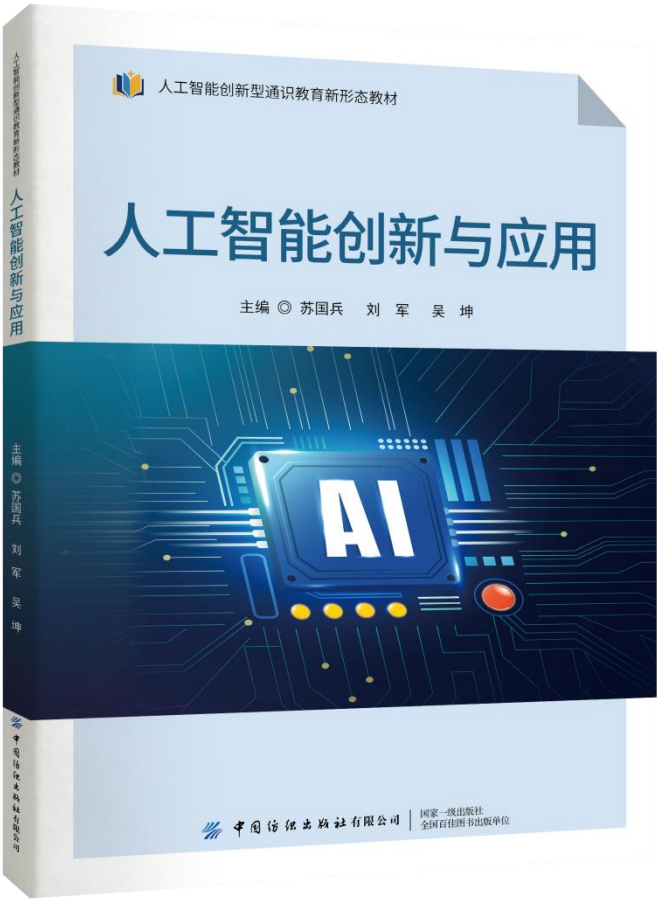
人工智能创新与应用



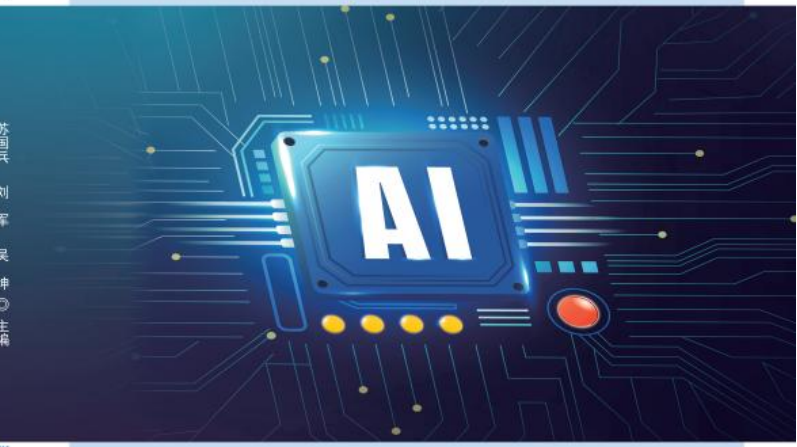
类目：人工智能类
书名：人工智能创新与应用
主编：苏国兵 刘 军 吴 坤
出版社：中国纺织出版社
开本：大 16 开
书号：978-7-5229-4163-9
使用层次：通用
出版时间：2026 年 10 月
定价：55.80 元
印刷方式：四色
是否有资源：有

