# 基于生成式大语言模型的人工智能专利识别研究*

□ 王新成　郭彧铭　李　垣　于晓宇

领域编辑推荐语：

有效地识别 AI 专利，对于当前相关领域的学术研究具有重要现实意义。论文提出基于 LLMs 的 AI 专利识别框架，显著提升了识别的有效性，具有推广性。

——贾良定

**摘　要：**人工智能（AI）专利作为技术创新的关键载体，承载着关键技术路径与应用发展脉络。面对当前 AI 专利数量激增、技术表述日益复杂、创新边界持续演变的现实挑战，传统识别方法在时效性、覆盖度与一致性等方面均面临显著局限。本文梳理既有专利识别方法的理论基础与适用情境，进而提出了一种基于生成式大语言模型（LLMs）的 AI 专利识别框架，并开展了系统的实证评估。研究采用统一的数据集和实验设置，对 LLMs 与传统机器学习、深度学习方法进行了比较。实验结果表明，基于 LLMs 构建的 AI 专利识别方法显著优于传统对照方法，尤其在处理技术边界模糊、术语动态演进、标注数据稀缺等复杂专利文本场景时，该方法展现出更强的适应性与稳健性，凸显其在新兴技术识别任务中的独特优势。本研究提出的 AI 专利识别框架不仅推动了专利文本智能分类的方法创新，而且也为 AI 技术演进分析、创新能力度量以及创新政策评估等实证研究提供了可靠的研究工具，具有重要的理论价值与实践意义。本研究用于 AI 专利识别的模型架构说明、提示工程及运行代码可访问开源代码库（https：//github. com/guoyumingsjtu/llm - patent - classifier）获取。

**关键词：**人工智能专利；专利识别；大语言模型；机器学习

*　国家自然科学基金面上项目：商业生态视角下先动企业与跟随企业数字化转型战略协同机制的研究（项目编号：[72372102]）；国家自然科学基金重点项目：人工智能时代科创企业技术战略与商业模式共演成长路径研究（项目编号：[72532008]）。

## 一、引言

专利文本作为系统记载技术方案、技术功效及其应用场景的关键信息载体，不仅为识别与解析技术创新内涵提供了底层数据支撑，而且更成为分析技术演进路径、度量创新能力、评估政策效果的重要实证依据（Mansfield，1986；王霄等，2022；Kriesch & Losacker，2024）。基于专利数据，学界已逐步发展出涵盖技术强度、创新能力、发明者合作网络及技术扩散路径在内的多维度测度体系（尤树洋等，2023）。在人工智能领域，专利更因其高度凝结技术细节与丰富的技术信息，成为开展 AI 实证研究不可或缺的重要数据来源（Arts et al.，2023；Hou et al.，2023）。然而，随着 AI 专利数量激增与文本复杂度持续提升，研究者在高效、准确识别 AI 专利方面遇到较大的挑战（Miric et al.，2023）。

首先，AI 专利申请量的规模化增长使传统识别方法陷入可扩展性困境。近年来，AI 专利申请量急剧增加，基于人工筛查与静态规则的传统识别方法越来越难以适应海量数据与技术快速迭代的现实要求，这导致传统识别方法的成本急剧增加和效率持续走低，陷入可扩展性的困境（Gong et al.，2025）。其次，AI 专利文本在结构与表达上具有区别于传统技术领域的特殊复杂性。该类文本通常包含大量涉及算法架构、模型参数与数据处理流程的功能性描述，其核心创新点常隐含于具体实施细节之中，缺乏明确的技术标签。AI 领域普遍存在多词术语、专有缩略语以及灵活的同义技术表述，致使仅依赖关键词匹配或局部语法特征的方法难以准确识别其语义内核，从而引发系统性的误判与漏判（Aceves & Evans，2024）。最后，作为通用目的技术，AI 的强渗透性与技术融合性进一步加剧了专利识别的边界模糊问题。新兴技术范式与专业术语层出不穷，而传统分类体系与主题词表的更新往往滞后于技术实践，致使基于固定词典或静态分类号的识别方法均存在明显局限，难以稳定、有效地刻画快速变迁中的技术真实图景（Giczy，2022）。

现有专利识别方法可分为两类。第一类是传统的专利识别方法，包括人工判读、基于国际专利分类号（IPC）和关键词检索三类方法。此类方法具备实现成本低、可解释性强与结果可复核性高等优势；然而，其在识别跨句语义关联与隐含技术表达方面能力有限，难以有效应对术语快速演变及跨领域技术融合带来的复杂性，易出现覆盖不全与更新滞后等问题。因此，该类方法更适用于术语体系相对稳定、分类边界明确的专利领域。第二类为监督式机器学习方法，包括基于文本特征的传统机器学习与基于深度学习的识别模型。这类方法通常将文本转化为向量表示后，采用朴素贝叶斯、支持向量机或随机森林等分类器进行识别，其建模过程相对透明，计算资源需求适中，便于进行误差分析与稳健性检验，在样本有限或需快速部署的场景中具有一定可行性；然而，其在处理跨句依赖与长语境方面存在局限，对同义替换、术语变体及文体差异的适应能力较弱（Rathje et al.，2024）。深度学习方法以分布式表示为理论基础，通过端到端训练自动提取语义特征，典型模型包括用于局部特征提取的卷

积神经网络（CNN）、用于序列建模的循环神经网络（RNN），以及基于变换器架构的预训练语言模型（Giczy，2022）。尽管深度学习模型在长文本与复杂语义任务中通常表现更优，但其对数据规模与计算资源的要求较高，跨语言、跨体裁迁移时需额外调整。此外，模型决策机制具有黑箱特性，结果解释与溯源难度较大，对研究结果的解释与可重复性构成挑战（Zhou et al.，2024）。

针对 AI 专利呈现出的新特性以及既有识别方法的局限性，本文将生成式大语言模型（Large Language Models，LLMs）应用于 AI 专利识别任务。LLMs 基于变换器架构，借助自注意力机制实现全局语境建模，并通过大规模自监督预训练获得层次化语义表示（Dathathri et al.，2024）。在恰当的提示构建与少样本示例引导下，LLMs 无需进行参数微调即可实现高效而精准的文本识别（Singhal et al.，2023）。本文设计了一套可审计、可迁移的操作流程：构建融合“规则—证据—结论”的提示框架以约束识别边界，采用零样本提示、少样本提示以及基于角色的提示策略辅以多数投票与上下文扰动以增强模型的泛化能力。在统一训练集与测试集的设置下，本文系统比较了 LLMs 与深度学习模型、传统机器学习模型的性能。实验结果表明，LLMs 在准确率与 F1 值上整体占优，尤其在精确率与召回率的均衡方面表现突出。此外，在训练数据集不断收缩的情况下，LLMs 并未出现性能坍塌，始终维持较高的表现。本文构建的专利识别方法与既有方法形成互补体系，传统的专利识别方法适用于术语与分类体系相对稳定的技术领域，传统机器学习与深度学习在成本与可解释性之间取得折中，但对长文本中隐含语义线索的捕捉能力有限；而在标注资源稀缺、文本异质性强且技术边界动态演化的情境下，LLMs 展现出更强的适应能力。

本文的研究贡献体现在以下三个方面。第一，本文系统梳理并比较了传统专利识别方法与基于有监督式机器学习的专利识别方法的适用范围与优劣，指出了不同方法在识别过程中可能产生的误判与漏判情境，从而为后续研究在工具选择、参数设定与结果比较方面提供了具体的指引。第二，本文将生成式大语言模型引入 AI 专利识别任务，围绕“规则—证据—结论”的推理链条对提示工程进行结构化与外显化设计，并结合多数投票等稳健性机制，提出一套可复现的专利识别流程。第三，本文所构建的方法在既定数据与对照框架下实现了对现有专利识别方法的显著性能提升，并在标注样本相对稀疏时仍能保持稳健，从而缓解了传统路径对大规模高质量标注的依赖；同时，该框架具有可迁移性，可拓展至其他非结构化文本分类任务，为数据稀缺条件下的高效准确识别提供了可行路径。

## 二、专利识别方法综述

通过对现有文献的系统性回顾可以发现，当前专利识别方法正在经历由传统方法向新兴识别方法的范式转变。传统方法主要包括人工判读、基于 IPC 分类号和关键词检索三类，而新兴方法则以有监督的机器学习方法为主，包括基于文本特征的传统机器学习方法和深度学习方法。尽管已有少量研究尝试将大语言模型

