Therefore, the core loophole in the protection of commercial secrets lies in the fact that the blueprints were not made public. While the economic value of the blueprints, the significant losses caused, and the bribery used to obtain them are all important factors, the most direct loophole is that the blueprints were not made public. Answer: B X

Figure 59: Error case of task 2-5.

✓ **Chinese** ✓ **Task 3-1: Correct Case** ✓ **English**

问题:优先权文本记载:“支架的轴向弹性模量为10~20GPa”,中国申请修改为“径向刚度刚度 $\geq 15GPa$ ”,说明书新增测试方法显示径向刚度与轴向模量存在非线性关系。该修改的隐蔽错误是?

Question: The priority text states that "the axial elastic modulus of the bracket is 10-20GPa", and the Chinese application is modified to "the radial compressive stiffness is greater than or equal to 15GPa". The new test method in the specification shows that there is a nonlinear relationship between radial stiffness and axial modulus. What is the hidden error of this modification?

A.轴向模量与径向刚度的物理量纲不同。  
A. The physical dimensions of axial modulus and radial stiffness are different.  
B.优先权文件未提供径向刚度测试数据。  
B. The priority document does not provide radial stiffness test data.  
C.15GPa阈值与原始参数范围无直接对应关系。  
C. There is no direct correlation between the 15GPa threshold and the original parameter range.  
D.未说明测试方法的边界条件。  
D. The boundary conditions for the test method are not specified.

**gpt-4o-mini:**  
C.15GPa阈值与原始参数范围无直接对应关系。  
C. There is no direct correlation between the 15GPa threshold and the original parameter range.

**Response:**  
A. The physical dimensions of axial modulus and radial stiffness differ. This option is correct because axial modulus and radial stiffness are indeed different physical quantities, although there may be some relationship between them, their units and dimensions are distinct.

B: The priority document does not provide radial stiffness test data. This option might be a hidden error, but it does not directly indicate the error in the modification.

C: There is no direct correspondence between the 15GPa threshold and the original parameter range. This option highlights a potential issue, that is, the modified parameters may not align with the original parameters, but it does not directly address the error in the physical quantity.

D: The boundary conditions of the test method are not specified. This option could also be a hidden error, as the boundary conditions of the test method may affect the test results, but it does not directly involve an error in the physical quantity.

Question:Priority doc shows "battery capacity retention  $\geq 80\%$  after 500 cycles at 1C." Continuation claims: " $\geq 90\%$  retention after 300 cycles at 2C." Spec only tests 1C cycling. Core deception?

A.C-rate doubling accelerates degradation nonlinearly.  
B. Cycle count reduction invalidates long-term performance.  
C. Capacity retention depends on voltage cutoff.  
D.No data supports 2C rate performance.

**gpt-4o-mini:** ✓  
D.No data supports 2C rate performance.

**Response:**  
A. C-rate doubling accelerates degradation nonlinearly: This option suggests that moving from 1C to 2C would lead to a significant increase in degradation, which is a valid concern in battery performance.  
B. Cycle count reduction invalidates long-term performance: This option implies that fewer cycles (300 at 2C vs. 500 at 1C) may not provide a reliable measure of long-term performance, which is also a valid point.  
C. Capacity retention depends on voltage cutoff: While voltage cutoff can affect capacity retention, it does not directly address the core issue of comparing performance at different C-rates and cycle counts.  
D. No data supports 2C rate performance: This option points out that the data provided does not support the claims made at the 2C rate, which is a critical observation.

Figure 60: Correct case of task 3-1.

3564

3565

3566

3567

3568

3569

3570

3571

3572

3573

3574

3575

3576

3577

3578

3579

3580

3581

3582

3583

3584

3585

3586

3587

3588

3589

3590

3591

3592

3593

3594

3595

3596

3597

3598

3599

3600

3601

3602

3603

3604

3605

3606

3607

3608

3609

3610

3611

3612

3613

3614

3615

3616

3617

**Task 3-1: Error Case**

**Chinese** **English**

**Question:** The priority document states: "Gene editing efficiency  $\geq 70\%$ ," while the Chinese application modifies it to: "Knockout efficiency in CHO-K1 cells  $\geq 80\%$ ." It is known that Example 1 using HEK293 cells achieved an efficiency of 75%, and Example 2 using CHO-K1 cells achieved an efficiency of 65%. What is the biological pitfall of this modification?

**gpt-4-mini:**  A. 修复后细胞类型特异性排除原始数据支持。  
A. The modified cell type-specific exclusion of the original data is supported.

**gpt-4-mini:**  A. 优先权未记载CHO-K1细胞的具体参数。  
C. The specific parameters of CHO-K1 cells are not recorded in the priority.

**gpt-4-mini:**  A. 不同细胞系的编辑机制一致性。  
D. The consistency of editing mechanisms in different cell lines has not been verified.

**gpt-4-mini:**  A. 修复后细胞类型特异性排除原始数据支持。  
A. The modified cell type-specific exclusion of the original data is supported.

**Error Category: Reasoning error**  
**Error Reason:**

A. The modified cell type-specific exclusion of original data supports : This option is reasonable because the modified requirement is for CHO-K1 cells, and the original data (75% efficiency of HEK293 cells) does not support this specific cell type requirement.  
B. Efficiency improvement lacks evidence of transfection condition optimization: The mention of transfection condition optimization in this option is not related to the issue as it was not mentioned in the problem.  
C. Priority is not recorded for specific parameters of CHO-K1 cells: Although the priority text does not specify parameters, the core of the issue lies in comparing efficiencies rather than the lack of parameters.  
D. The consistency of editing mechanisms across different cell lines has not been verified: While editing mechanisms may differ across cell lines, the key to the issue is comparing efficiencies, not the consistency of mechanisms.

**Question:** The description of the invention's solution for creating an external metal-to-metal seal between the pin and box connection of a threaded tubular, it explains that a separate metal seal base is added to the external surface of the pin. It further elaborates on how this allows for the repair or replacement of the seal if damaged. When evaluating the clarity and sufficiency of this description of the invention's key feature, which option is correct?

**gpt-4-mini:**  A. The description is incomplete as it doesn't mention how the added metal seal base affects the overall weight and strength of the pipeline connection.  
B. There are unclear claims regarding the durability of the new seal formed with the separate metal seal base as compared to the prior art seals.  
C. The description clearly conveys the concept and functionality of the invention's key feature, and there is no error in the way it is presented.  
D. The description is incorrect because it assumes that the repair process of the seal is always feasible without considering potential manufacturing tolerances.

**gpt-4-mini:**  C. The description clearly conveys the concept and functionality of the invention's key feature, and there is no error in the way it is presented.

**Error Category: Reasoning error**  
**Error Reason:**

A. This option suggests that the description is incomplete because it does not address the impact of the added metal seal base on the overall weight and strength of the pipeline connection. While these factors are important in a comprehensive evaluation of the invention, the question specifically asks about the clarity and sufficiency of the description of the key feature, which is the addition of the metal seal base itself.  
B. This option points out that there are unclear claims regarding the durability of the new seal compared to prior art seals. The question does not indicate that durability is a key feature of the description, so this option may not be the most relevant.  
C. This option states that the description clearly conveys the concept and functionality of the invention's key feature, indicating that there is no error in the presentation. If the description effectively communicates how the separate metal seal base functions and its benefits, this option could be valid.  
D. This option claims that the description is incorrect because it assumes the repair process is always feasible without considering manufacturing tolerances. It critiques an assumption rather than the clarity of the description.

Figure 61: Error case of task 3-1.

**Task 3-2: Correct Case**

**Chinese** **English**

**Question:** The alloy ratio of an orthopedic screw patent is exactly the same as the formula record of a customer's customized product (not patented) from a smelter 15 years ago. What is missing from this patent?