人工智能创新型通识教育新形态教材

人工智能创新与应用

苏国兵　刘　军　吴　坤◎主编





中国纺织出版社有限公司
国家一级出版社
全国百佳图书出版单位

ISBN 978-7-5229-4183-9
定价：55.80元

中国纺织出版社官方微博

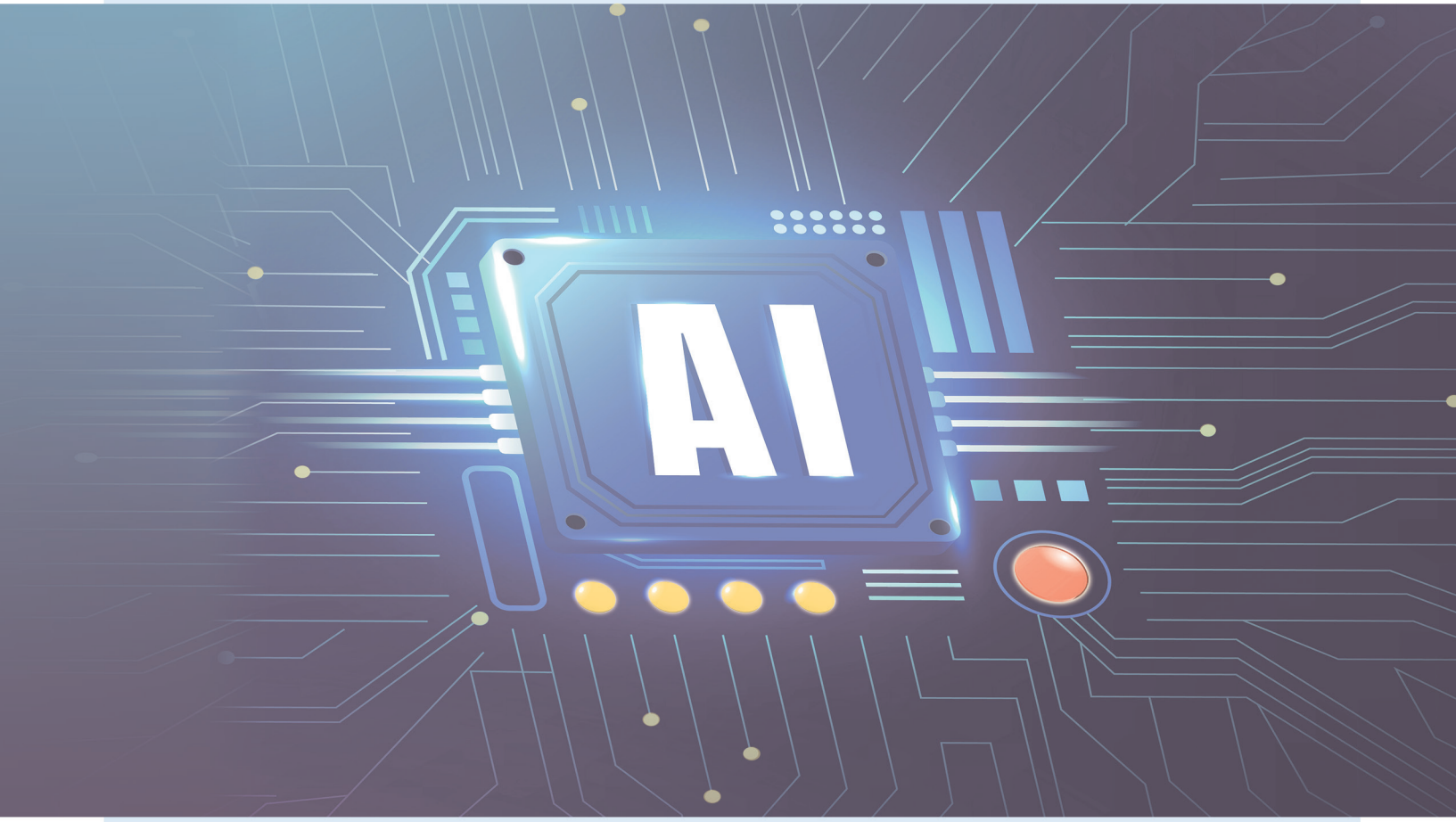
中国纺织出版社官方微信

人工智能创新与应用
苏国兵　刘　军　吴　坤◎主编
中国纺织出版社有限公司

人工智能创新型通识教育新形态教材

人工智能创新与应用

苏国兵　刘　军　吴　坤 ◎ 主编



中国纺织出版社有限公司

内 容 提 要

本书围绕人工智能基本理论及相关应用创新展开系统阐述，旨在帮助读者全面掌握人工智能相关知识与应用技能。

全书共分为11章，从人工智能的基础概念切入，阐述了人工智能通识和AI驱动的现代创新方法论相关知识；分章节介绍了通用对话大模型创新应用，基于RAG的智能知识库创新，AI多模态绘画与视觉内容创新，AI智能视频生成与剪辑创新，AI虚拟数字人创新，AI智能体自主任务创新，AI智慧教育个性化创新，AI智慧零售精准营销创新和AI工业视觉创新等具体创新实践相关知识。

本书从基础开始，由浅入深进行学习，可帮助读者逐步理清人工智能的核心技术和发展趋势。

本书既可以作为培养人工智能应用型人才的课程配套教材，也适用于人工智能初学者和对人工智能基础理论感兴趣的广大读者阅读参考。

图书在版编目(CIP)数据

人工智能创新与应用 / 苏国兵，刘军，吴坤主编. 北京：中国纺织出版社有限公司，2026. 10. --ISBN 978-7-5229-4163-9

Ⅰ. TP18

中国国家版本馆CIP数据核字第2026D0A241号

责任编辑：郭　婷　周　航　责任校对：王蕙莹　责任印制：储志伟

中国纺织出版社有限公司出版发行

地址：北京市朝阳区百子湾东里A407号楼　邮政编码：100124

销售电话：010-87155895　传真：010-87155801

http://www.c-textilep.com

中国纺织出版社天猫旗舰店

官方微博 http://weibo.com/2119887771

三河市祥达印刷包装有限公司　各地新华书店经销

2026年10月第1版第1次印刷

开本：210×285　1/16　印张：17

字数：390千字　定价：55.80元

凡购本书，如有缺页、倒页、脱页，由本社图书营销中心调换

《人工智能创新与应用》

编委会

主　编　苏国兵　刘　军　吴　坤

副主编　冯　倩　周　洁



PREFACE

前言

随着 ChatGPT 等生成式人工智能产品进入公众视野，人工智能技术展示了其强大的力量，DeepSeek 等国产大模型的发展更是催生了整个社会关注人工智能的热潮。今天，人工智能正以惊人的速度渗透到社会生产生活的各个领域，从文本创作、图像设计到视频制作、音频生成，再到智能化办公场景的应用，人工智能正在重塑我们的工作与学习方式，我们已经切身感受到了人工智能对人们的思维模式和生活方式的改变。人工智能技术必将对人类社会生产生活产生重大而深远的影响，掌握人工智能相关知识与技能，已成为新时代人才不可或缺的核心素养。