技术引入专利识别任务，但是尚未有研究系统性地比较大语言模型与既有识别方法的性能差异。本部分回顾既有专利识别方法的基本思路、适用情景及其优势与劣势。

**（一）传统专利识别方法**

1. 人工判读

人工判读作为专利分析的基础方法，主要依赖领域专家对专利文本进行逐篇解读与专业识别。该方法在技术发展初期或专利数量有限时具有独特优势，能够深入解析技术发展脉络，精准识别核心创新点。然而，随着 AI 领域专利数据规模的指数级增长与技术复杂度的不断提升，人工判读的局限性日益凸显。首先，在专业能力层面，现代专利分类体系日趋复杂且持续演进，加之 AI 技术本身具有跨学科、跨领域的特性，以及专利文本中普遍存在的跨语言表述差异，显著提高了专业判读的准入门槛，对分析人员的多领域知识储备与实践经验提出了更高的要求。其次，在可扩展性方面，面对数以万计且持续快速增长的 AI 专利，人工方法的时间与人力成本难以持续。这不仅导致处理效率的急剧下降，更使得对新增数据的实时更新与系统性跟踪变得不可实现。最重要的是，人工判读方法在结果的一致性与可复现性方面存在本质缺陷（Cole，2024）。不同判读者之间的主观认知差异、判定标准的不统一，以及标准版本迭代带来的规则变化，都会对结果的稳定性与可比性产生显著影响。这些因素共同制约了人工判读方法在 AI 专利识别这一要求高效、精准且可验证的研究场景中的适用性。

2. 基于 IPC 分类号的识别方法

该方法以国际专利分类号（IPC）等现有技术分类体系为基础，采用规则驱动的思路对专利样本进行分类。常见的方法包括直接依据专利文献标注的主副分类号进行判定，或预先构建与目标技术相关的分类号清单进行归类。在技术体系成熟、术语标准统一、技术边界相对固定的传统产业中，该方法展现出良好的适应性（Lafond & Kim，2019）。然而，在 AI 专利这一特定场景下，该方法面临显著挑战。作为通用目的技术，人工智能具有高度跨学科、跨领域特性，其技术演进速度快、创新形态多样，与既有分类体系之间存在着结构性张力。IPC 的制度设计旨在服务专利文献的组织、检索与审查分工，并非为新兴技术的即时标引而设。其条目调整通常需经历一个较长的周期，即等到相关专利积累到一定规模，且既有分类体系在归类效率与检索成本方面问题凸显后，方能启动 IPC 体系的调整，这就决定了其更新节奏难以跟上快速迭代的技术前沿。同时，出于专利检索与审查管理的实务需要，IPC 归类常以专利的主要应用场景或技术载体为锚点，往往未对其中嵌入的通用技术模块进行标注。

因此，采用 IPC 识别 AI 专利时存在以下局限。第一，IPC 更新迟滞与技术演进脱节。AI 技术迭代迅速、范式更替频繁，而 IPC 的修订周期较长，难以及时为新出现的 AI 技术提供分类指引（Verendel，2023）。尽管部分国家已建立 AI 相关分类索引（如中国《关键数字技术专利分类体系（2023）》将人工智能单列为一级分类），但整体上分类体系的更新速度仍滞后于技术发展节奏。第二，应用场景导向导致 AI 通用技术模块被掩蔽。AI 常作为通用技术模块嵌入医疗、制造、交通等具体应用领域，专利文本

的核心叙述与权利要求多围绕应用问题与技术载体展开。在以检索与审查为导向的归类逻辑下，此类专利更可能被归入应用领域类目，而非 AI 专门类目，致使仅依赖 AI 相关分类号清单的识别策略容易遗漏“场景主导型”AI 创新。例如，基于深度学习的医学影像诊断装置可能主要归类于医疗器械类目。

3. 关键词检索

基于关键词检索的识别方法通过构建领域关键词表或主题词库，在专利标题、摘要等结构化字段中进行术语匹配，以此实现专利分类与筛选。该方法具有实现成本低、逻辑可解释性强和结果可复现性好等优势，在技术体系成熟、术语标准统一的传统产业中表现良好。然而，将关键词检索方法应用于人工智能专利识别时，其静态匹配机制与 AI 技术的动态特性之间存在矛盾。AI 领域的技术演进速度快，新术语、新架构与新范式持续涌现，使得静态预设的关键词表难以保持时效性，导致对新兴技术主题的识别存在滞后。

同时，AI 术语普遍具有高度的语境依赖性和语义多义性。同一术语在不同技术子领域或应用场景中可能承载差异化含义，而单纯依赖词面匹配无法有效识别这种语义变化，容易产生误判。例如，“模型”“训练”“推理”等词语在 AI 专利中通常对应特定的机器学习建模与推断机制，但在统计仿真、控制优化或设备调试等非 AI 语境中也被广泛使用且语义指向存在不同，从而使基于词面匹配的识别方法容易产生误判。此外，AI 专利文本中普遍存在的复合型技术表述、跨语言混合使用以及不同申请人采用的表述差异，都显著增加了构建完备关键词列表的复杂性与维护成本。现有研究指出，在 AI 这类技术快速演进、学科交叉深入的领域，预设关键词列表难以全面捕捉动态变迁的技术热点与新兴分支（Damioli et al.，2024）。若仅依赖有限的关键词进行识别，容易低估 AI 创新活动的实际范围与规模。

因此，在 AI 专利识别实践中，关键词检索方法更适合作为初步筛选工具或辅助判据。其有效性需要与语义理解、表示学习等先进文本分析技术相结合，通过构建动态更新的术语识别机制，才能缓解因术语快速演化与语境多样性带来的偏差（Damioli et al.，2024）。

**（二）有监督式机器学习的识别方法**

随着机器学习技术的不断发展，数据驱动的文本分类方法已被广泛应用于专利识别领域。基于文本特征提取和模型架构的差异，现有监督学习方法可分为两类：基于人工设计特征的传统机器学习方法和基于预训练语言模型的深度学习方法。下文将分别对这两类方法进行系统阐述。

1. 基于人工设计特征的传统机器学习

该方法通过将专利文本转化为结构化特征表示，再结合传统分类器实现专利识别（Giczy，2022）。具体而言，研究者通常从专利摘要、权利要求等关键字段中提取 TF－IDF、n－gram 等统计特征，构建专利文本的数值化表征，继而采用支持向量机、朴素贝叶斯、随机森林等模型建立分类决策边界（LeCun et al.，2015；Qaiser & Ali，2018；Kamateri et al.，2023；Yun & Geum，2020）。将这一方法应用于 AI 专利识别时，其内在的特征构建方式面临三重特殊挑战。首先，该方法依赖固定的特征词典，但 AI 技术术

语快速迭代，静态词典难以适应新概念的持续涌现，导致特征表达不完整。其次，该方法基于词面匹配的特征提取机制无法区分同一术语在不同 AI 子领域中的语义差异，造成特征表示失真。最后，该方法使用的局部文本特征难以捕捉 AI 专利中常见的跨领域技术融合特征，如算法优化与硬件加速的协同创新，导致对复杂技术逻辑的表征不足（Jiang & Goetz，2025）。

为提升专利识别的效能，研究者从四个维度推进方法创新：通过动态更新机制缓解特征词典的时效性滞后；引入领域知识增强特征语义表示；采用集成学习策略提升对长尾技术的覆盖范围；结合主动学习降低高质量标注数据的获取成本（Yun & Geum，2020）。这些改进虽在一定程度上缓解了特征表征的不足，但由于人工特征工程固有的局限性，仍难以充分捕捉 AI 专利中复杂的语义关联和技术演进特征，在跨领域应用时表现出明显的性能局限（Zhang et al.，2025）。

2. 深度学习

近年来，深度学习凭借较强的表征学习能力，逐步成为专利文本识别的前沿方法。相较于依赖人工设计特征的传统做法，深度模型通过多层神经网络自动学习分布式表示，能够更充分地捕捉跨句、跨段的语义依存关系，在面向长文本的专利分类任务中表现出更高的适配性与识别力。实证研究表明，在数据同质、评估框架一致的条件下，深度学习模型在准确率、召回率等关键指标上通常能够显著超越传统基线方法。

从技术演进的角度看，深度学习在专利识别中的应用经历了从浅层网络到深层架构的持续发展。早期研究主要基于词向量嵌入与卷积神经网络（CNN）相结合的方法，通过 CNN 提取局部文本特征，在多项专利分类任务中取得了优于传统机器学习模型的效果（Li et al.，2018；Ebrahimi et al.，2022）。随着对序列建模需求的增强，研究者开始引入长短期记忆网络（LSTM）等循环神经网络结构，通过其门控机制捕捉文本中的长距离依赖关系，在技术语境复杂、语义连贯性强的专利识别场景中表现出色（Choudhury et al.，2019）。