**gpt-4-mini:**  A. 充分公开  
A. Full disclosure  
B. 实用性  
B. Practicality  
C. 创造性  
C. Creativeness  
D. 新颖性  
D. Novelty

**gpt-4-mini:**  D. 新颖性  
D. Novelty

**Response:**

A. Full Disclosure: The patent application must detail the invention in sufficient detail so that a person skilled in the relevant field can understand and implement it.  
B. Practicality: The invention must have the potential for practical application.  
C. Creativity: The invention must represent a significant advancement over existing technology and cannot be obvious.  
D. Novelty: The invention must be new and not already disclosed.

**Question:** An AI startup patented a neural network architecture in 2022. Evidence shows the same architecture was fully implemented in an open-source GitHub repository with 500+ stars 14 months pre-filing. What invalidates the patent?

**gpt-4-mini:**  A. Novelty  
B. Non-obviousness  
C. Utility  
D. Enablement

**gpt-4-mini:**  A. Novelty

**Response:**

A. Novelty: For a patent to be granted, the invention must be novel, meaning it must not have been publicly disclosed before the patent application was filed. In this case, since the same architecture was fully implemented in an open-source GitHub repository 14 months before the patent filing, it indicates that the architecture was already known to the public. This directly challenges the novelty requirement.  
B. Non-obviousness: This criterion assesses whether the invention is sufficiently inventive or if it would have been obvious to someone skilled in the field at the time of the patent application. While the prior art (the GitHub repository) could also affect this, the primary issue here is the lack of novelty.  
C. Utility: This requirement states that the invention must have a specific, substantial, and credible utility. The information provided does not suggest that the architecture lacks utility, so this is not a relevant factor in invalidating the patent.  
D. Enablement: This requirement means that the patent application must provide enough detail for someone skilled in the art to make and use the invention. The existence of the open-source implementation suggests that the architecture is indeed enabled, so this is not a reason for invalidation.

Figure 62: Correct case of task 3-2.

3618

3619

3620

3621

3622

3623

3624

3625

3626

3627

3628

3629

3630

3631

3632

3633

3634

3635

3636

3637

3638

3639

3640

3641

3642

3643

3644

3645

3646

3647

3648

3649

3650

3651

3652

3653

3654

3655

3656

3657

3658

3659

3660

3661

3662

3663

3664

3665

3666

3667

3668

3669

3670

3671

X Chinese Task 3-2: Error Case English

问题:某手术机器人专利的力反馈算法,被证实与某国际会议3年前公开的未申请专利的演示视频技术方案相同。该专利无效的核心问题是?

Question: The force feedback algorithm patented by a surgical robot has been proved to be the same as the unpatented demonstration video technology scheme disclosed at an international conference three years ago. What is the core problem of the invalidation of this patent?

A.技术启示  
A. Technological implications  
B.工业适用性  
B. Industrial applicability  
C.附图说明缺陷  
C. Illustrative defects in the drawings  
D.新颖性  
D. Novelty

gpt-4o-mini: ✗  
D.新颖性  
D. Novelty

**Error Category: Reasoning error**  
**Error Reason:** To determine the core issue of invalidating a surgical robot patent, we need to analyze the relationship between options and the reasons for patent invalidation.

A. Technical Inspiration: This usually refers to the source of technical inspiration or insight, which may be related to the validity of the patent but is not the direct core issue leading to its invalidation. ✓

B. Industrial Applicability: This refers to whether the invention can be applied in industry. Even if a technology is theoretically feasible, it may still lead to patent invalidation if it cannot be practically applied. However, the question does not mention the applicability of the technology. ✗

C. Defects in Drawings and Descriptions: This involves the clarity and completeness of drawings and descriptions in the patent document. If the drawings or descriptions are unclear, it may result in patent invalidation, but this point is not mentioned in the question.

D. Novelty: Novelty refers to whether the invention is new and has not been disclosed before the application for the patent. The question mentions that the force feedback algorithm of the patent is identical to an unpatented demonstration video technology from three years ago, which directly addresses the issue of novelty.

Question: A patent claims a "new" drug dosage but admits it causes severe side effects in 90% of patients. What is the rejection reason?

A. Non-obviousness  
A. Non-obviousness  
B. Ethical issues  
B. Ethical issues  
C. Enablement failure  
C. Enablement failure  
D. Lack of utility  
D. Lack of utility

**Error Category: Reasoning error**  
**Error Reason:** A. This criterion assesses whether the invention is sufficiently different from existing knowledge or products. If the dosage is truly new and not obvious to someone skilled in the field, this might not be the rejection reason.

B. While the severe side effects could raise ethical concerns, patent law primarily focuses on the technical aspects of the invention rather than ethical implications. Therefore, this is unlikely to be the primary reason for rejection. ✗

C. This criterion requires that the patent application must provide enough information for someone skilled in the art to make and use the invention. If the dosage causes severe side effects, it might not be considered adequately enabled, but this is more about the practical application rather than the inherent utility of the drug. ✗

D. This is the most relevant option. A patent must demonstrate that the invention has a specific, substantial, and credible utility. If a drug dosage causes severe side effects in 90% of patients, it may not be considered useful or beneficial, leading to a rejection based on lack of utility.

Figure 63: Error case of task 3-2.

✓ Chinese Task 3-3: Correct Case English

问题:请从A, B, C, D四个选项中选出与下述专利最相似的专利序号,该序号是?一种刷头及吸尘器。刷头包括:连接管(1),一端用于连接电源,另一端为吸气口……(后略)

Question: Please select the patent number that is most similar to the following patent from the four options A, B, C and D. What is the number? A brush head and vacuum cleaner. The brush head includes: a connecting tube (1), one end for connecting to a negative pressure source, the other end is a suction port...(omit)

A.一种多功能洁面仪,包括外壳(110)、刷头组件、电机(130)与电源组件,外壳(110)内设有电源组件、电机(130)与支撑电机(130)的支架……(后略)

B. A multi-functional facial cleanser, including a housing (110), a brush head assembly, a motor (130) and a power supply assembly. The housing (110) is provided with a power supply assembly, a motor (130) and a bracket supporting the motor (130)...(omit)

C.一种具有吸尘功能的宠物毛刷及其使用方法,适用于宠物护理设备领域……(后略)

D. A pet brush with hair absorption function and its use method, suitable for pet care equipment field...(omit)

A.一种多功能洁面仪,包括外壳(110)、刷头组件、电机(130)与电源组件,外壳(110)内设有电源组件、电机(130)与支撑电机(130)的支架……(后略)

B. A pet brush with hair absorption function and its use method, suitable for pet care equipment field...(omit)

C.一种吸水刷头,包括刷头壳体(1),所述刷头壳体(1)具有与真空源连通以产生吸引流的吸入通道……(后略)

D.本发明公开了一种多功能化妆刷,包括第一刷头、第二刷头、第三刷头……(后略)

D. The invention discloses a multi-functional makeup brush, including a first brush head, a second brush head, a third brush head...(omit)

gpt-4o-mini: ✗  
C.一种吸水刷头,包括刷头壳体(1),所述刷头壳体(1)具有与真空源连通以产生吸引流的吸入通道……(后略)

C. A water-absorbing brush head, including a brush head housing (1), the brush head housing (1) having an intake channel connected to a vacuum source for generating suction flow...(omit)

Question:Please select the most similar patent number from A, B, C and D. Which number is?Provided is an adipose tissue preservation solution, comprising: polyethylene glycol 400, human albumin, norfloxacin, low molecular dextran, and Ac-DEVD-CHO. The present invention uses the polyethylene glycol 400 as a base component for the first time; said substance has good compatibility with an oil-type substance, and may also separate out a lipid-type substance after being compatible with a water-type substance; and human albumin is used as a nutritional component of adipose mesenchymal stem cell preservation solution...(omit)

A. Provided in the present invention are a mesenchymal stem cell injection, a preparation method thereof and the use in preparing drugs for treating dilated cardiomyopathy in children...(omit)

B. The present invention provides a mesenchymal stem cell injection, a preparation method thereof, and application thereof in preparing a drug for treating ulcerative colitis...(omit)

C. Provided in the present invention is a cell freezing medium for clinical use. In particular, the cell freezing medium of the present invention comprises the following components...(omit)

D. The present invention provides a mesenchymal stem cell injection, a preparation method thereof, and application thereof in preparing a diabetes drug...(omit)

gpt-4o-mini: ✗  
D. The present invention provides a mesenchymal stem cell injection, a preparation method thereof, and application thereof in preparing a diabetes drug...(omit)

Figure 64: Correct case of task 3-3.