党的二十大报告提出要实施科教兴国战略，强化现代化建设人才支撑，并强调要深化教育领域综合改革，加强教材建设和管理。为响应党中央号召，深入贯彻全国教育大会和《教育强国建设规划纲要(2024—2035 年)》精神，更好构建适应社会需求的人工智能通识课程教育体系，我们在充分进行调研和论证的基础上，精心编写了这本《人工智能创新与应用》教材。

与同类书相比，本书具备以下特点：

循序渐进，由浅入深

本书为人工智能的入门教材，以人工智能的整个处理流程为主线，逐一介绍了人工智能所涉及的相关技术及其创新应用，帮助读者形成对人工智能知识体系及其应用领域的轮廓性认识，为读者在人工智能领域的继续深造学习奠定基础。

技能教育与思政教育紧密结合

在讲解人工智能专业知识的同时，紧密结合思政教育主旋律，立足专业知识角度潜移默化引导学生提升思政综合素养。

实操性强，注重应用

通过大量贴近应用的创新实践案例，强化学习者的动手能力，助力所学知识能够直接应用于实际工作场景。

本书旨在为各类院校“人工智能导论”“人工智能应用实务”等相关课程提供清晰的教学指引，帮助教师有序开展教学活动，引导学生循序渐进地掌握人工智能技术的核心要点与应用方法。

由于编者能力有限，书中难免存在不妥之处，恳请读者朋友提出宝贵意见，不胜感激。

苏国兵　刘军　吴坤

2026 年 7 月

目录

CONTENTS

第 1 章　人工智能通识进阶 / 1

1.1　技术迭代：人工智能发展与范式升级 / 1

1.2　核心体系：前沿人工智能技术架构 / 4

1.3　规范准则：AI 创新伦理与合规体系 / 7

1.4　本章实训：AI 模型能力边界测试与分析 / 9

第 2 章　AI 驱动的现代创新方法论 / 11

2.1　方法革新：传统创新模式的 AI 赋能改造 / 11

2.2　思维升级：AI 时代创新核心思维逻辑 / 15

2.3　实施评估：AI 辅助创新需求与可行性分析 / 20

2.4　本章实训：场景化创新问题定义实践 / 27

第 3 章　通用对话大模型创新应用 / 29

3.1　问题导入：传统任务型对话系统的应用瓶颈 / 29

3.2　探索分析：大模型场景价值与创新空间 / 30

3.3　技术支撑：大模型原理与对话交互技术 / 34

3.4　实践应用：提示词设计与智能内容生成 / 37

3.5　评价与反思：模型缺陷与应用创新优化 / 46

3.6　本章实训：对话大模型应用能力测试与优化 / 50

第 4 章 基于 RAG 的智能知识库创新 / 53

4.1 问题导入：传统文档知识管理的复用难题 / 53
4.2 探索分析：RAG 技术价值与知识库创新场景 / 54
4.3 技术支撑：向量检索与检索增强生成技术 / 57
4.4 实践应用：私有知识库搭建与智能问答优化 / 59
4.5 评价与反思：检索缺陷与知识生态创新 / 65
4.6 本章实训：RAG 知识库搭建与问答效果测试 / 68

第 5 章 AI 多模态绘画与视觉内容创新 / 71

5.1 问题导入：传统视觉创作的效率与创意局限 / 71
5.2 探索分析：AIGC 视觉产业发展与创新机遇 / 72
5.3 技术支撑：扩散模型与多模态生成技术 / 74
5.4 实践应用：AI 视觉创作与二次优化设计 / 78
5.5 评价与反思：创作伦理与视觉创新升级 / 85
5.6 本章实训：AI 绘画全流程创作与优化 / 89

第 6 章 AI 智能视频生成与剪辑创新 / 91

6.1 问题导入：传统视频制作的效率门槛问题 / 91
6.2 探索分析：AI 视频生产变革与轻量化创新场景 / 92
6.3 技术支撑：视频生成与智能剪辑技术 / 94
6.4 实践应用：AI 全流程视频制作与优化 / 98
6.5 评价与反思：内容缺陷与视频创新升级 / 108
6.6 本章实训：AI 视频全流程制作 / 110

第 7 章 AI 虚拟数字人创新 / 113

7.1 问题导入：真人商业服务的成本与时效短板 / 113
7.2 探索分析：数字人商业化场景与发展趋势 / 113
7.3 技术支撑：数字人建模与实时交互技术 / 117
7.4 实践应用：数字人场景方案设计与制作 / 121
7.5 评价与反思：交互缺陷与合规创新优化 / 129
7.6 本章实训：数字人形象定制与应用 / 131

第 8 章 AI 智能体自主任务创新 / 133

8.1 问题导入：传统 AI 被动交互的功能局限 / 133
8.2 探索分析：自主智能演进与智能体创新场景 / 134
8.3 技术支撑：任务拆解与智能协同技术 / 138
8.4 实践应用：旅游智能体设计与搭建 / 141
8.5 评价与反思：决策缺陷与智能创新趋势 / 163
8.6 本章实训：智能体任务拆解与搭建 / 165

第 9 章 AI 智慧教育个性化创新 / 167

9.1 问题导入：传统同质化教学的育人短板 / 167
9.2 探索分析：智慧教育发展与 AI 创新切入点 / 167
9.3 技术支撑：学情画像与智能助教技术 / 171
9.4 实践应用：个性化教学方案与工具设计 / 175
9.5 评价与反思：教育风险与 AI 创新边界 / 198
9.6 本章实训：教学方案设计与智能助学工具搭建 / 201