该领域的关键突破来自 Transformer 架构及预训练语言模型的广泛应用。Transformer 通过自注意力机制实现对文本全局上下文的高效建模，并在大规模语料上以自监督方式学习通用语言表示，经任务特定微调后能够快速适配下游专利分类任务，显著提升模型的泛化能力（Gentzkow et al.，2019）。以 BERT 为代表的双向编码器模型及其改进版本（如 ALBERT、RoBERTa 等）通过更高效的参数利用和训练策略优化，在多个专利数据集上不断刷新性能纪录。与此同时，针对专利文本的结构化特性，研究者还开发了 PatentBERT 等领域专用模型，通过在专利语料上进行继续预训练，进一步增强对权利要求、技术方案等关键内容的语义理解能力（王霄等，2022）。

针对 AI 专利的特性，深度学习展现出独特的优势：首先，通过自注意力机制，模型能够准确识别技术方案的核心创新点，区分基础理论突破与工程实现优化；其次，预训练语言模型对技术术语的上下文感知能力，使其能够准确理解同一概念在不同 AI 子领域中的具体含义；最后，模型的架构灵活性使其能够适应 AI

技术的快速演进，通过持续学习及时捕捉新兴技术方向。然而，该方法同样面临挑战：其一，模型对训练数据的规模和质量要求较高，而 AI 领域高质量、细粒度的专利标注数据相对稀缺；其二，深度学习模型的可解释性较差，难以提供令人信服的识别依据；其三，模型在跨技术领域迁移时性能不够稳定，如在自然语言处理专利数据上训练的模型，直接应用于计算机视觉领域时可能出现性能衰减。

表 1 系统总结了现有专利识别方法的基本特性、适用情景及其优势与局限。

**表 1　　现有专利识别方法适用情景与优劣势**

| 方法范畴 | | 适用情景 | 优势 | 劣势 |
|---|---|---|---|---|
| 传统专利识别方法 | 人工判读 | 小样本或高敏感度任务；边界界定与标注规范制定；复杂疑难样本复核与抽样质检 | • 细读与情境把握强，能识别隐含意图与边界判例<br>• 证据链清晰、可审计、便于异议复核<br>• 可沉淀判定标准，作为标注规范标准 | • 依赖专家经验与标准一致性，主观差异难免<br>• 成本高、效率低，难以规模化覆盖<br>• 跨语种、跨法域复核成本与协调成本较高 |
| | 基于 IPC 的规则分类 | 术语标准化高、技术边界稳定的成熟领域（如机械、医药、半导体）；需要标准统一与审计留痕的场景 | • 标准化与透明度高，易于复现与跨库对接<br>• 与监管统计标准与历史序列衔接顺畅<br>• 运行与维护成本低 | • 分类更新滞后，新兴技术与交叉应用覆盖不足<br>• 单一维度标签难刻画复合属性与多场景嵌入<br>• 各法域实践差异影响一致性与可比性 |
| | 关键词检索 | 候选集构建与高召回初筛；宜与语义表示或领域本体、下游分类器联合使用 | • 部署快捷、复现性好，便于快速扩容调参<br>• 初筛效率高，可按需调控阈值实现高召回或高精度<br>• 便于与人工复核、统计标准衔接 | • 词表静态，难及时吸纳新术语与表述变体<br>• 忽视上下文以及语义歧义，容易出现漏判和误判<br>• 跨语种写法与复合短语处理困难 |
| 有监督式机器学习 | 基于文本特征的传统机器学习 | 中等规模数据与算力受限场景；强调稳定与可解释输出；与规则/深度模型的混合模型 | • 资源需求适中，易上线与维护<br>• 可解释性较好，误差归因清晰，利于模型调试与优化 | • 跨句/长文本语境刻画不足，对同义改写与体裁差异敏感<br>• 跨域迁移依赖人工调参与特征工程<br>• 对概念边界动态变化适应较弱 |
| | 深度学习 | 大规模、结构复杂文本；语种与体裁贴合时优势显著；对准确性要求高且具备流程留痕与输出规范的应用 | • 上下文理解较强，适配长文本<br>• 较能识别隐含线索与术语变体<br>• 可通过再预训练/微调贴合领域语料 | • 数据与算力要求高，部署与运维成本较大<br>• 模型内部机理不易外显，解释与溯源压力较高<br>• 跨语种/体裁迁移需再适配与流程治理 |

## 三、人工智能专利识别的挑战

人工智能专利作为技术创新的重要载体，系统记录了 AI 技术的算法架构、技术功效及其应用场景，长期以来被视为衡量技术进步与创新绩效的关键依据（Wu et al.，2025）。在人工智能引领的新一轮科技革命背景下，已成为学术界与产业界共同关注的话题。然而，随着 AI 技术复杂度的提升和专利规模的扩大，现有 AI 专利识别方法面临以下四个核心挑战。

第一，AI 技术边界动态延展与跨域融合带来的界定难题。人工智能作为通用目的技术，其创新活动横跨机器学习、计算机视觉、自然语言处理等多个技术领域，并深度渗透至金融、医疗、制造等应用场景（Babina et al.，2024）。这种跨领域特性使得传统的单一维度分类标准难以有效捕捉技术的复合性与交叉性，专利的技术边界呈现持续演化态势。

第二，AI 专利数据规模与时效性要求的双重压力。全球 AI 专利数量呈指数级增长，更新频率不断加快，这对识别方法的海量数据处理能力和实时响应能力提出了极高要求。传统依赖人工筛查或静态规则的方法不仅面临高昂的成本压力，更在时间维度上面临稳健性挑战。此外，样本量的激增意味着传统通过抽样标注得到的训练集更难以涵盖不断扩张的技术边界，依赖于已标注数据的监督式模型将面临更大的性能挑战。

第三，AI 专利文本复杂性与语义深度的理解障碍。AI 专利文本通常包含大量专业术语、算法描述和功能性权利要求，关键技术特征常隐含在复杂的语法结构和专业表述中。同时，多词短语、缩略语变体以及跨语言表述差异进一步增加了准确理解的难度，使得基于词面匹配或局部特征的方法容易产生偏差。

第四，现有技术路径的内在局限性。传统专利识别的方法虽具透明性优势，但难以适应快速演进的技术术语；传统机器学习方法在可解释性和计算效率方面表现良好，却对长文本中的隐含语义线索捕捉不足；深度学习方法虽在表征能力上表现突出，但仍受限于数据需求、计算成本及模型可解释性等多重约束。

在此背景下，LLMs 为突破上述困境提供了新的技术路径。基于 Transformer 架构的 LLMs 通过自注意力机制实现对长文本的全局语义建模，能够在不依赖大规模标注数据的前提下，通过精心设计的提示策略和少量示例完成高质量的专利识别（喻理等，2025）。具体而言，LLMs 在应对 AI 专利识别挑战时展现出独特优势：一方面，其强大的语义理解能力能够有效处理术语变体、同义表述和跨体裁文本，提升对专利中隐含技术特征的识别精度；另一方面，预训练阶段积累的丰富知识为其在跨领域场景中的稳定表现提供了坚实基础。更重要的是，通过构建“规则—证据—结论”的可追溯推理框架，将识别标准明确化、示例选择多样化、输出结果结构化，LLMs 能够在保持识别性能的同时增强过程的透明度和可复现性（Shanahan et al.，2023）。基于此，本文将在后续章节详细阐述基于 LLMs 的 AI 专利识别框架的设计理念、实现路径与实证评估结果，以期为推动 AI 专利分析的方法创新提供有益参考。

## 四、基于 LLMs 的 AI 专利识别模型

本节构建基于 LLMs 的 AI 专利识别的框架，整体研究流程如图 1 所示。研究首先构建高质量的专利数据集并进行专业标注，继而系统阐述 LLMs 在 AI 专利识别中的具体应用方法，最后通过严谨的实证比较验证所提框架的有效性。