3672

3673

3674  Chinese Task 3-3: Error Case English

3675

3676 问题: 请从A, B, C, D四个选项中选出与下述专利最相似的专利序号, 该序号是本申请适于图像处理技术领域, 提供了一种脑萎缩程度的定量检测方法、检测装置及终端设备, 包括…… (后略)

3677 Question: Please select the patent number that is most similar to the following patent from the four options A, B, C and D. What is the patent number? This application is applicable to the field of image processing technology, and provides a quantitative detection method, detection device and terminal equipment for brain atrophy degree, including... (omit)

3678 A.本发明公开了一种图像特征的提取方法及装置…… (后略)

3679 A. The invention discloses an image feature extraction method and device... (omit)

3680 B.一种病变部位识别方法、识别装置、计算机装置及…… (后略)

3681 B. A method for identifying a lesion site... (omit)

3682 C.本申请适用于图像处理技术领域, 提供了一种…… (后略)

3683 C. This application is applicable to the field of image processing technology and provides a method... (omit)

3684 D.本申请公开了一种目标检测方法和装置…… (后略)

3685 D. D: This application discloses a target detection method and device... (omit)

3686

3687 gpt-4o-mini: ✘

3688 B.一种病变部位识别方法、识别装置、计算机装置及…… (后略)

3689 B. A method for identifying a lesion site... (omit)

3690

3691

3692

3693

3694

3695

3696

3697

3698

3699

3700

3701  Chinese Task 3-4: Correct Case English

3702

3703 问题: 丙公司技术员张某, 完成本职工作之余, 利用业余时间、自行购买材料研发“新型焊接工具”, 未利用丙公司物质技术条件, 公司与张某也未约定权属。该工具的专利申请权属于?

3704 Question: Zhang, a technician of Company C, developed a "new welding tool" by himself using spare time and materials purchased by himself after completing his own work, without using the material and technical conditions of Company C, and the company and Zhang did not agree on ownership. Who owns the patent application right of the tool?