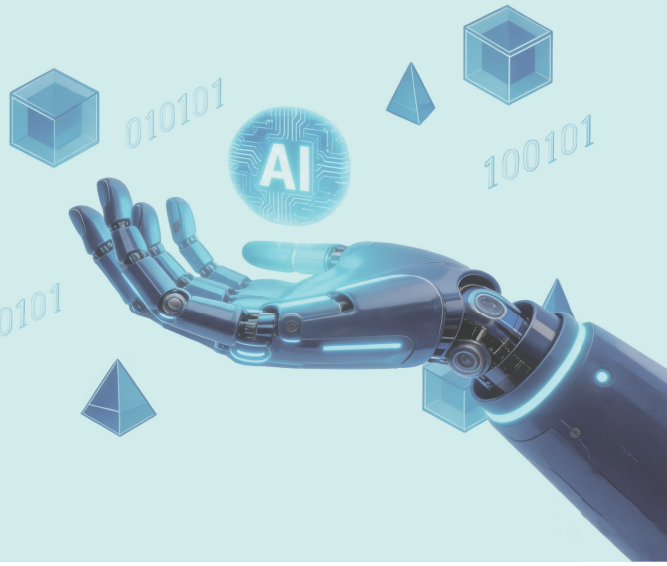
第 10 章 AI 智慧零售精准营销创新 / 203

10.1 问题导入：传统零售粗放运营的经营痛点 / 203
10.2 探索分析：零售数字化转型与 AI 增效场景 / 204
10.3 技术支撑：客流分析与智能推荐技术 / 207
10.4 实践应用：零售数据解析与营销方案优化 / 210
10.5 评价与反思：数据偏差与零售持续创新 / 231
10.6 本章实训：零售数据解析与营销方案优化 / 234

第 11 章 AI 工业视觉创新 / 237

11.1 问题导入：传统工业产品识别的精度与效率瓶颈 / 237
11.2 探索分析：智能制造转型与工业 AI 视觉创新空间 / 237
11.3 技术支撑：工业视觉智能化的关键技术 / 240
11.4 实践应用：工业产品识别方案设计与模拟部署 / 243
11.5 评价与反思：工况适配与工程创新优化 / 257
11.6 本章实训：工业产品识别方案设计与测试 / 260

参考文献 / 262



第 1 章 人工智能通识进阶

人工智能历经数十年发展，从实验室理论探索全面走向产业落地，尤其是近几年大模型、多模态、智能体等技术爆发式迭代，显著推动了技术形态、应用逻辑与创新范式的变革。本章将系统梳理人工智能的发展脉络、前沿技术架构，讲解人工智能创新必须遵守的伦理与合规准则，并配套实训内容，帮助同学们建立完整、前沿且合规的人工智能基础认知，为后续人工智能创新方法、场景应用的学习筑牢根基。



1.1 技术迭代：人工智能发展与范式升级

人工智能并非一蹴而就的新兴技术，其发展伴随技术路线、算力水平、应用需求的变化不断迭代。

1.1.1 人工智能技术迭代历程

人工智能的核心目标是让机器模拟、延伸乃至拓展人类智能。自 1956 年正式定名以来，人工智能整个发展历程呈现出波浪式前进、阶段性跃迁的特征，行业也经历过繁荣、低谷与再度崛起的多个周期，如图 1-1 所示。

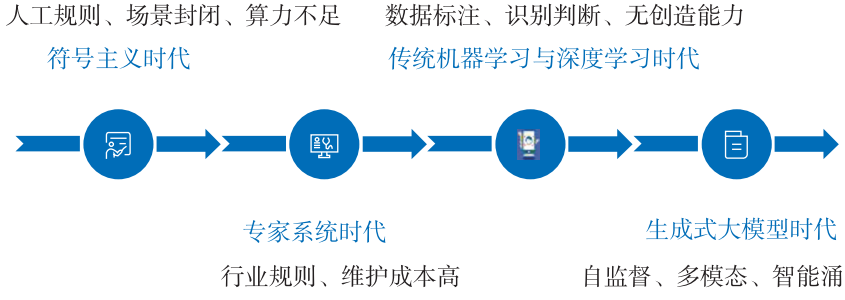


图 1-1 人工智能技术迭代历程

第一阶段为符号主义时代(20 世纪 50 年代—70 年代)

这是人工智能的萌芽期。1956 年达特茅斯会议首次提出“人工智能(Artificial Intelligence,AI)”概念,标志人工智能学科正式诞生。符号主义学派核心观点是人类认知的本质是符号运算,人类智能可拆解为标准化逻辑符号与推理规则,机器通过符号逻辑推演即可复现智能行为。当时取得了逻辑推理程序、早期对话机器人等成果,它们能够完成数学定理证明、简单问答等任务。但受限于早期算力不足,同时规则无法覆盖现实世界的模糊问题与常识场景,这一阶段的 AI 只能在封闭、简单的场景中工作,最终发展陷入停滞。

第二阶段是专家系统时代(20 世纪 80 年代—90 年代初)

这一时期研究者放弃了打造通用智能的目标,转向垂直领域应用落地,依托“知识工程”构建专家系统。技术人员将行业专家的经验、知识整理为固定规则,嵌入系统中,应用在医疗诊断、金融咨询、工业运维等场景。专家系统实现了 AI 的初步商业落地,但短板十分突出,例如,行业知识更新时需要人工逐条修改规则,维护成本极高;AI 无法自主学习新内容,面对超出规则范围的问题效果大幅下降,这也让 AI 第二次进入发展低谷。