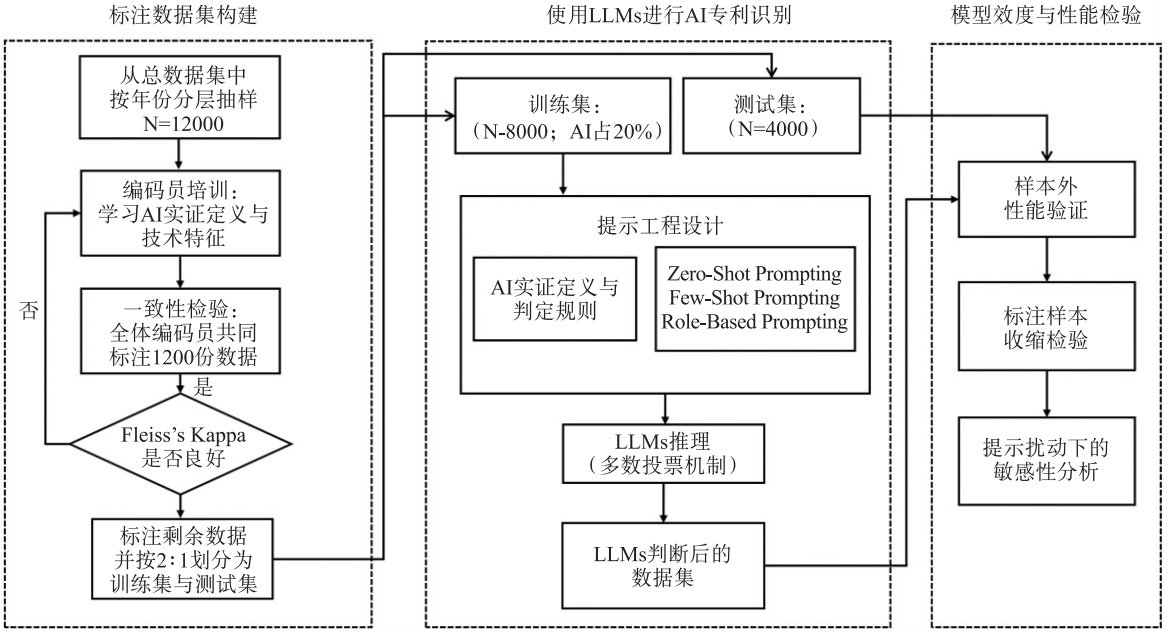


**图 1　基于 LLMs 的 AI 专利识别方法实施流程**

### （一）数据来源与样本构建

本研究数据来源于 IncoPat 全球专利数据库。该数据库涵盖 157 个国家及地区的逾 1.5 亿件专利文献，提供 330 余个结构化检索字段，以其全面的数据覆盖和专业的结构化管理，成为专利研究领域的重要数据来源（喻理等，2025）。本文提取了 2018—2024 年中国上市公司的中文专利记录作为研究样本。这一时间窗口具有特殊意义，它完整涵盖了 AI 技术从理论突破到产业化应用的关键发展阶段，期间 AI 专利申请量呈现爆发式增长，为研究提供了丰富且具有代表性的样本基础。

为确保研究样本的质量与代表性，建立了系统的数据清洗与验证流程。一方面，通过整合上市公司年报、工商登记信息与专利申请人数据，构建了精确的专利主体识别图谱，确保每项专利都能准确归属于对应的上市公司主体。另一方面，采用申请号与公开号双重校验机制，有效识别并剔除重复申报的专利记录。尤其是，考虑到专利审查周期的特点，保留了所有处于有效、实质审查和授权状态的专利记录，仅排除明确失效的专利，这保证了样本的技术时效性。在文本质量方面，实施了多层次的过滤策略：不仅排除摘要缺失的记录，而且还通过正则表达式匹配技术结合人工抽检，系统过滤了包含乱码、格式错误或内容不完整的异常样本。经过这些严谨的处理流程，最终构建的数据集兼具时效性、完整性与清洁度，为后续建模分

析奠定了可靠基础。

**（二）标注数据集构建**

为构建高质量的标注数据集，本研究组建了由三位具有相关研究背景的博士研究生组成的编码小组。基于统计功效分析，采用分层随机抽样方法从总体样本中选取 12000 条专利数据构成标注样本集，确保各年度样本量与当年专利产出规模成比例，保证了时间维度的代表性。由于专利摘要在各类专利数据库中均为披露且规范化的文本字段，能够在有限篇幅内凝练技术要旨与关键特征，并且已被国内外学者广泛用作主要文本分析来源（Miric et al.，2023），本研究因此以专利摘要作为数据标注与模型设计的主要文本依据。

编码团队接受了系统的培训，培训内容涵盖理论基础与实践操作两个维度。在理论层面，深入学习了 Giczy 等（2022）提出的包含机器学习、计算机视觉、自然语言处理等八个核心领域的 AI 技术分类框架，以及 Cockburn 等（2018）建立的 AI 技术实证定义体系与典型应用场景判例。在实践层面，通过典型案例分析、模拟标注练习和小组讨论，统一了识别标准。

为量化评估编码一致性，我们采用随机抽样方法选取 1200 条专利数据（占标注样本的 10%）进行预标注实验。三位编码员独立完成标注后，计算得到 Fleiss's Kappa 系数为 0.85（$p < 0.001$），达到“几乎完全一致”的统计标准，证明编码团队对 AI 专利的识别标准掌握具有高度一致性。基于这一质量保证，团队采用“独立标注—交叉验证—争议协商”的协作模式完成剩余样本的标注工作，确保每项专利都经过至少两次独立识别。

本研究从标注数据集中随机抽取 4000 条数据作为测试集，其余 8000 条数据作为训练集。针对 AI 专利在总体样本中占比较低的实际情况，为规避类别不平衡对模型训练造成的潜在偏差，本研究借鉴 Miric 等（2023）的先进经验，采用合成少数类过采样技术对 AI 相关专利进行智能扩增，将训练集中 AI 专利的占比提升至 20%。这一处理既保持了样本分布的原始特征，又有效增强了模型对少数类样本的识别能力。

**（三）基于 LLMs 的模型架构**

为提升方法在不同模型谱系下的稳健性与泛化性，本文选取两类在学界与产业界具有代表性的通用 LLMs 作为对照对象：DeepSeek – reasoner（由 DeepSeek 团队研发）与 Qwen – plus – latest（由阿里云研发）。其中，DeepSeek – reasoner 在长上下文与指令执行的一致性方面表现稳健；Qwen – plus – latest 在技术文本的理解与任务化输出上具有良好适配性。本文通过 API 对其进行访问并评估，旨在于同一提示与结构化输出规则之下检验方法的有效性与可迁移性。本节将在提示工程设计等方面详细介绍使用 LLMs 对 AI 专利识别的过程。

1. 提示工程设计

在本研究中，提示工程是连接判定规则与模型输出的关键环节。在明确判定标准和输出格式的前提下，精心设计的提示结构能够引导模型在既定边界内作出判断。本文依次构造了零样本提示（Zero – Shot Prompting）、少样本提示（Few – Shot Prompting）与基于角色的提示（Role – Based Prompting）三种策略，在不同程度上平衡对标注数据的依赖程度、性能以及对边界样本的敏感性。

第一，零样本提示不向模型提供具体判例，而是通过文字说明的方式，给出 AI 专利的定义以及纳入或排除标准。该标准与前文编码团队所学习的内容保持一致，本文对既有文献所提出的 AI 技术进行归纳，将其中可操作的判定规则转化为提示文本的主体内容，并以自然语言形式明确应重点关注的技术特征与典型应用场景。此种策略不依赖已有标注数据，易于部署和迁移，但对文本表达方式与体裁差异较为敏感，在接近边界的样本上稳健性有限。

第二，少样本提示是在零样本设计的基础上，进一步将少量已标注的数据作为“正例”“反例”嵌入提示中，以帮助模型形成更清晰的边界。具体而言，LLMs 对某一摘要作出判断之前，先向其呈现少量已标注的代表性判例，使其在统一判定标准之下，学习边界两侧的特征。本文采用对称式设计：在构造判例时，从正类与负类两侧分别选取数量相同的近邻样本，并在提示中以交替、并列的方式予以呈现，避免示例偏向单一方向而诱发系统性偏差。近邻样本主要通过基于 ChineseBERT 的编码器从训练集中检索获得：将摘要编码为语义向量后，按余弦相似度选取各类内部最接近的若干条（类内最近邻 K 条），同时在类内筛选过程中加入去冗余约束以保持判例的信息密度与多样性，以此构成判例集合。相较于仅提供零散示例或随机样本，对称式少样本提示有助于降低示例呈现的片面性，提高模型对边界样本的区分能力，并使后续的判断依据更具可解释性，但其过度拟合的风险较高，消耗的词元（Token）更多。

第三，基于角色的提示通过为模型设定稳定的“专业身份”，强化其阅读路径与输出形式的规范性。本文在提示中将模型明示为“负责筛查 AI 相关技术的专利分析员”或“承担技术梳理工作的研究人员”，并要求其按照类似人工审读的顺序展开。通过这种角色设定与流程约束，模型被持续提醒“必须依据文本作出审慎判断”，而非仅凭一般常识给出直觉性回答，有助于抑制“想象性补全”和风格化发挥。但是基于角色的提示并不保证“专家式”的严谨性，即个别输出可能在语气上符合专业话语，却在证据引用与推理链条上仍显薄弱，呈现出某种“表演性专业”，且若角色设定过于具体，也可能无意中引入特定立场或偏好，对边界样本产生系统性偏差。

本研究分别使用上述三种提示工程构建用于 AI 专利识别的 LLMs 模型，并在相同测试集上验证比较，以筛选出 AI 专利识别情境下的最佳策略。提示工程的具体设计如表 2 所示。