3705 A.丙公司, 因张某是公司员工。

3706 A. Company C, because Zhang is an employee of the company.

3707 B.张某, 因非职务发明创造。

3708 B. Zhang, because it is not an invention made in the course of his duties.

3709 C.双方共有, 因张某员工身份关联。

3710 C. The two parties share the same identity due to Zhang's employee status.

3711 D.归当地政府, 鼓励创新。

3712 D. Return to the local government and encourage innovation.

3713

3714 gpt-4o-mini: ✘

3715 B.张某, 因非职务发明创造

3716 B.Zhang, because it is not an invention made in the course of his duties

3717

3718

3719

3720

3721

3722

3723

3724

3725

3726

3727

3728

3729

3730

3731

3732

3733

3734

3735

3736

3737

3738

3739

3740

3741

3742

3743

3744

3745

3746

3747

3748

3749

3750

3751

3752

3753

3754

3755

3756

3757

3758

3759

3760

3761

3762

3763

3764

3765

3766

3767

3768

3769

3770

3771

3772

3773

3774

3775

3776

3777

3778

3779

3780

3781

3782

3783

3784

3785

3786

3787

3788

3789

3790

3791

3792

3793

3794

3795

3796

3797

3798

3799

3800

3801

3802

3803

3804

3805

3806

3807

3808

3809

3810

3811

3812

3813

3814

3815

3816

3817

3818

3819

3820

3821

3822

3823

3824

3825

3826

3827

3828

3829

3830

3831

3832

3833

3834

3835

3836

3837

3838

3839

3840

3841

3842

3843

3844

3845

3846

3847

3848

3849

3850

3851

3852

3853

3854

3855

3856

3857

3858

3859

3860

3861

3862

3863

3864

3865

3866

3867

3868

3869

3870

3871

3872

3873

3874

3875

3876

3877

3878

3879

3880

3881

3882

3883

3884

3885

3886

3887

3888

3889

3890

3891

3892

3893

3894

3895

3896

3897

3898

3899

3900

3901

3902

3903

3904

3905

3906

3907

3908

3909

3910

3911

3912

3913

3914

3915

3916

3917

3918

3919

3920

3921

3922

3923

3924

3925

3926

3927

3928

3929

3930

3931

3932

3933

3934

3935

3936

3937

3938

3939

3940

3941

3942

3943

3944

3945

3946

3947

3948

3949

3950

3951

3952

3953

3954

3955

3956

3957

3958

3959

3960

3961

3962

3963

3964

3965

3966

3967

3968

3969

3970

3971

3972

3973

3974

3975

3976

3977

3978

3979

3980

3981

3982

3983

3984

3985

3986

3987

3988

3989

3990

3991

3992

3993

3994

3995

3996

3997

3998

3999

4000

4001

4002

4003

4004

4005

4006

4007

4008

4009

4010

4011

4012

4013

4014

4015

4016

4017

4018

4019

4020

4021

4022

4023

4024

4025

4026

4027

4028

4029

4030

4031

4032

4033

4034

4035

4036

4037

4038

4039

4040

4041

4042

4043

4044

4045

4046

4047

4048

4049

4050

4051

4052

4053

4054

4055

4056

4057

4058

4059

4060

4061

4062

4063

4064

4065

4066

4067

4068

4069

4070

4071

4072

4073

4074

4075

4076

4077

4078

4079

4080

4081

4082

4083

4084

4085

4086

4087

4088

4089

4090

4091

4092

4093

4094

4095

4096

4097

4098

4099

4100

4101

4102

4103

4104

4105

4106

4107

4108

4109

4110

4111

4112

4113

4114

4115

4116

4117

4118

4119

4120

4121

4122

4123

4124

4125

4126

4127

4128

4129

4130

4131

4132

4133

4134

4135

4136

4137

4138

4139

4140

4141

4142

4143

4144

4145

4146

4147

4148

4149

4150

4151

4152

4153

4154

4155

4156

4157

4158

4159

4160

4161

4162

4163

4164

4165

4166

4167

4168

4169

4170

4171

4172

4173

4174

4175

4176

4177

4178

4179

4180

4181

4182

4183

4184

4185

4186

4187

4188

4189

4190

4191

4192

4193

4194

4195

4196

4197

4198

4199

4200

4201

4202

4203

4204

4205

4206

4207

4208

4209

4210

4211

4212

4213

4214

4215

4216

4217

4218

4219

4220

4221

4222

4223

4224

4225

4226

4227

4228

4229

4230

4231

4232

4233

4234

4235

4236

4237

4238

4239

4240

4241

4242

4243

4244

4245

4246

4247

4248

4249

4250

4251

4252

4253

4254

4255

4256

4257

4258

4259

4260

4261

4262

4263

4264

4265

4266

4267

4268

4269

4270

4271

4272

4273

4274

4275

4276

4277

4278

4279

4280

4281

4282

4283

4284

4285

4286

4287

4288

4289

4290

4291

4292

4293

4294

4295

4296

4297

4298

4299

4300

4301

4302

4303

4304

4305

4306

4307

4308

4309

4310

4311

4312

4313

4314

4315

4316

4317

4318

4319

4320

4321

4322

4323

4324

4325

4326

4327

4328

4329

4330

4331

4332

4333

4334

4335

4336

4337

4338

4339

4340

4341

4342

4343

4344

4345

4346

4347

4348

4349

4350

4351

4352

4353

4354

4355

4356

4357

4358

4359

4360

4361

4362

4363

4364

4365

4366

4367

4368

4369

4370

4371

4372

4373

4374

4375

4376

4377

4378

4379

4380

4381

4382

4383

4384

4385

4386

4387

4388

4389

4390

4391

4392

4393

4394

4395

4396

4397

4398

4399

4400

4401

4402

4403

4404

4405

4406

4407

4408

4409

4410

4411

4412

4413

4414

4415

4416

4417

4418

4419

4420

4421

4422

4423

4424

4425

4426

4427

4428

4429

4430

4431

4432

4433

4434

4435

4436

4437

4438

4439

4440

4441

4442

4443

4444

4445

4446

4447

4448

4449

4450

4451

4452

4453

4454

4455

4456

4457

4458

4459

4460

4461

4462

4463

4464

4465

4466

4467

4468

4469

4470

4471

4472

4473

4474

4475

4476

4477

4478

4479

4480

4481

4482

4483

4484

4485

4486

4487

4488

4489

4490

4491

4492

4493

4494

4495

4496

4497

4498

4499

4500

4501

4502

4503

4504

4505

4506

4507

4508

4509

4510

4511

4512

4513

4514

4515

4516

4517

4518

4519

4520

4521

4522

4523

4524

4525

4526

4527

4528

4529

4530

4531

4532

4533

4534

4535

4536

4537

4538

4539

4540

4541

4542

4543

4544

4545

4546

4547

4548

4549

4550

4551

4552

4553

4554

4555

4556

4557

4558

4559

4560

4561

4562

4563

4564

4565

4566

4567

4568

4569

4570

4571

4572

4573

4574

4575

4576

4577

4578

4579

4580

4581

4582

4583

4584

4585

4586

4587

4588

4589

4590

4591

4592

4593

4594

4595

4596

4597

4598

4599

4600

4601

4602

4603

4604

4605

4606

4607

4608

4609

4610

4611

4612

4613

4614

4615

4616

4617

4618

4619

4620

4621

4622

4623

4624

4625

4626

4627

4628

4629

4630

4631

4632

4633

4634

4635

4636

4637

4638

4639

4640

4641

4642

4643

4644

4645

4646

4647

4648

4649

4650

4651

4652

4653

4654

4655

4656

4657

4658

4659

4660

4661

4662

4663

4664

4665

4666

4667

4668

4669

4670

4671

4672

4673

4674

4675

4676

4677

4678

4679

4680

4681

4682

4683

4684

4685

4686

4687

4688

4689

4690

4691

4692

4693

4694

4695

4696

4697

4698

4699

4700

4701

4702

4703

4704

4705

4706

4707

4708

4709

4710

4711

4712

4713

4714

4715

4716

4717

4718

4719

4720

4721

4722

4723

4724

4725

4726

4727

4728

4729

4730

4731

4732

4733

4734

4735

4736

4737

4738

4739

4740

4741

4742

4743

4744

4745

4746

4747

4748

4749

4750

4751

4752

4753

4754

4755

4756

4757

4758

4759

4760

4761

4762

4763

4764

4765

4766

4767

4768

4769

4770

4771

4772

4773

4774

4775

4776

4777

4778

4779

4780

4781

4782

4783

4784

4785

4786

4787

4788

4789

4790

4791

4792

4793

4794

4795

4796

4797

4798

4799

4800

4801

4802

4803

4804

4805

4806

4807

4808

4809

4810

4811

4812

4813

4814

4815

4816

4817

4818

4819

4820

4821

4822

4823

4824

4825

4826

4827

4828

4829

4830

4831

4832

4833

4834

4835

4836

4837

4838

4839

4840

4841

4842

4843

4844

4845

4846

4847

4848

4849

4850

4851

4852

4853

4854

4855

4856

4857

48

3726

3727

3728

3729

3730

3731

3732

3733

3734

3735

3736

3737

3738

3739

3740

3741

3742

3743

3744

3745

3746

3747

3748

3749

3750

3751

3752

3753

3754

3755

3756

3757

3758

3759

3760

3761

3762

3763

3764

3765

3766

3767

3768

3769

3770

3771

3772

3773

3774

3775

3776

3777

3778

3779

✓ Task 3-5: Correct Case

# Patent Applications Awaiting Examination

```
<ApplicationNumber> 12763093 </ApplicationNumber>
<Title> METHODS AND APPARATUS FOR ENSURING
COMPATIBILITY ON A HIGH PERFORMANCE SERIAL BUS
</Title>
<Abstract> A data communications system is disclosed having at least
one Legacy cloud coupled to at least one Beta cloud. The system further
having at least one BOSS node and at least one border node. A method
for ensuring compatibility is disclosed comprising determining when the
BOSS node is idle, determining whether the last packet transmitted by
any border node was an Alpha format packet if the BOSS node is idle,
and unlocking the Legacy cloud if the last packet transmitted by the
border node was not an Alpha format packet. </Abstract>
<Background> <SOI> BACKGROUND OF THE INVENTION
<EOH>1. Field of the Invention The present invention relates to data
communications...(omit)
```

A. Allowed

B. Rejected

gpt-4o-mini: ✓

To determine whether the patent application should be allowed or rejected, we need to analyze the content of the application, including the title, abstract, background, claims, and summary.

Analysis:

1. Title: "METHODS AND APPARATUS FOR ENSURING COMPATIBILITY ON A HIGH PERFORMANCE SERIAL BUS" The title suggests a focus on compatibility in data communication systems, which is a relevant and important area in technology.
2. Abstract: The abstract describes a system that ensures compatibility between different types of data packets in a high-performance serial bus system. It outlines a method involving a BOSS node and border nodes to manage packet transmission and compatibility.
3. Background: The background provides context about the need for compatibility in data communication systems, particularly with the evolution of standards like IEEE 1394-1995 and its successors...(omit)

Answer: allowed

Figure 67: Correct case of task 3-5.

3757

3758

3759

3760

3761

3762

3763

3764

3765

3766

3767

3768

3769

3770

3771

3772

3773

3774

3775

3776

3777

3778

3779

✗ Chinese Task 3-4: Error Case English

問題: 我公司員工孫某的下列發明中, 哪個不屬於職務發明?

Question: Which of the following inventions made by Sun, an employee of Company V, is not a service invention?

A. 利用單位保密技術資料研发的 “量子通信設備”。

A. "Quantum communication equipment" developed using unit confidential technical data.

B. 执行單位任務研发的 “人臉識別系統”。

B. The "face recognition system" developed by the executing unit to carry out its tasks.

C. 留職後 1 年內研发的 “新能電池”。

C. "New energy battery" developed within one year after leaving the job.

D. 主要利用單位資金研发的 “智能穿戴設備”。

D. "Smart wearable devices" developed mainly with unit funds.

gpt-4o-mini: ✗

C. "New energy battery" developed within one year after leaving the job.

Error Category: **Consistency error**

Error Reason:

When analyzing this issue, we need to understand the definition of "service invention". A service invention refers to an invention created by an employee while performing their duties, typically closely related to their job responsibilities, tasks, or resources of the organization.

A. "Quantum communication equipment" developed using the organization's confidential technical data- This is a service invention because it utilizes the organization's confidential technical data.

B. "Facial recognition system" developed while performing organizational tasks- This is also a service invention because it was developed while performing organizational tasks.