第三阶段为传统机器学习与深度学习时代(20 世纪 90 年代—2020 年)

随着数据量增长、算力提升,AI 彻底摆脱了“人工编写规则”的模式,进入数据驱动阶段。早期以传统机器学习为主,依靠算法从标注数据中提取特征,完成垃圾邮件识别、图像分类、语音转写等任务。2012 年之后,深度神经网络迎来突破,深度学习成为主流,模型依托多层神经网络自动挖掘数据特征,图像识别、人脸识别、语音交互等新技术快速普及。这一阶段的 AI 以感知、判断、分析为核心,我们称之为判别式 AI,它能精准识别事物、预测结果,但不具备主动创造能力。

第四阶段便是生成式大模型时代(2020 年至今)

海量数据、超大算力与 Transformer(变换器)架构的结合,催生了参数规模达千亿、万亿级别的大语言模型,AI 迎来范式革命。这个时期的模型不再局限于“识别与判断”,而是具备了文本创作、图像生成、视频制作、代码编写等创造能力,同时出现多模态融合、自主智能体等延伸技术。AI 从单一工具逐步向多功能协同载体演变,全面赋能各行各业的创新活动。

1.1.2 传统 AI 与生成式 AI 范式差异

1. 传统 AI 与生成式 AI 的本质区别

在深入讨论生成式 AI 之前,需要厘清传统 AI(判别式 AI)与生成式 AI 在技术范式上的根本差异。这一差异不仅体现在技术实现层面,更决定了各自的能力边界与应用场景。

(1)传统机器学习以判别式模型为主导

判别式模型的核心任务是学习从输入数据 x 到输出标签 y 的映射关系 $P(y \mid x)$。语音识别系统将音频信号映射为文字序列,人脸识别系统将面部图像映射为身份标识,垃圾邮件分类器将邮件内容映射为“垃圾/非垃圾”两类标签。这类系统擅长在给定输入条件下做出判断或决策,但其

能力被限定在训练数据所覆盖的范围内；传统判别式模型大多遵循“一专一能”的开发思路，多数场景下单个模型仅针对单一目标任务训练，跨任务复用能力弱。

(2) 生成式模型的学习目标完全不同

生成式模型的核心任务是学习数据的概率分布 $P(x)$ 或条件分布 $P(x \mid y)$，从而能够从该分布中采样生成新的数据样本。当输入文字描述时，图像生成模型可以从学到的图像分布中采样，生成符合描述的新图像；当输入提示词时，语言模型可以从语言的概率分布中逐词采样，生成连贯的新文本。判别式模型回答“这是什么”，生成式模型回答“这可以是什么”，判别式模型专注于分类与预测，生成式模型专注于创造与合成。

这种差异的深层含义在于判别式模型只能基于已知模式进行判断，而生成式模型能够通过学习数据的内在结构创造出未曾出现过的新内容。这也是为什么生成式 AI 被广泛视为人工智能发展史上最重大的范式跃迁之一。

为直观说明两种模型的差异，以识别“这是山羊还是绵羊”为例：

判别式模型：要确定一个羊是山羊还是绵羊，判别式模型的方法是从历史数据中学习到模型，然后通过提取这只羊的特征来预测出这只羊是山羊的概率和是绵羊的概率。

生成式模型：是根据山羊的特征首先学习出一个山羊的模型，然后根据绵羊的特征学习出一个绵羊的模型，然后从这只羊中提取特征，放到山羊模型中看概率是多少，再放到绵羊模型中看概率是多少，哪个概率大就是哪个。

从技术实现路径来看，判别式模型通常采用有监督学习范式，依赖输入–输出配对数据进行训练，模型学习的是“条件概率”$P(y \mid x)$；而生成式模型则主要采用自监督学习范式，通过“预测被遮挡的词”“判断两句话是否连续”等前置任务从无标注文本中学习语言的统计规律，再通过微调适应下游任务。这种实现路径的差异，构成了范式升级的技术根基。

2. 范式升级的四个维度

(1) 数据需求维度



范式升级的四个维度

传统监督学习依赖大量人工标注的结构化数据。训练一个图像分类器需要成千上万张标注好类别的图片，标注过程耗时费力且成本高昂。生成式 AI 通过自监督学习范式，可以在预训练阶段直接利用互联网上未经标注的海量文本、图像和视频学习数据的内在规律。尽管后续的指令微调、人类偏好对齐等阶段仍然需要人工标注数据，但预训练阶段对大规模标注数据的依赖已大幅降低，这是相较于传统监督学习的显著进步。

(2) 任务泛化维度

判别式 AI 遵循“一任务一模型”的模式，如翻译模型处理翻译、识别模型处理识别，不同任务之间不存在能力共享。生成式 AI 则展现出任务泛化的特性，如同一个大语言模型既能完成文本翻译，又能生成代码、进行对话、撰写文案，甚至通过少量样本学习快速适应新任务，无须针对每个任务重新训练。

(3)交互方式维度

传统 AI 的交互以封闭式任务为主，用户提出具体问题，系统返回确定性结果。生成式 AI 则支持开放式、多轮次的交互方式，用户可以通过自然语言连续对话引导 AI 完成复杂任务，系统能够根据上下文动态调整输出策略。