**表 2　　提示工程设计**

| 策略 | 简介 | 具体设计 |
| --- | --- | --- |
| Zero – Shot Prompting | 仅通过文字向模型说明 AI 的定义及判断标准 | 请阅读下方给出的中文专利摘要，并判断它是否属于“AI 相关专利”。在本任务中，请严格依据以下定义与判断标准：【AI 相关专利定义】<br>请只输出一个数字：<br>1：属于 AI 相关专利；<br>0：不属于 AI 相关专利。<br>不要输出任何解释或其他文字。<br>以下是需要判断的专利摘要：【待判断专利摘要】 |

续表

| 策略 | 简介 | 具体设计 |
| --- | --- | --- |
| Few – Shot Prompting | 嵌入少量典型判例，以对称方式向模型展示代表性样本 | 你将阅读下方给出的中文专利摘要，并判断它们是否属于“AI 相关专利”。<br>请首先理解以下定义与判断标准：【AI 相关专利定义】<br>请学习以下样例，该样例展示了专利的摘要以及对应其是否属于 AI 相关专利：<br>【示例 1：正类（标签 =1）】<br>【示例 2：正类（标签 =1）】<br>【示例 3：负类（标签 =0）】<br>【示例 4：负类（标签 =0）】<br>现在，请在理解上述定义与示例的基础上，对下方“待判断摘要”进行分类。请只输出一个数字：<br>1：属于 AI 相关专利；<br>0：不属于 AI 相关专利。<br>不要输出任何其他内容。<br>以下是需要判断的专利摘要：【待判断专利摘要】 |
| Role – Based Prompting | 根据具体任务赋予模型专业角色以聚焦相关领域知识 | 你现在的身份是一名长期从事技术情报分析的专利审查员，负责筛选“AI 相关专利”。请你根据以下定义与判断标准，阅读下方的中文专利摘要，并给出是否属于“AI 相关专利”的判断。AI 相关专利的定义与判断标准如下：【AI 相关专利的定义】<br>请站在“专利分析员”的立场上，先根据上述标准完成判断，但在最终输出中只给出一个数字：<br>1：属于 AI 相关专利；<br>0：不属于 AI 相关专利。<br>不要输出任何解释或附加文字。<br>以下是需要判断的专利摘要：【待判断专利摘要】 |

2. 多数投票机制

在单次推理层面上，即便在判定标准相对稳定、解码温度等生成参数保持不变的条件下，LLMs 对边界样本的输出仍可能出现一定程度的偶然波动。为抑制这种由单次调用带来的不确定性，本文引入多数投票机制：围绕同一待判读的专利摘要，组织若干次推理，将每次输出的二元标签加以记录，并依据简单多数原则确定最终结论，从而在总体上提升样本外识别结果的稳健性与可依赖性。

具体而言，本文在前述三类提示策略的基础上，刻意营造多重“判读视角”。在不改变实质判定规则与有效证据信息的前提下，本文对提示文本实施幅度可控的微调：对于少样本提示，适度打乱正、负判例的呈现顺序，并在预设相似度阈值内以等价样本予以替换；对于零样本与基于角色的提示，则在不影响含义的前提下，对说明语句的先后与措辞作轻微调整。

由此获得的一组判断结果，可视为在同一判定标准之下、基于不同提示策略与证据编排方式形成的多种“观测”，在统计意义上彼此之间具有一定程度的近似独立性。

从方法论角度看，多数投票并非简单地重复若干次同质调用，而是在统一规则与标准的约束下，对提示模板与证据组织方式进行小幅度抽样，再通过表决机制汇聚各次判断，据以削弱单一提示实例所带来的偶然性与系统性偏差。

**（四）对比模型构建**

为确保比较研究的规范性与结果的可比性，本研究在统一的实验设置下，系统对比了 LLMs 方法与两类主流专利识别方法的性能表现：基于文本特征的传统机器学习方法和深度学习方法。所有对比模型均使用相同的数据集进行训练与验证，以消除数据差异对性能评估的影响。

在传统机器学习方法方面，本研究选取了随机森林（Random Forest）和支持向量机（SVC）作为代表性模型。随机森林通过构建多个决策树并采用集成学习策略，有效提升了模型的泛化能力；支持向量机则基于结构风险最小化原则，通过核函数将文本特征映射到高维空间以实现最优分类边界。两种模型均以经过 TF－IDF 加权的文本特征向量作为输入，体现了基于人工特征工程的经典范式。

在深度学习方法方面，本研究选择了 SciBERT 和 ChineseBERT 两个预训练语言模型。SciBERT 基于大规模科学文献语料进行预训练，其词汇表和语言表征更贴近学术文本特点，适用于科学技术领域的文本理解任务；ChineseBERT 则是专门针对中文语言特点设计的预训练模型，通过融入汉字结构特征和中文语言规律，对中文专利文本的语义理解具有天然优势（Masclans et al.，2025）。这两种模型代表了当前专利文本分析中先进的深度学习方法。

## 五、模型效度与性能分析

本节旨在统一测试框架之下，系统呈现所构建模型在中文 AI 专利识别任务中的性能优势及可推广性。首先，对比 LLMs 与多种监督式机器学习方法的样本外识别表现，以评估不同范式的相对优劣；其次，在标注样本规模逐步收缩的设定下，考察各类模型对训练数据依赖程度的差异，贴近管理学研究中普遍存在的“小样本、高标注成本”应用情境；最后，通过提示扰动等方式，对所提出 LLMs 识别框架的稳健性与结果可重复性予以检验。

**（一）样本外模型性能评估**

为保证评价的可比性，本文在统一数据与五折交叉验证框架下，对三类方法的样本外性能进行系统评估。评价采用四项二分类常用指标：Accuracy 衡量总体判定的正确比例；Precision 度量被判为“AI”的样本中真阳性的占比（误报越少越高）；Recall 反映真实“AI”样本被识别出来的比例（漏报越少比例越高）；F1 为 Precision 与 Recall 的调和平均，体现二者的综合权衡。就本研究场景而言，Precision 对应“避免将非 AI 判作 AI”的治理诉求，Recall 对应“尽可能不漏识 AI 专利”的覆盖诉求，F1 则体现二者之间的可接受均衡（Xie，2022）。各模型的具体性能指标如表 3 所示。

表 3　　各模型的具体性能指标

| 模型类型 | 模型名称 | 提示工程策略 | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|---|---|
| 传统机器学习 | Random Forest | | 0.922 | 0.782 | 0.843 | 0.811 |
| | SVC w. RBF kernel | | 0.912 | 0.739 | 0.821 | 0.778 |
| 深度学习 | SciBERT | | 0.906 | 0.796 | 0.711 | 0.751 |
| | ChineseBERT | | 0.938 | 0.838 | 0.858 | 0.847 |
| LLMs | DeepSeek – reasoner | Zero – Shot Prompting | 0.863 | 0.915 | 0.860 | 0.887 |
| | | Few – Shot Prompting | 0.933 | **0.960** | 0.889 | **0.923** |
| | | Role – Based Prompting | 0.875 | 0.911 | 0.872 | 0.891 |
| | Qwen – plus – latest | Zero – Shot Prompting | 0.894 | 0.915 | 0.867 | 0.890 |
| | | Few – Shot Prompting | **0.938** | 0.935 | **0.906** | 0.920 |
| | | Role – Based Prompting | 0.912 | 0.906 | 0.879 | 0.892 |

实证结果显示，LLMs 在各项性能指标上居于领先地位。具体而言，两类 LLMs 在不同提示策略下的 F1 均已明显超出基线模型，其中 DeepSeek – reasoner 在少样本提示下的 F1 达到 0.923，Qwen – plus – latest 在同一策略下的 F1 为 0.920，对应的 Accuracy 分别为 0.933 与 0.938，呈现出兼具精确率与召回率的较高水准。即便在不提供示例的零样本条件下，两类 LLMs 的 F1 也保持在 0.887 至 0.890 的区间，已优于所有对照模型，显示出在仅依托任务描述与判定规则时，对中文专利文本仍具有较强的理解与归类能力。在对比模型中，ChineseBERT 表现次之，其准确率为 0.938，F1 分数为 0.847；而基于传统机器学习的随机森林和支持向量机模型虽然提供了可解释性强、计算成本低的基线方案，但在整体性能上与深度学习方法和大语言模型存在明显差距。尤其是，SciBERT 模型由于在中文专利文本上的适应性不足，其召回率仅为 0.711，显著制约了整体性能表现。上述性能排序在多个评估维度中保持一致，表明 LLMs 的性能优势并非源自某一指标的偶然波动，而是建立在其对复杂语义的深层理解能力之上。为进一步验证上述性能是否显著差异，本文采用配对自助法对模型间的 F1 差异进行检验，即使用同一组样本同步抽取对比模型的预测与真实标签（以保持“配对”结构），计算对比模型间 F1 的差值（ΔF1），进而形成 ΔF1 的经验分布，然后构造置信区间进行双侧检验。结果表明，LLMs 与其他模型的 ΔF1 有显著差异。本研究采用的结构化提示模板以及多数投票机制，共同为 LLMs 准确捕捉专利文本中的隐含技术特征提供了有力支撑，从而显著提升了模型在样本外数据上的稳健性。