C. "New energy battery" developed within one year after leaving the organization (not related to previous work) - This does not qualify as a service invention because it was developed after leaving the organization and is unrelated to previous work. ✓

D. "Smart wearable device" primarily funded by the organization- This is a service invention because it mainly utilized organizational funds.

Question: Amgen had an ongoing research project on PCSK9 - binding antibodies. A researcher, Dr. Smith, who was part of this project, left Amgen and joined a competitor. Before leaving, Dr. Smith had made some key preliminary discoveries related to a new antibody structure. After joining the competitor, Dr. Smith further developed this antibody structure and the competitor filed for a patent. Amgen claims that it should own the patent rights. Which of the following is the most likely basis for Amgen's claim?

A. Dr. Smith's work was a continuation of Amgen's research project, and he used Amgen's confidential information.

B. Amgen was the first to start researching PCSK9 - binding antibodies.

C. The competitor had no right to hire Dr. Smith as he was still under a non-compete agreement with Amgen.

D. Amgen had already filed a provisional patent application on a related concept.

gpt-4o-mini: ✗

A. Dr. Smith's work was a continuation of Amgen's research project, and he used Amgen's confidential information.

Error Category: **Priority error**

Error Reason:

A. Dr. Smith's work was a continuation of Amgen's research project, and he used Amgen's confidential information. This option suggests that Dr. Smith's new work was directly related to the research he conducted at Amgen and that he utilized proprietary information from Amgen. If this is the case, Amgen could argue that the patent rights belong to them because the new discoveries were built upon their confidential research.

B. Amgen was the first to start researching PCSK9 - binding antibodies. While being the first to research a topic can establish a claim to the field, it does not automatically grant patent rights to subsequent discoveries made by former employees unless they are directly tied to the original research and confidential information.

C. The competitor had no right to hire Dr. Smith as he was still under a non-compete agreement with Amgen. If Dr. Smith was indeed under a non-compete agreement, Amgen could potentially pursue legal action against both Dr. Smith and the competitor. However, this does not directly address the patent rights issue regarding the antibody structure.

D. Amgen had already filed a provisional patent application on a related concept. Filing a provisional patent application can establish a priority date for an invention, but it would depend on whether the provisional application covers the specific antibody structure developed by Dr. Smith. If it does, Amgen could claim rights based on their prior application. ✗

Figure 68: Error case of task 3-4.

3780

3781

3782

3783

3784

3785

3786

3787

3788

3789

3790

3791

3792

3793

3794

3795

3796

3797

3798

3799

3800

3801

3802

3803

3804

3805

3806

3807

3808

3809

3810

3811

3812

3813

3814

3815

3816

3817

3818

3819

3820

3821

3822

3823

3824

3825

3826

3827

3828

3829

3830

3831

3832

3833

X Task 3-5: Error Case

# Patent Applications Awaiting Examination

<ApplicationNumber> 12765091 </ApplicationNumber>

<Title> METHOD AND DEVICE OF ASSIGNING RING IDENTIFIER

</Title>

<Abstract> A control message having stored therein a link identifier for identifying a link such as a fiber or cable and a ring identifier for identifying a ring (or a control message having stored therein a ring identifier) is transmitted to a neighbor node. A ring identifier preliminarily assigned to a link identified by the link identifier received from the neighbor node (or a ring identifier preliminarily assigned to a link having received the control message) is compared with a ring identifier received from the neighbor node to select either one of the ring identifiers according to specified selection criteria. The ring identifier selected is assigned to the link identified. </Abstract>

<Background> <SOH> BACKGROUND <EOH> A large number of transmission networks have adopted a configuration having a plurality of ring networks R 1 and R 2 ... (omit)

A. Allowed

B. Rejected

gpt-4o-mini: X

B. Rejected

Error Category: Hallucination Error

Error Reason:

- Novelty: The application describes a method and device for assigning ring identifiers in a network, which involves transmitting control messages, comparing identifiers, and selecting based on specified criteria. We need to check if this method is novel compared to existing technologies.
- Non-obviousness: The claims must not be obvious to someone skilled in the art. The application references existing protocols and methods, such as Link Management Protocol (LMP) and other related documents. If the method described is an obvious extension of existing technologies, it may be rejected.
- Utility: The application appears to have a clear utility in managing network resources efficiently, which is a positive aspect for allowing the patent.
- Claims Analysis: The claims detail specific steps and components involved in the method and device. If these claims are sufficiently detailed and not merely abstract ideas or general concepts, they may be allowed.
- Prior Art: The background section references existing technologies and protocols. If similar methods have been patented or published, this could impact the novelty and non-obviousness of the application.

Figure 69: Error case of task 3-5.

✓ Chinese Task 4-1: High-quality Case English

# 权利要求

1. 液压深松铲，包括支撑臂(2)和深松头(3)，深松头(3)安装在支撑臂(2)底端，其特征在于，还包括固定装置、提升装置、套筒(1)和两组破土片(4)，支撑臂(2)上下滑动穿过套筒(1)内部，套筒(1)通过固定装置安装在旋耕机前端，固定装置用于对套筒(1)位置固定…… (后略)

1. Hydraulic deep loosening shovel, including support arm (2) and deep loosening head (3), the deep loosening head (3) is installed at the bottom of the support arm (2); Its features include a fixing device, a lifting device, a sleeve (1), and two sets of soil breaking pieces (4). The support arm (2) slides up and down through the interior of the sleeve (1), and the sleeve (1) is installed at the front end of the rotary tiller through the fixing device. The fixing device is used to fix the position of the sleeve (1), and the lifting device is installed on the fixing device and pushed to the side of the sleeve (1)...

gpt-4o-mini:

本发明涉及一种液压深松铲，主要包括支撑臂和深松头，深松头安装在支撑臂底端。该液压深松铲还包括固定装置、提升装置、套筒和两组破土片。支撑臂上下滑动穿过套筒内部，套筒通过固定装置安装在旋耕机前端，固定装置用于固定套筒位置。提升装置安装在固定装置上并推动设置在套筒侧方，用于带动支撑臂升降移动。两组破土片分别安装在深松头的左右两侧……(后略)

The present invention relates to a hydraulic deep loosening shovel, mainly comprising a support arm and a deep loosening head, wherein the deep loosening head is installed at the bottom end of the support arm. The hydraulic deep loosening shovel also includes a fixing device, a lifting device, a sleeve, and two sets of soil breaking pieces. The support arm slides up and down through the inside of the sleeve, and the sleeve is installed at the front end of the rotary tiller through a fixing device, which is used to fix the position of the sleeve. The lifting device is installed on the fixed device and pushed to the side of the sleeve to drive the support arm to move up and down. Two sets of soil breaking pieces are installed on the left and right sides of the deep loosening head respectively...

# Claims

1. A method for controlling a blind spot detector on a motorcycle, the method comprising: detecting, from lane markings shown in a field of view of a camera of a lane position sensor, a lateral change in a position of the motorcycle, wherein the lateral change comprises the motorcycle moving towards an edge of a lane on a first side of the motorcycle; and adjusting, based on the lateral change detected from the lane markings shown in the field of view of the camera of the lane position sensor, a width of a detection zone of the blind spot detector, wherein: the adjusting the width of the detection zone comprises decreasing the width of the detection zone on the first side of the motorcycle, and the width of the detection zone is further adjusted by steering the blind spot detector.