(4)能力边界维度

判别式 AI 能力的上限由训练数据中已标注的知识范围决定，而生成式 AI(尤其是大语言模型)在参数规模突破特定阈值后，会涌现出未经过专门训练的能力，如类比推理、代码生成、多步逻辑推演等。这一现象被称为“智能涌现”，其机制目前仍是人工智能研究的前沿课题。

上述数据需求、任务泛化、交互方式、能力边界四个维度并非彼此孤立，而是相互关联。例如，自监督学习范式降低了对标注数据的依赖，进而使得模型能够在更多任务上实现泛化。



1.2 核心体系：前沿人工智能技术架构

当下主流 AI 创新并非单一模型独立运行，而是形成了以大模型为底座，多模态、RAG、智能体等技术协同配合的完整技术体系。

1.2.1 大模型与多模态技术基础

大语言模型(Large Language Model，LLM)是当前整个 AI 技术生态的重要基础，绝大多数对话、创作、逻辑推理类 AI 应用，都搭建在大模型之上。如 GPT 系列、文心一言、DeepSeek、讯飞星火等，都属于大语言模型。

大语言模型的核心是基于 Transformer 架构训练的巨型神经网络，训练过程可以简单理解为模型学习互联网海量文本数据中的语言逻辑、知识体系、表达习惯。

与人类的语义理解机制不同，大语言模型本质上是基于海量文本中词语间的统计关联来生成回答。关于这种统计学习是否构成真正的“语义理解”，学界存在根本性分歧。外在主义观点认为，只要模型行为上表现出理解和推理能力即可被视为具备某种形式的理解；而内在主义观点则认为，模型缺乏意向性和真实世界指涉，其“理解”只是统计模式的表象。这一问题目前仍是人工智能哲学与认知科学的前沿争议课题。

早期大语言模型仅能处理文字，也就是单模态模型。但现实世界的信息是多元的，包含文字、图片、语音、视频等多种形式，由此衍生出多模态技术。多模态的核心目标是打破不同信息形态的壁垒，让 AI 同时理解、处理、生成文本、图像、音频、视频等多种类型数据，实现“多模态融合”。

例如，单模态大模型只能看懂一段文字；而多模态模型可以看懂一张图片并配文，根据文字描述作画，听语音并整理成文档，解析视频画面并总结内容。其核心技术逻辑是模态对齐，将文字、图像、语音等不同类型的信息，转化为统一的向量表达，让模型能够在同一个逻辑空间中识

别、关联不同模态的数据，最终完成跨模态交互与生成。

在 AI 创新实践中，大模型负责核心逻辑与内容生成，多模态技术负责拓展信息交互的形式，二者结合让 AI 应用从纯文本场景，延伸到视觉、音频、视频等全场景，也是视觉创作、数字人、视频生成等创新应用的技术根基。

1.2.2　RAG 与智能体核心技术逻辑

大语言模型存在两个难以根治的固有局限：一是知识滞后，模型的知识截止于训练数据的时间，无法获取实时信息；二是模型幻觉，面对陌生知识、专业领域内容时，容易编造看似合理但完全错误的信息。同时，传统大模型只能被动接收指令、完成单一任务，不具备自主规划、分步执行复杂任务的能力。而 RAG 和智能体两项技术，恰好针对性解决了上述问题，也是企业级、场景化 AI 创新的两大核心组件。

1. RAG

检索增强生成(Retrieval-Augmented Generation，RAG)。我们可以把 RAG 理解为给大模型外接了一座专属动态知识库。它不需要对大模型本身做微调，而是在模型生成回答之前，先从外部知识库检索相关资料，把权威、实时的内容作为上下文交给大模型，引导模型基于真实资料输出答案，以此解决知识滞后、幻觉、专业领域知识不足等问题。

RAG 的完整工作流程分为三个核心环节，三者依次衔接，构成闭环(图 1-2)。



图 1-2　RAG 的完整工作流程

①检索环节：将用户的问题转化为向量，在向量数据库中匹配、检索相似度最高的文档片段。

②增强环节：把检索到的权威文档和用户原始问题整合在一起，形成完整提示词。

③生成环节：将整合后的内容输入大模型，由模型结合检索资料生成准确回答。

RAG 广泛应用于企业私有知识库、行业文档问答、政务资料查询等场景。

2. 智能体

如果说 RAG 是给大模型补充“知识”，那么 AI 智能体就是给大模型赋予“行动能力”。传统 AI 是被动交互，用户下达一条指令，AI 完成一项对应任务；而智能体是主动智能体，面对复杂、多步骤的任务时，能够自主拆解目标、规划执行步骤、调用各类工具、分步完成任务，还能根据执行结果动态调整方案。

智能体的核心运行逻辑包含任务拆解、思维链决策和工具调用三大模块。

(1)任务拆解是智能体运行的起点

指将用户提出的复杂、模糊的任务自上而下分解为若干可独立执行的子任务，并识别子任务之间的先后依赖关系，确保执行路径合理有序。

(2)思维链决策是智能体的认知核心

思维链技术引导模型在给出最终答案前，先生成一系列中间推理步骤，将复杂问题逐层推导。在智能体语境下，思维链决策具备判断每个子任务应由内部知识完成还是调用外部工具、在多步推理过程中验证中间结果的合理性、在执行异常时动态调整后续策略这三重功能。