基于提示工程视域，不同策略选择之间呈现出较为清晰的性能梯度。综合来看，少样本提示在两类 LLMs 上均取得最优表现，其 F1 相较于零样本提示有小幅度提升，且 Accuracy 有一定抬升。这表明，在判定标准已经得到明确界定的前提下，适度嵌入若干经严格筛选的正反“判例”，有助于模型更为清晰地勾勒 AI 与非 AI 专利的边界。基于角色的提示在多数情形下优于零样本提示，却略逊于少样本提示，说明通过赋予 LLMs“AI 专利审核者”等专业身份，可在一定程度上强化模型对任务语境与责

任边界的把握，但若缺乏足量且多样的示例支撑，其增益仍较为有限。综观三类策略，在统一判定规则与多数投票机制的共同约束之下，通过增加任务的结构化信息并引入具有代表性的示例，LLMs 的判别性能以一种平滑且可解释的方式稳步提升。

此外，不同模型对语料和文本体裁的适应性差异也值得关注。在预训练语言模型组内，基于中文语料训练的 ChineseBERT 显著优于主要基于英文科学文献训练的 SciBERT，这凸显了领域适配的重要性。这一发现为资源受限场景下的模型选择提供了重要参考：当无法使用大语言模型时，选择与目标任务语种一致、领域相关的预训练模型是较为理想的替代方案。同时，传统机器学习方法虽然在绝对性能上不占优势，但其良好的可解释性和较低的计算成本，使其在模型诊断、误差分析和快速原型开发等场景中仍具有独特的应用价值。

在此基础上，本文进一步借助混淆矩阵刻画基于 LLMs 识别方案的错误结构，如图 2 所示。结果表明，各模型在误报与漏识之间呈现出相对均衡的误差约束：在少样本提示条件下，模型在维持较低误报水平的同时进一步收敛漏识规模，显示对称式判例的嵌入有助于其更为稳健地锚定判别边界，从而减弱边界样本的偶然摇摆；相较而言，其余提示策略下的误差主要体现为漏识略有抬升，但该幅度总体仍属可控，并未实质削弱模型对真实 AI 专利的主体覆盖。总体来看，基于 LLMs 的识别框架在样本外数据上仍能兼顾精确性与覆盖性，维持较为优异且稳健的总体性能。

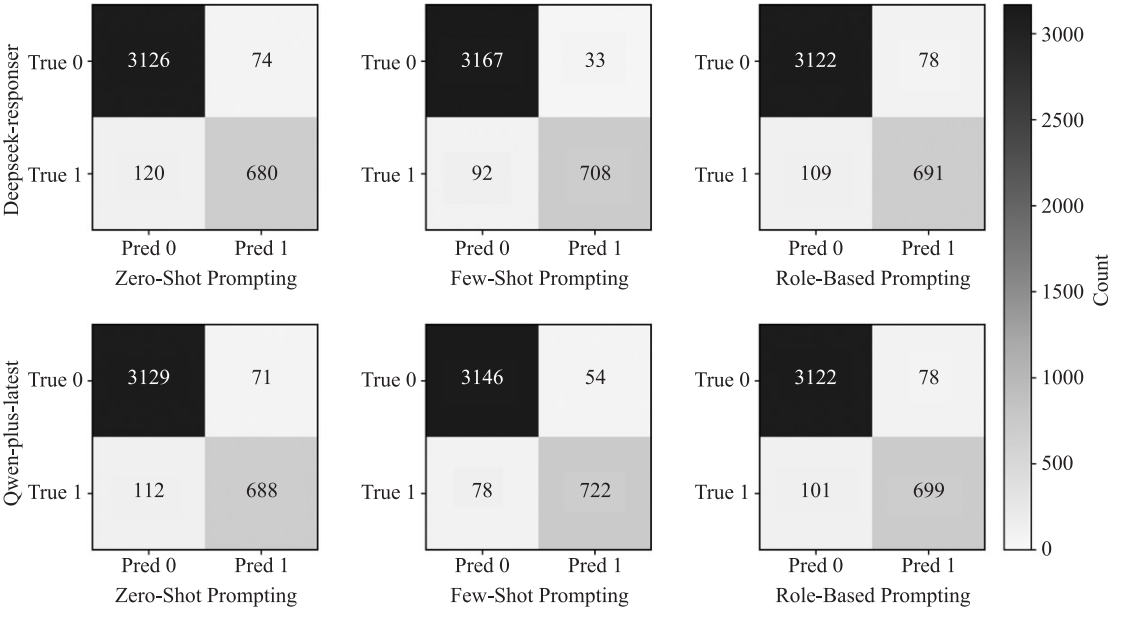


**图 2　不同提示策略下基于 LLMs 的 AI 专利识别模型混淆矩阵**

### （二）标注样本收缩下的性能表现

实证研究中，高质量的数据标注往往成本高昂且规模受限，“小样本、高标注成本”已非个案而近乎常态。为检验模型在此约束条件下的可用性与稳健性，本文进一步通过设计对比实验，考察在可用训练数据逐步减少时，各模

型性能的演化轨迹及其差异。LLMs 仅在少样本提示下需要依托已标注数据，因此本文选取两种基于少样本提示策略的 LLMs，并结合前文结果，从传统机器学习与深度学习两组方法中分别挑选表现较优的 Random Forest 与 ChineseBERT 作为参照模型。随后，将可用标注训练集自 6000 条依次压缩至 4000 条、2000 条与 1000 条，考察四类模型性能指标上的变动（见表 4）。

**表 4　　标注样本收缩下的性能表现**

| 模型名称 | 训练集规模 | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|---|
| Random Forest | 6000 | 0.798 | 0.718 | 0.815 | 0.763 |
| ChineseBERT | | 0.840 | 0.765 | 0.865 | 0.812 |
| DeepSeek – reasoner（Few – Shot Prompting） | | 0.938 | 0.912 | 0.935 | 0.923 |
| Qwen – plus – latest（Few – Shot Prompting） | | 0.898 | 0.898 | 0.936 | 0.917 |
| Random Forest | 4000 | 0.754 | 0.671 | 0.755 | 0.711 |
| ChineseBERT | | 0.804 | 0.741 | 0.785 | 0.762 |
| DeepSeek – reasoner（Few – Shot Prompting） | | 0.934 | 0.878 | 0.970 | 0.922 |
| Qwen – plus – latest（Few – Shot Prompting） | | 0.895 | 0.910 | 0.916 | 0.913 |
| Random Forest | 2000 | 0.686 | 0.592 | 0.690 | 0.637 |
| ChineseBERT | | 0.698 | 0.608 | 0.690 | 0.646 |
| DeepSeek – reasoner（Few – Shot Prompting） | | 0.930 | 0.873 | 0.965 | 0.917 |
| Qwen – plus – latest（Few – Shot Prompting） | | 0.895 | 0.906 | 0.920 | 0.913 |
| Random Forest | 1000 | 0.574 | 0.473 | 0.560 | 0.513 |
| ChineseBERT | | 0.560 | 0.459 | 0.555 | 0.502 |
| DeepSeek – reasoner（Few – Shot Prompting） | | 0.936 | 0.908 | 0.935 | 0.921 |
| Qwen – plus – latest（Few – Shot Prompting） | | 0.891 | 0.912 | 0.906 | 0.909 |
| DeepSeek – reasoner（Role – Based Prompting） | 0 | 0.875 | 0.911 | 0.872 | 0.891 |
| Qwen – plus – latest（Role – Based Prompting） | | 0.912 | 0.906 | 0.879 | 0.892 |

实验结果呈现出较明显的性能分化趋势。就 F1 而言，随着标注样本量从 6000 条降至 1000 条，Random Forest 模型的 F1 从 0.763 下降至 0.513，ChineseBERT 模型从 0.812 跌至 0.502，表明两者在标注稀缺环境下均表现出显著的性能塌陷。相比之下，两类 LLMs 在同一标注梯度下始终将 F1 维持在较高水平，准确率也持续接近或高于 0.89。尤其是，当标注规模缩减为零、模型完全无需训练样本时，通过角色提示机制调优的两种 LLMs 依然明显优于监督式模型在不同标注规模下的表现。