2. detecting, from lane markings shown in a field of view of a camera of a lane position sensor, a lateral change in a position of the motorcycle, wherein the lateral change comprises the motorcycle moving towards an edge of a lane on a first side of the motorcycle... (omit)

gpt-4o-mini:

The present invention relates to a method and system for enhancing blind spot detection on motorcycles. The method involves detecting lateral changes in the motorcycle's position within its lane using a lane position sensor equipped with a camera. Based on this detection, the width of the blind spot detection zones on either side of the motorcycle is dynamically adjusted. Specifically, when the motorcycle moves closer to the edge of a lane, the detection zone on that side is narrowed, while the opposite side's detection zone is correspondingly widened. The system includes one or more sensors to monitor these zones, a controller to process the lane markings and adjust the detection zones in real-time, and a signaling device to alert the rider of any vehicles detected within these zones. The alerts can be provided through haptic, audible, or visual signals, ensuring the rider is informed of potential hazards in their blind spots. This innovative approach aims to improve rider safety by providing timely and accurate awareness of surrounding vehicles.

Figure 70: High-quality case of task 4-1.

3834	Chinese	Task 4-1: Low-quality Case	English
3835			
3836	X	Chinese	English
3837			
3838	# 权利要求		# Patent Applications Awaiting Examination
3839	1. 一种旋耕机耐磨刀片，包括机壳(1)，其特征在于：所述机壳(1)的左右两侧面均固定设置有侧板(2)，所述机壳(1)的上端面安装有驱动机构(3)，所述驱动机构(3)贯穿于机壳(1)，所述驱动机构(3)的内部安装有转轴(4)，所述转轴(4)的左右两端分别安装在左右两侧板(2)的内侧，所述转轴(4)的外侧设置有安装组件(20)； 所述安装组件(20)包括安装环板(5)、刀片本体(9)、耐磨层(14) …… (后略)		# Claims
3840			A system comprising: one or more processors; and one or more non-transitory computer-readable media storing instructions executable by the one or more processors, wherein the instructions, when executed, cause the system to perform operations comprising: determining a position of a vehicle in an environment, the position comprising two-dimensional coordinates and a yaw value; determining a heading vector associated with the vehicle based at least in part on the position, wherein the heading vector comprises a heading direction; determining a sideslip vector associated with the vehicle based at least in part on a direction of motion of the vehicle and a velocity, wherein the sideslip vector comprises a sideslip direction that is different from the heading direction; determining a curvature of the vehicle by: determining a predicted heading vector associated with the vehicle based at least in part on a predicted position of the vehicle; determining a predicted sideslip vector associated with the vehicle based at least in part on a predicted direction of motion of the vehicle and a predicted velocity; and determining the curvature based at least in part on a difference between the predicted sideslip vector and the sideslip vector and a difference between the predicted heading vector and the heading vector...(omit)
3841	#Claims of Rights		
3842	1. A wear-resistant blade for a rotary tiller, comprising a housing (1), characterized in that: side plates (2) are fixedly provided on both sides of the housing (1), a driving mechanism (3) is installed on the upper end surface of the housing (1), the driving mechanism (3) penetrates through the housing (1), a rotating shaft (4) is installed inside the driving mechanism (3), and the left and right ends of the rotating shaft (4) are respectively installed on the inner sides of the left and right side plates (2), and an installation component (20) is provided on the outer side of the rotating shaft (4); The installation component (20) comprises an installation ring plate (5), a blade body (9), and a wear-resistant layer (14)...(omit)		
3843			
3844			
3845			
3846			
3847			
3848	gpt-4o-mini:		
3849	本发明涉及一种旋耕机耐磨刀片及其旋耕机。所述耐磨刀片包括机壳、侧板、驱动机构和转轴。转轴外侧设有安装组件，安装组件包括安装环板、刀片本体和耐磨层。安装环板的正面设有限位槽一，槽内固定有限位凸块，并开设有多个螺孔以便于安装。刀片本体的右端设有安装部，安装部上端面开设有限位槽二和螺孔一，螺孔二与螺孔一相互匹配并配有锁紧螺栓… (后略)		The present invention relates to a system and method for retrieving patient information utilizing large language models (LLMs). The system comprises a processor and a memory that stores instructions for processing natural language queries received from users. Upon receiving a query, the system inputs it into an LLM to generate a computer language query composed of multiple nodes. These nodes are then mapped to entries in a patient database, and a database response is obtained based on this mapping. The system generates a final database query from the database response and queries the patient database accordingly. The user response is then transmitted to a graphical user interface, which includes generating a natural language response through the LLM. This response generation involves analyzing a user profile, training a classifier with exemplary query data, and classifying the user profile into language groupings. The system also includes features for generating user interface data structures and identifying atomic elements within queries to enhance the accuracy and relevance of the information retrieved...(omit)
3850			
3851			
3852			
3853			
3854			
3855			
3856			
3857			
3858			
3859			
3860			
3861			
3862			
3863			
3864			
3865			
3866			
3867			
3868			
3869			
3870			
3871			
3872			
3873			
3874			
3875			
3876			
3877			
3878			
3879			
3880			
3881			
3882			
3883			
3884			
3885			
3886			
3887			
3888			
3889			
3890			
3891			
3892			
3893			
3894			
3895			
3896			
3897			
3898			
3899			
3900			
3901			
3902			
3903			
3904			
3905			
3906			
3907			
3908			
3909			
3910			
3911			
3912			
3913			
3914			
3915			
3916			
3917			
3918			
3919			
3920			
3921			
3922			
3923			
3924			
3925			
3926			
3927			
3928			
3929			
3930			
3931			
3932			
3933			
3934			
3935			
3936			
3937			
3938			

Figure 71: Low-quality case of task 4-1.

Chinese	Task 4-2: High-quality Case	English
<p># 独立权利要求项</p> <p>1. 自动麻将机，其特征在于，包括环形推牌装置，所述环形推牌装置包括推牌槽，所述推牌槽包括基座、环形直立挡板和环形挡牌侧板，所述环形直立挡板和环形挡牌侧板设在所述基座上，所述环形挡牌侧板位于所述环形直立挡板的外侧，所述环形直立挡板的弯曲角度大于180°，所述推牌槽设有进牌口和出牌口（后略） 请根据给定的独立权利要求项生成其对应的所有的从属权利要求。</p> <p>#Independent claim:</p> <p>An automatic mahjong machine, characterized in that it comprises a circular card pushing device, the circular card pushing device comprising a card pushing groove, the card pushing groove comprising a base, a circular upright baffle, and a circular blocking side plate. The circular upright baffle and the circular blocking side plate are arranged on the base, and the circular blocking side plate is located on the periphery of the circular upright baffle. The bending angle of the circular upright baffle is greater than 180°, and the card pushing groove is provided with a card inlet and a card outlet. Please generate all corresponding dependent claims based on the given independent claims</p> <p>gpt-4o-mini:</p> <p>根据给定的独立权利要求项，以下是可能的从属权利要求：</p> <ol style="list-style-type: none"> <li>根据权利要求1所述的自动麻将机，其特征在于，所述环形推牌装置的推牌槽的基座上设有驱动装置，用于驱动推牌槽内的麻将牌沿推牌槽移动。</li> <li>根据权利要求1所述的自动麻将机，其特征在于，所述环形直立挡板和环形挡牌侧板由耐磨损材料制，以提高使用寿命。</li> <li>根据权利要求1所述的自动麻将机，其特征在于，所述进牌口和出牌口的设计使得麻将牌在推牌过程中不易卡住……（后略） dependent claims:2. The automatic mahjong machine according to claim 1, characterized in that a driving device is provided on the base of the pushing slot of the circular pushing device, for driving the mahjong tiles in the pushing slot to move along the pushing slot.3. The automatic mahjong machine according to claim 1, characterized in that the annular upright blocking plate and the annular blocking plate side plate are made of wear-resistant materials to improve their service life.4. The automatic mahjong machine according to claim 1, characterized in that the design of the card inlet and outlet makes it difficult for mahjong tiles to get stuck during the pushing process... (omit)</li> </ol>	<p># Independent Claim</p> <p>A scanning mirror assembly, comprising: a two-dimensional micro-electromechanical system (MEMS) scanning mirror; a first pair of piezoelectric electrodes coupled to the MEMS scanning mirror through a first pair of looped torsion springs, wherein the first pair of piezoelectric electrodes drives the MEMS scanning mirror to rotate around a first axis; and a second pair of piezoelectric electrodes coupled to the MEMS scanning mirror through a second pair of looped torsion springs, wherein the second pair of piezoelectric electrodes drives the MEMS scanning mirror to rotate around a second axis orthogonal to the first axis, wherein at least one of the first pair of looped torsion springs or the second pair of looped torsion springs are multiple-looped torsion springs. a two-dimensional micro-electromechanical system (MEMS) scanning mirror; a first pair of piezoelectric electrodes coupled to the MEMS scanning mirror through a first pair of looped torsion springs, wherein the first pair of piezoelectric electrodes drives the MEMS scanning mirror to rotate around a first axis; and a second pair of piezoelectric electrodes coupled to the MEMS scanning mirror through a second pair of looped torsion springs, wherein the second pair of piezoelectric electrodes drives the MEMS scanning mirror to rotate around a second axis orthogonal to the first axis, wherein at least one of the first pair of looped torsion springs or the second pair of looped torsion springs are multiple-looped torsion springs. Please generate all dependent claims corresponding to the given independent claim.</p> <p>gpt-4o-mini:</p> <p>Here are several dependent claims that correspond to the provided independent claim regarding the scanning mirror assembly:</p> <p>Dependent Claim 2: The scanning mirror assembly of claim 1, wherein the multiple-looped torsion springs are configured to provide enhanced rotational stability and precision for the MEMS scanning mirror.</p> <p>Dependent Claim 3: The scanning mirror assembly of claim 1, wherein the first pair of piezoelectric electrodes is configured to operate in a push-pull manner to achieve a greater range of motion for the MEMS scanning mirror... (omit)</p>	