(3)工具调用是智能体实际执行的能力基础

智能体通过结构化指令调用外部工具或 API(Application Programming Interface，应用程序编程接口)，如搜索引擎、数据库查询、代码解释器、第三方应用等，获取实时信息、执行计算操作或完成跨系统协作，并将工具返回的结果纳入后续推理与生成中。

三大模块形成“拆解→决策→执行→再决策”的动态循环(图1-3)，相互协同、缺一不可，共同构成智能体自主完成复杂任务的技术基石。

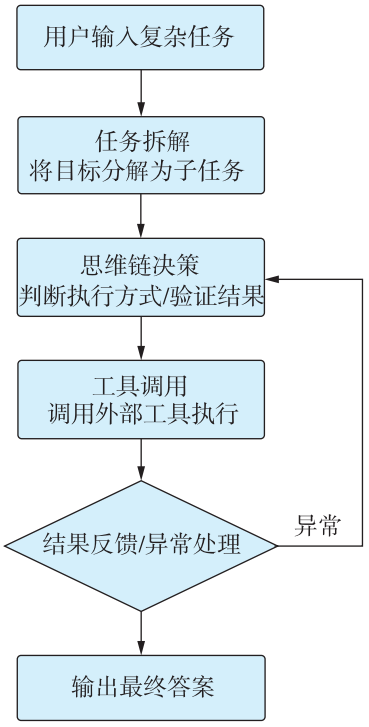


图 1-3　智能体核心运行逻辑

例如，面对“撰写一份市场调研报告”这类复杂任务，智能体不会直接输出内容，而是先拆解为“收集行业数据→整理竞品信息→分析市场趋势→撰写报告初稿→修改优化”多个子任务；再通过思维链逐一判断每个子任务的完成方式；同时自主调用文档检索、数据统计、翻译等工具，全程自主推进，最终完成整体任务。如今办公自动化、智能助理、自动化运维等创新场景，均以智能体技术为核心。

在实际创新落地中，RAG 和智能体常常组合使用，智能体负责整体任务规划与执行，RAG 为智能体提供实时、专业的知识支撑，二者与大模型、多模态技术融合，构成了当前最主流的 AI 技术组合架构，如图 1-4 所示。



图 1-4　分层总架构图



1.3 规范准则：AI 创新伦理与合规体系

技术是一把双刃剑，人工智能在释放巨大创新价值的同时，也带来了伦理风险、隐私泄露、知识产权侵权、内容违规等一系列问题。当前我国已出台多项法律法规与伦理准则，伦理先行、合规底线是所有 AI 创新活动的前提。

1.3.1　AI 应用伦理风险与约束规范

2026 年 5 月 19 日，全国网络安全标准化技术委员会（SAC/TC260，简称“网安标委”）在 2026 年中国网络文明大会人工智能赋能网络文明建设分论坛上正式发布《人工智能应用伦理安全指引 1.0》（TC260-005）。该文件系统梳理了人工智能（AI）应用可能引发的六类结构性伦理影响，并提出九项伦理安全原则。这一文件的出台，标志着我国 AI 伦理治理从宏观倡导转向落地化的具体规范。

1. 六类伦理安全影响

根据《人工智能应用伦理安全指引 1.0》的梳理，AI 应用可能在以下六个维度对社会产生深层影响。

（1）人类主导权影响

随着 AI 系统自主决策能力的增强，人类对社会运行关键环节的主导地位可能受到挑战。尤其在智能体系统能够自主完成多步骤任务的背景下，决策权归属问题变得日益紧迫。

（2）公共秩序影响

深度伪造、AI 生成虚假信息等手段的普及，可能削弱社会运行所依赖的信息真实性基础。当任何人都能以较低成本生成难以辨伪的合成内容时，社会信任机制将面临重大考验。

（3）个体认知与社会价值影响

过度依赖 AI 可能导致个体认知能力的退化，或使个体与现实社会脱节。该问题在教育领域尤为突出，如学生若长期依赖 AI 完成学习任务，可能损害学习的真实性与创造性。

（4）社会分化和歧视影响

算法模型的训练数据中如果蕴含社会偏见（包括文化、性别、地域、阶层等维度），这些偏见可能被模型放大并通过决策输出固化。在招聘筛选、信贷审批、刑事司法等场景中，算法偏见可能造成实质性不公。

（5）生命健康与基本权益影响

AI 在医疗诊断、自动驾驶、公共安全等领域的应用直接关系人的生命安全和基本权益。错误决策的后果可能极为严重，且责任认定面临复杂的法律挑战。

(6) 可持续生态影响

大规模 AI 模型训练、推理运行会消耗大量电力与水资源。随着 AI 应用规模化普及，技术研发与落地的生态环境成本应当纳入技术全生命周期评估框架。

2. 九大伦理安全原则

针对上述风险，《人工智能应用伦理安全指引 1.0》提出了九项原则，包括增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明、保护隐私安全、确保可控可信、敏捷共治、普惠共享。这意味着高校应加强 AI 伦理教育，将普惠共享理念融入课程设计与人才培养全过程。

1.3.2 AI 创新合规与知识产权要求

在开展 AI 创新项目、开发 AI 应用、使用 AI 生成作品时，除了遵守伦理准则，还必须厘清合规要求与知识产权边界。

1. AI 创新合规

AI 创新的整体合规要求，覆盖数据使用、模型调用、服务上线三大环节，如图 1-5 所示。



图 1-5 AI 创新合规要求

(1) 数据使用合规要求

在数据使用上，训练、投喂给 AI 的数据集、文档、素材必须具备合法来源，禁止使用盗版数据、非法爬取的网络数据、涉密数据；涉及第三方版权内容、个人隐私数据时，必须获得授权。