综合上述证据，可以得出如下结论：一方面，LLMs 对标注数据量的边际依赖性明显低于传统监督学习方法，其识别能力主要来源于预训练阶段获得的语义知识及结构化提示机制，而非对有限标注样本的高度拟合；另一方面，在“小样本、高标注成本”情境下，LLMs 不仅在标注相对充足时表现更优，即使在标注极为稀缺甚至完全缺失时，仍能保持有效的识别性能，相比监督学习方法表现出更强的数据效率优势。

### （三）提示扰动下的结果敏感性分析

在本文所构建的识别框架中，模型架构、参数设定与训练数据均保持恒定，可变的因素

是提示文本本身。换而言之，相较于传统监督式机器学习模型中的“变量选择”“函数形式”等方法自由度，LLMs 场景下最大可变空间恰恰体现在提示工程之上（Li et al. , 2024）。若在数据、定义与判定规则以及提示工程策略不变的前提下，模型输出对某一条具体提示的偶然措辞高度敏感，则据此构造的度量不够稳健，其可复现性与可审计性较差（Goli & Singh, 2024）。为此，对合理提示扰动的敏感性分析是评估 LLMs 标注工具效度的重要一环（Carlson & Burbano, 2025; Ampel et al. , 2025）。

鉴于前述对比实验，采用少样本提示的 DeepSeek – reasoner 在各候选方法中整体性能最优，本文据此将其确立为基准模型与基准提示策略。在此基础上，首先构造一条“元提示”，在元提示中对判别任务的定义、AI 专利的纳入与排除标准以及输出格式予以严格固定，仅将语序、表达结构与措辞风格作为可变因素。为避免由同一模型自生成提示而导致的风格收敛与偏好固化，并引入相对外生的语言扰动，本文调用异于主模型的 LLMs（ChatGPT – 5. 1 与 Claude Sonnet – 4. 5），依照该元提示自动生成 20 条具体提示文段，形成一组在任务含义与决策准则上等价但在语言组织与指令结构上系统变化的“提示族”。

在获得上述提示族后，本文在样本规模、提示策略（少样本提示）、阈值设定以及多数投票机制等条件完全不变的前提下，逐一将 20 条提示应用于 DeepSeek – reasoner，对同一训练集进行学习并在同一测试集上完成 AI 专利判别任务。该设计实质上是将提示工程视作一种可显式控制的方法自由度，即在模型与数据固定的前提下，仅沿“合理提示扰动”这一维度进行低幅度探索，以考察模型性能对提示表述的敏感性，并由此评估所构建识别框架的内部一致性与方法效度。图 3 展示了 20 次独立测试下的 F1 值与 Accuracy 分布。

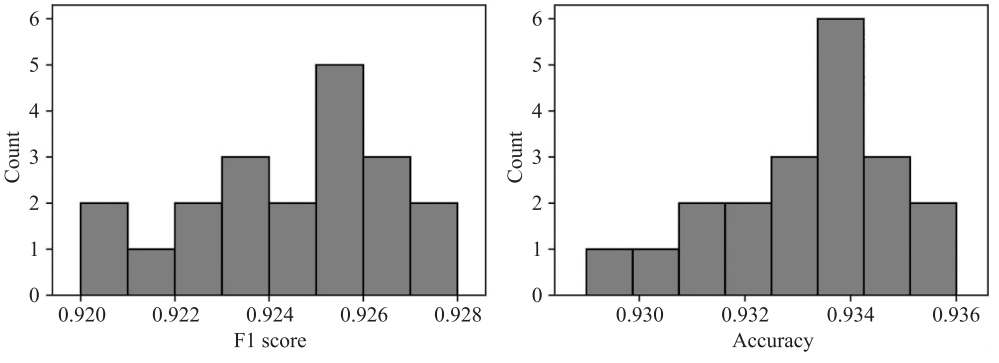


**图 3　提示族扰动下 DeepSeek – reasoner（Few – Shot Prompting）模型性能指标的分布特征**

以上结果表明，本研究所构建的识别模型在全部 20 种提示配置下均表现出良好的稳定性。DeepSeek – reasoner 模型的 F1 值始终围绕基准提示在 0. 005 范围内波动，Accuracy 的变化幅度也在 0. 004 以内。这表明，模型的分类效果并非依赖于特定提示的措辞表达，而是在一组语义同构但结构多样化的提示设计下均能保持结果一致。因此，该识别框架在方法学上具

有较好的稳健性和可推广性。

## 六、研究总结与意义

### （一）研究总结

本研究立足于人工智能技术快速发展背景下专利识别的理论需求与实践挑战，针对传统方法在应对 AI 技术快速迭代、文本复杂度高等特性时的不足，系统探讨了大语言模型在此领域的应用价值与实现路径。研究首先揭示了 AI 专利区别于传统技术领域专利的独特性：技术术语动态演进、创新内容分散于长文本、技术边界持续扩展。这些特征使得基于固定规则或传统机器学习的方法面临显著挑战。在此背景下，本研究验证了大语言模型在理解和处理此类复杂文本方面的独特优势——其强大的语义理解能力和上下文把握水平，使其能够有效捕捉 AI 专利中的核心创新要素。

通过构建“规则—证据—结论”一体化的识别框架，并在此基础上系统考察零样本提示、少样本提示与基于角色的提示三类策略，本文实现了大语言模型在 AI 专利识别任务中的可控嵌入与精细调节。实证结果显示，在统一判定标准与多数投票机制的约束下，少样本提示整体表现最佳，基于角色的提示次之，零样本提示则提供了标注成本最低但仍具竞争力的基线，由此勾勒出一条从“最低标注依赖”到“最高识别性能”的连续配置谱系。进一步的训练集样本收缩分析表明，与随机森林、ChineseBERT 等对照模型在标注样本减少时性能明显坍塌不同，两类大语言模型在训练数据大幅缩减乃至缺失的条件下，仍能维持较高水平的 Accuracy 与 F1，凸显出其对人工标注依赖度较低、适用于标注资源稀缺但任务复杂情境的特征。

### （二）研究意义

本研究在 AI 专利识别及管理实证研究方法方面具有重要的方法论贡献。Bergh 等（2022）提出，方法论贡献的价值在于其能提升研究过程的严谨性、透明度与可复核性，以增强研究过程和结果的可信性，或以更严格的标准重新验证已有结论，推动知识的累积。为系统评估方法论的贡献，Bergh 等（2022）提出一个包含两个互补维度的分析框架：一是基于方法生命周期的视角，考察方法是否实现引介与转移、改进与扩展、适用边界的澄清以及实践经验的沉淀与传播；二是结合对既有方法的创新程度与潜在受众广度，认为方法的创新程度越高、适用范围越广，其方法论贡献就越大。基于此框架，本文的贡献主要体现在以下三个方面。

第一，在“改进与扩展”与“适用边界的澄清”方面，本研究对既有专利识别方法体系进行了系统性的比较分析，在中文专利文本识别任务中验证了传统方法与机器学习方法的性能差异，从而扩展了现有方法比较的维度和深度。研究进一步追溯了性能差异产生的结构性根源，揭示了方法假设与文本生成机制之间的内在张力，并系统识别了不同方法容易出现误判与漏判的情景，从而在中文专利识别领域明确了传统方法与机器学习方法各自的适用范围与优劣，完成了对两类方法适用边界的澄清。

第二，在“引介与转移”与“实践经验的沉淀与传播”方面，本研究将生成式大语言模型这一新兴技术工具引介到专利识别这一特定管理学实证研究场景，实现了新兴技术方法跨

学科的转移与适配，并基于专利文本的语义结构与任务需求，重新设计了适应性的推理框架。针对大语言模型应用中普遍存在的“提示工程黑箱化”与“结果不可复核”问题，本研究通过将“规则—证据—结论”的推理链条标准化、提示模板系统化、验证策略多元化（零样本/少样本/角色提示）以及引入多数投票等稳健性机制，把原本依赖研究者个体经验的提示构建过程，转化为可明确描述、可重复操作、可第三方审阅的标准化流程。这一过程不仅沉淀了关键操作经验，更形成了可被后续研究使用与拓展的框架，实现了研究方法的传播。

第三，从方法创新幅度与受众广度来看，本文提出的基于生成式人工智能的 AI 专利识别方法，在实证数据上实现了相较于现有方法的显著性能提升，且在标注数据稀缺时仍保持稳健，从而有效降低了对大规模标注数据的依赖，拓展了研究方法的可行性边界。此外，该框架具备良好的跨任务迁移能力，可延伸至其他非结构化文本分类与测度任务，为数据稀缺条件下的高效准确识别提供了技术路径。