Figure 72: High-quality case of task 4-2

3888  
3889  
3890  
3891  
3892  
3893  
3894  
3895  
3896  
3897  
3898  
3899  
3900  
3901  
3902  
3903  
3904  
3905  
3906  
3907  
3908  
3909  
3910  
3911  
3912  
3913  
3914  
3915  
3916  
3917  
3918  
3919  
3920  
3921  
3922  
3923  
3924  
3925  
3926  
3927  
3928  
3929  
3930  
3931  
3932  
3933  
3934  
3935  
3936  
3937  
3938  
3939  
3940  
3941

Chinese      Task 4-2: Low-quality Case      English

# 独立权利要求项  
1. 一种用于性感带的刺激装置(1)，包括：至少一个压力场生成器(2)，该至少一个压力场生成器(2)具有：至少一个第一腔室(3)；所述第一腔室(3)具有单个开口，和驱动部件(6)，所述驱动部件(6)改变所述第一腔室(3)的体积；以改变所述压力场生成器(2)中的总压力，其中，所述总压力包括经调节的相对于参照压力的正压或负压；和使所述驱动部件(6)开动的控制装置(7)；所述压力场生成器(2)没有阀门；所述刺激装置(1)是手持的装置。  
请根据给定的独立权利要求项生成其对应的所有从属权利要求。  
#Independent claims  
1. A stimulation device (1) for a sex belt, comprising: at least one pressure field generator (2), the at least one pressure field generator (2) having: at least one first chamber (3); the first chamber (3) has a single opening and a driving component (6) that changes the volume of the first chamber (3) to alter the total pressure in the pressure field generator (2), wherein the total pressure includes modulated positive or negative pressure relative to a reference pressure; And a control device (7) for activating the driving component (6); The pressure field generator (2) does not have a valve; The stimulation device (1) is a handheld device. Please generate all dependent claims corresponding to the given independent claims.

gpt-4o-mini:  
根据给定的独立权利要求项，以下是可能的从属权利要求：  
1. 根据权利要求1所述的刺激装置(1)，其中所述第一腔室(3)的开口具有可调节的直径，以便调节所述压力场生成器(2)的输出压力。  
2. 根据权利要求1所述的刺激装置(1)，其中所述驱动部件(6)为电动驱动装置，能够通过电信号控制其动作。  
3. 根据权利要求1所述的刺激装置(1)，其中所述控制装置(7)包括一个用户界面，允许用户选择不同的刺激模式……（后略）  
Based on the given independent claims, the following are possible dependent claims:1. The stimulation device (1) according to claim 1, wherein the opening of the first chamber (3) has an adjustable diameter for adjusting the output pressure of the pressure field generator (2).2. The stimulation device (1) according to claim 1, wherein the driving component (6) is an electric driving device capable of controlling its action through electrical signals.3. The stimulation device (1) according to claim 1, wherein the control device (7) comprises a user interface that allows the user to select different stimulation modes... (omit)

# Independent Claim  
A semiconductor device comprising: a substrate; an alignment marker embedded in the substrate, the alignment marker comprising tantalum (Ta) and having a line edge roughness of less than 2.5 nm; a plurality of transistors in the substrate; and a plurality of metal interconnects and metal vias connected to the plurality of transistors, a substrate; an alignment marker embedded in the substrate, the alignment marker comprising tantalum (Ta) and having a line edge roughness of less than 2.5 nm; a plurality of transistors in the substrate; and a plurality of metal interconnects and metal vias connected to the plurality of transistors.  
Please generate all dependent claims corresponding to the given independent claim.

gpt-4o-mini:  
Here are several dependent claims that could correspond to the independent claim provided:  
Dependent Claim 1: The semiconductor device of claim 1, wherein the alignment marker is positioned at a predetermined location within the substrate to facilitate alignment during manufacturing processes.  
Dependent Claim 2: The semiconductor device of claim 1, wherein the plurality of transistors includes at least one type selected from the group consisting of NMOS transistors, PMOS transistors, and FinFET transistors.  
Dependent Claim 3: The semiconductor device of claim 1, wherein the metal interconnects are made of copper (Cu) or aluminum (Al).  
Dependent Claim 4: The semiconductor device of claim 1, wherein the substrate is made of silicon (Si) or silicon-on-insulator (SOI).  
Dependent Claim 5: The semiconductor device of claim 1, wherein the line edge roughness of the alignment marker is less than 2.0 nm.  
Dependent Claim 6: The semiconductor device of claim 1, further comprising a dielectric layer between the alignment marker and the plurality of transistors.  
Dependent Claim 7: The semiconductor device of claim 1, wherein the alignment marker is formed using a deposition technique selected from the group consisting of sputtering, chemical vapor deposition (CVD), and atomic layer deposition (ALD)... (omit)

Figure 73: Low-quality case of task 4-2.

3912  
3913  
3914  
3915  
3916  
3917  
3918  
3919  
3920  
3921  
3922  
3923  
3924  
3925  
3926  
3927  
3928  
3929  
3930  
3931  
3932  
3933  
3934  
3935  
3936  
3937  
3938  
3939  
3940  
3941

Chinese      Task 4-3: Correct Case      English

请从A, B, C, D四个选项中选出题目对应的答案。  
某专利保护工业机器人手臂的齿轮传动关节结构，核心为三级减速齿轮组。下列哪种改进最可能规避侵权？  
Question: A patent protects the gear transmission joint structure of an industrial robot arm, with the core being a three-stage reduction gear set. Which of the following improvements is most likely to avoid infringement?  
A. 改用谐波减速器替代齿轮组  
B. 将齿轮材料从钢换成钛合金  
C. 增大齿轮模数以提升强度  
D. 增加润滑油注入口数量  
D. Increase gear module to enhance strength.

gpt-4o-mini:  在这个问题中，我们需要考虑如何通过改进来规避对某专利的侵权。专利的核心是三级减速齿轮组，因此我们需要寻找一种改进方式，能够在不使用该专利的核心技术的情况下，达到类似的效果。  
A. 改用谐波减速器替代齿轮组：谐波减速器是一种不同于传统齿轮组的减速装置，使用了不同的工作原理，因此这种改进最可能规避侵权。  
B. 将齿轮材料从钢换成钛合金：虽然材料的改变可能会影响性能，但核心的齿轮结构和工作原理没有改变，因此这并不能有效规避侵权。  
C. 增大齿轮模数以提升强度：这仍然是在使用齿轮组的设计，只是改变了齿轮的尺寸和强度，依然可能侵犯专利。  
D. 增加润滑油注入口数量：这只是对现有设计的一个小改动，并没有改变齿轮组的基本结构和工作原理，因此也无法规避侵权。  
综上所述，最有效的改进方式是选项A，即改用谐波减速器替代齿轮组。  
In this issue, we need to consider how to avoid infringement of a certain patent through improvement. The core of the patent is the three-stage reduction gear set, so we need to find an improved way to achieve similar functionality without using the core technology of the patent. A. Replacing gear sets with harmonic reducers: Harmonic reducers are a type of reduction device different from traditional gear sets, using different working principles, so this improvement is most likely to avoid infringement. B. Replacing the gear material from steel to titanium alloy: Although the change in material may affect performance, the core gear structure and working principle have not changed, so this cannot effectively avoid infringement. C. Increasing the gear module to enhance strength: This is still using the design of gear sets, only changing the size and strength of the gears, which may still infringe on patents. D. Increasing the number of lubricant injection ports: This is only a minor modification to the existing design and does not change the basic structure and working principle of the gear set, so infringement cannot be avoided. In summary, the most effective improvement method is option A.

Question: Please select the correct answer from A, B, C, and D.  
A patented autonomous drone navigation system (US 10,555,123) requires "LiDAR-based obstacle avoidance with 360° point cloud fusion" and "dynamic path recalculation every 0.1 seconds." Which redesign avoids infringement under the doctrine of equivalents?  
A. Using stereo cameras instead of LiDAR but achieving similar resolution  
B. Increasing path recalculation frequency to 0.05 seconds  
C. Combining LiDAR with ultrasonic sensors for redundancy  
D. Adopting pre-mapped static routes without real-time recalculation

gpt-4o-mini:  To determine which redesign avoids infringement under the doctrine of equivalents, we need to analyze each option in relation to the patented features of the autonomous drone navigation system.  
1. Understanding the Patent Claims: The patent requires "LiDAR-based obstacle avoidance with 360° point cloud fusion." It also specifies "dynamic path recalculation every 0.1 seconds."  
2. Evaluating Each Option:  
A. Using stereo cameras instead of LiDAR but achieving similar resolution: This option substitutes LiDAR with stereo cameras. While it may achieve similar resolution, it does not use LiDAR, which is a specific requirement of the patent. This could potentially avoid infringement.  
B. Increasing path recalculation frequency to 0.05 seconds: This option changes the frequency of recalculation but still uses dynamic path recalculation. Since it still falls under the same functional requirement of the patent, it does not avoid infringement.  
C. Combining LiDAR with ultrasonic sensors for redundancy: This option still uses LiDAR, which is a key element of the patent. Therefore, it does not avoid infringement.  
D. Adopting pre-mapped static routes without real-time recalculation: This option eliminates the dynamic recalculation aspect entirely, which is a critical feature of the patent. By not recalculating paths dynamically, it avoids the infringement of that specific claim.  
3. Conclusion: The best option that avoids infringement under the doctrine of equivalents is D, as it fundamentally changes the method of navigation by removing the dynamic recalculation requirement.  
Final answer: D

Figure 74: Correct case of task 4-3.

3942

3943

3944

3945

3946

3947

3948

3949

3950

3951

3952

3953

3954

3955

3956

3957

3958

3959

3960

3961

3962

3963

3964

3965

3966

3967

3968

3969

3970

3971

3972

3973

3974

3975

3976

3977

3978

3979

3980

3981

3982

3983

3984

3985

3986

3987

3988

3989

3990

3991

3992

3993

3994

3995

X      Chinese      Task 4-3: Error Case      English

请从A, B, C, D四个选项中选出题目对应的答案。  
某专利保护机床刀具的三点式径向定位夹具,下列哪种改进最可能规避侵权?  
Please refer to A, B. Choose the answer corresponding to the question from the four options C and D.  
Question: A three-point radial positioning fixture for machine tool tools protected by a certain patent. Which of the following improvements is most likely to avoid infringement?  
A. 加装温度传感器  
A. Install temperature sensor.  
B. 使用碳化钨镀层夹具  
B. Use tungsten carbide coating fixture.  
C. 增加定位点至四个  
C. Increase the number of positioning points to four.  
D. 设计为V型槽轴向夹紧结构  
D. Designed as a V-shaped groove axial clamping structure.

**gpt-4o-mini: X**

C. 增加定位点至四个  
C. Increase the number of positioning points to four.

**Error Category: Composite error**  
**Error Reason:** The most effective way to avoid patent infringement is usually to modify or replace the key features of the patent. The core feature of the three-point radial positioning fixture mentioned in the title is the design of its three-point positioning. A. Installing a temperature sensor - this is not related to the positioning function of the fixture and may not effectively avoid infringement. B. Using tungsten carbide coating fixes - this is mainly a change in material and may not affect the core design of the patent. C. Increasing the number of positioning points to four directly changes the design features of the original patent from three points to four points, which may effectively avoid infringement. D. Designed as a V-shaped groove axial clamping structure - although this is a design change, it may still be similar in function to the original patent. Taking all factors into consideration, increasing the number of positioning points to four is the most direct and effective way to avoid infringement.

Please select the correct answer from A, B, C, and D.  
Question: A medical device patent (US 9,878,210) covers a "wearable insulin pump with real-time glucose monitoring using subcutaneous microdialysis." Which modification would most likely avoid infringement if the original patent claims specify "microdialysis membrane thickness of 10-15μm" and "feedback-controlled dosing based on glucose levels"?

A. Replacing microdialysis with non-invasive Raman spectroscopy for glucose detection.  
B. Using a 16μm microdialysis membrane with identical feedback algorithms.  
C. Adding Bluetooth connectivity to share data with smartphones.  
D. Implementing an open-loop dosing system without glucose feedback.

**gpt-4o-mini: X**

A. Replacing microdialysis with non-invasive Raman spectroscopy for glucose detection.

**Error Category: Reasoning Error**  
**Error Reason:** To determine which modification would most likely avoid infringement of the patent covering a "wearable insulin pump with real-time glucose monitoring using subcutaneous microdialysis," we need to analyze each option in relation to the specific claims of the patent.  
A. Replacing microdialysis with non-invasive Raman spectroscopy for glucose detection - This modification changes the method of glucose detection entirely, moving away from microdialysis to a different technology. This would likely avoid infringement as it does not utilize the patented method X.  
B. Using a 16μm microdialysis membrane with identical feedback algorithms\*\* - This option still uses microdialysis, which is a key aspect of the original patent. The change in membrane thickness (from 10-15μm to 16μm) may not be sufficient to avoid infringement, especially since the feedback-controlled dosing based on glucose levels remains unchanged.  
C. Adding Bluetooth connectivity to share data with smartphones - While this adds a feature to the device, it does not change the core technology of glucose monitoring or dosing. Therefore, it is unlikely to avoid infringement.  
D. Implementing an open-loop dosing system without glucose feedback - This modification removes the feedback-controlled dosing aspect, which is a critical part of the original patent. By eliminating the feedback mechanism, this option could potentially avoid infringement. Considering the analysis, the most effective modification to avoid infringement is option X, as it completely changes the method of glucose detection, thus steering clear of the patented technology.

Figure 75: Error case of task 4-3.