总之，本研究不仅在方法层面构建了一套基于 LLMs、可复现且具有强适应性的 AI 专利识别框架，更重要的是，通过系统的生命周期视角，实现了对现有专利识别方法体系的改进、边界澄清与技术引介，显著提升了复杂文本识别任务的高效性与可靠性。同时，该方法展现出优良的性能稳定性和跨场景迁移能力，不仅为 AI 专利识别提供了系统化的解决方案，也为更广泛的非结构化文本测度与创新管理研究奠定了方法基础，从而推动了大语言模型技术与管理学研究的深入融合。同时，本研究存在以下问题有待研究。首先，基于中文专利数据的验证结论需要在多语言环境下进一步检验。其次，现有技术方案在参数优化和提示工程设计方面仍有提升空间。最后，面对人工智能技术的持续演进，如何建立动态适应机制以确保系统的长期有效性，是未来研究的重要方向。

接受编辑：贾良定
收稿时间：2025 年 10 月 8 日
接收时间：2026 年 3 月 11 日

**作者简介**

王新成，同济大学经济与管理助理教授，同济大学城市高质量发展与规划决策实验室；上海交通大学安泰经济与管理学院博士，主要研究方向为创新管理、供应链管理和数字创新，在《中国软科学》、*MIS Quarterly*、*Journal of Operations Management* 等刊物发表论文。

郭彧铭（通讯作者，E - mail：guoyuming@sjtu. edu. cn），现为上海交通大学安泰经济与管理学院博士生。主要研究方向为技术创新管理、战略管理。

李垣，上海交通大学安泰经济与管理学院讲席教授，西安交通大学管理学院博士，主要研究方向为创新管理，在《科研管理》、*Academy of Management Journal*、*Journal of Operations Management* 等刊物发表论文。

于晓宇，上海大学管理学院教授，上海交通大学安泰经济与管理学院博士，主要研究方向为数智创业、创新与战略管理，在《管理世界》《管理科学学报》、*Long Range Planning* 等刊物发表论文。

## 参考文献

［1］王霄、董峰、韩雪亮：《家族控制权对创新开放度的影响——基于社会情感财富理论的研究》，《管理学季刊》，2022 年第 7 卷。

［2］尤树洋、卢芳妹、贾良定：《组织德行如何驱动探索式创新——基于中国情境的实证研究》，《经济管理》，2023 年第 45 卷。

［3］喻理、王淑婷、李凡等：《劳动节约型创新、买方垄断与超额利润——兼论自动化背景下的工资支付保障机制》，《管理世界》，2025 年第 41 卷。

［4］Aceves, P., & Evans, J. A. 2024. Mobilizing conceptual spaces: How word embedding models can inform measurement and theory within organization science. *Organization Science*, 35: 788 – 814.

［5］Ampel, B. M., Yang, C. – H., Hu, J., & Chen, H. 2025. Large language models for conducting advanced text analytics information systems research. *ACM Transactions on Management Information Systems*, 16: 1 – 27.

［6］Arts, S., Cassiman, B., & Hou, J. 2023. Position and differentiation of firms in technology space. *Management Science*, 69: 7253 – 7265.

［7］Babina, T., Fedyk, A., He, A., & Hodson, J. 2024. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151: 103745.

［8］Bergh D D, Boyd B K, Byron K, et al. What constitutes a methodological contribution? ［J］. *Journal of Management*, 2022, 48 (7): 1835 – 1848.

［9］Carlson, N. A., & Burbano, V. 2025. The use of LLMs to annotate data in management research: Foundational guidelines and warnings. *Strategic Management Journal*, forthcoming.

［10］Choudhury, P., Wang, D., Carlson, N. A., et al. 2019. Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal*, 40: 1705 – 1732.

［11］Cockburn, I. M., Henderson, R., & Stern, S. 2018. The impact of artificial intelligence on innovation: An exploratory analysis. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), The economics of artificial intelligence: An agenda: 115 – 146. Chicago, IL: University of Chicago Press.

［12］Cole, R. 2024. Inter – rater reliability methods in qualitative case study research. *Sociological Methods & Research*, 53: 1944 – 1975.

［13］Damioli, G., Van Roy, V., Vértesy, D., et al. 2024. Drivers of employment dynamics of AI innovators. Technological Forecasting and Social Change, 201: 123249.

［14］Dathathri, S., See, A., Ghaisas, S., et al. 2024. Scalable watermarking for identifying large language model outputs. *Nature*, 634: 818 – 823.

［15］Ebrahimi, M., Chai, Y., Samtani, S., et al. 2022. Cross – lingual cybersecurity analytics in the international dark web with adversarial deep representation learning. *MIS Quarterly*, 46: 1209 – 1226.

［16］Gentzkow, M., Kelly, B., & Taddy, M. 2019. Text as data. *Journal of Economic Literature*, 57: 535 – 574.

［17］Giczy, A. V., Pairolero, N. A., & Toole, A. A. 2022. Identifying artificial intelligence (AI) invention: A novel AI patent dataset. *The Journal of Technology Transfer*, 47: 476 – 505.

［18］Goli, A., & Singh, A. 2024. Frontiers: Can large language models capture human preferences? *Marketing Science*, 43: 709 – 722.

［19］Gong, H., Zou, H., Liang, X., et al. 2025. A global dataset mapping the AI innovation from academic research to industrial patents. *Scientific Data*, 12: 1261.

［20］Hou, Y., Png, I. P., & Xiong, X. 2023. When stronger patent law reduces patenting: Empirical evi-

dence. *Strategic Management Journal*, 44: 977 – 1012.

[21] Jiang, L., & Goetz, S. M. 2025. Natural language processing in the patent domain: A survey. *Artificial Intelligence Review*, 58: 214.

[22] Kamateri, E., Salampasis, M., & Diamantaras, K. 2023. An ensemble framework for patent classification. *World Patent Information*, 75: 102233.

[23] Kriesch, L., & Losacker, S. 2024. A global patent dataset of bioeconomy – related inventions. *Scientific Data*, 11 (1): 1308.

[24] Lafond, F., & Kim, D. 2019. Long – run dynamics of the US patent classification system. *Journal of Evolutionary Economics*, 29: 631 – 664.

[25] LeCun, Y., Bengio, Y., & Hinton, G. 2015. Deep learning. *Nature*, 521: 436 – 444.

[26] Li, P., Castelo, N., Katona, Z., & Sarvary, M. 2024. Frontiers: Determining the validity of large language models for automated perceptual analysis. *Marketing Science*, 43: 254 – 266.

[27] Li, S., Hu, J., Cui, Y., et al. 2018. DeepPatent: Patent classification with convolutional neural networks and word embedding. *Scientometrics*, 117: 721 – 744.

[28] Mansfield, E. 1986. Patents and innovation: An empirical study. *Management Science*, 32: 173 – 181.

[29] Masclans, R., Hasan, S., & Cohen, W. M. 2025. Measuring the commercial potential of science. *Strategic Management Journal*, 46: 2199 – 2236.

[30] Miric, M., Jia, N., & Huang, K. G. 2023. Using supervised machine learning for large – scale classification in management research: The case for identifying artificial intelligence patents. *Strategic Management Journal*, 44: 491 – 519.

[31] Qaiser, S., & Ali, R. 2018. Text mining: Use of TF – IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 181: 25 – 29.

[32] Rathje, J., Katila, R., & Reineke, P. 2024. Making the most of AI and machine learning in organizations and strategy research: Supervised machine learning, causal inference, and matching models. *Strategic Management Journal*, 45 (10): 1926 – 1953.

[33] Shanahan, M., McDonell, K., & Reynolds, L. 2023. Role play with large language models. *Nature*, 623: 493 – 498.

[34] Singhal, K., Azizi, S., Tu, T., et al. 2023. Large language models encode clinical knowledge. *Nature*, 620: 172 – 180.

[35] Verendel, V. 2023. Tracking artificial intelligence in climate inventions with patent data. *Nature Climate Change*, 13: 40 – 47.

[36] Wu, L., Lou, B., & Hitt, L. M. 2025. Innovation strategy after IPO: How AI analytics spurs innovation after IPO. *Management Science*, 71: 2360 – 2389.

[37] Xie, J., Liu, X., Zeng, D. D., et al. 2022. Understanding medication nonadherence from social media: A sentiment – enriched deep learning approach. *MIS Quarterly*, 46: 341 – 372.

[38] Yun, J., & Geum, Y. 2020. Automated classification of patents: A topic modeling approach. *Computers & Industrial Engineering*, 147: 106636.

[39] Zhang, D., Zhou, L., Tao, J., et al. 2025. KETCH: A knowledge – enhanced transformer – based approach to suicidal ideation detection from social media content. *Information Systems Research*, 36: 572 – 599.

[40] Zhou, L., Schellaert, W., Martínez – Plumed, F., Moros – Daval, Y., Ferri, C., & Hernández – Orallo, J. 2024. Larger and more instructable language models become less reliable. *Nature*, 634: 61 – 68.