(2) 模型调用合规要求

在模型调用上，使用开源模型、商用 API 时，必须严格遵守对应平台的用户协议，区分个人非商用、商用场景，商用场景需按照规定付费、备案。

(3) 服务上线合规要求

在服务上线上，若开发的 AI 应用面向公众提供服务，尤其是生成式 AI 服务，需要按照国家要求完成安全评估、备案，落实内容审核机制。

2. 知识产权问题

知识产权问题是当前 AI 创新领域争议最多、风险最高的部分。

(1)训练数据的知识产权

AI 模型训练依赖海量文本、图片、音视频数据，未经版权方授权，擅自使用受著作权保护的作品训练模型，可能构成侵权行为。

(2)AI 生成内容的著作权归属

根据我国现行法规与司法实践：由人类主导构思、设定创作框架，仅借助 AI 工具辅助完成的内容，若满足独创性要求，用户依法享有著作权；完全由 AI 自动生成、无人类创造性投入的内容，不构成著作权法意义上的作品。因此，在使用 AI 辅助创作文案、设计画作、编写方案时，应当如实标注 AI 辅助情况，不得将 AI 生成内容冒充为完全原创作品。

(3)开源模型的使用合规

大量开源大模型、视觉模型附带开源协议，不同协议的约束不同，部分协议允许免费修改及非商用使用；部分协议要求修改后的衍生作品必须开源共享；还有部分协议明确禁止商用。在使用开源模型开展创新前，必须仔细阅读协议条款，避免违规商用或未经授权的二次分发。

《人工智能应用伦理安全指引 1.0》对 AI 开发者和服务提供者提出了明确的合规要求。开发者应将安全可控、公平公正、隐私保护等伦理要求嵌入系统设计；对具有高度自主决策能力的 AI 应用应评估失控风险并设置干预机制；建立事故信息回溯体系，确保责任可追溯。服务提供者应在高风险场景中保障 AI 仅承担辅助角色、不作为直接决策依据；持续监控 AI 服务运行状态；设置事故应急处置预案。使用者则应以审慎态度对待 AI 生成内容，避免过度依赖；妥善保护个人敏感信息；履行 AI 生成合成内容的标识义务。



1.4 本章实训：AI 模型能力边界测试与分析

实训目标

(1)实操体验主流通用大模型，测试其知识储备、逻辑推理、内容创作三大基础能力。

(2)识别大模型幻觉、知识滞后等典型缺陷，总结能力边界。

(3)简单体验 RAG 知识库工具，对比“纯大模型回答”与“RAG 增强回答”的差异。

(4)结合本章伦理、合规知识，分析测试过程中存在的潜在风险。

实训环境

(1)在线通用大模型(文心一言、通义千问、DeepSeek 等，选用 2 ~ 3 款即可)；

(2)轻量化免费 RAG 知识库工具(在线网页版，无须本地部署)；

(3)测试素材：过时知识问题、专业冷门知识、开放性创作题目、模糊逻辑问题、行业内部文档片段。

实训步骤

1. 基础能力测试(大模型)

分四类问题依次向大模型提问，记录每款模型的回答结果：一是基础常识与时事问题(测试知识储备与知识滞后)；二是数学逻辑、复杂推理问题(测试逻辑能力)；三是文案、短文创作问题(测试创作能力)；四是冷门专业知识、虚构问题(重点测试模型幻觉)。

2. 能力边界记录与分类

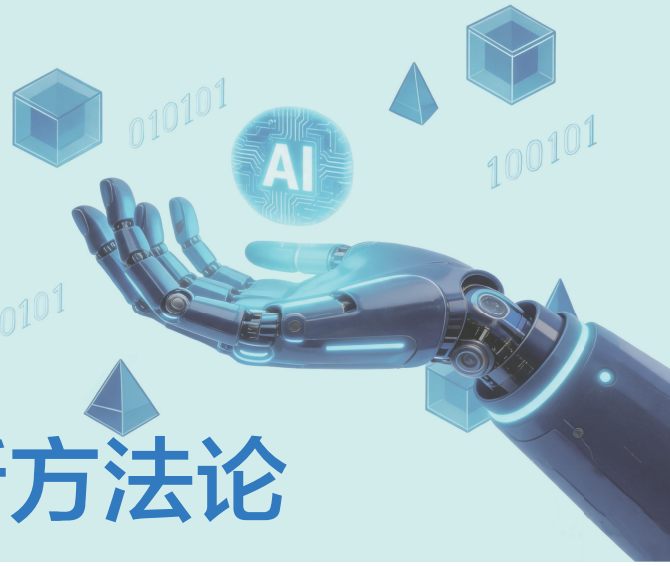
根据回答结果，将模型表现分为“回答准确”“回答有瑕疵”“明显编造(幻觉)”“无法回答”四类，梳理总结大模型擅长哪些任务、在哪些场景下容易出错、知识滞后体现在哪些方面。

3. RAG 对比测试

将一份冷门专业文档、过时资料上传至轻量化 RAG 知识库，分别向纯大模型和接入该知识库的 RAG 系统提问文档内的专业问题，对比两者回答的准确性，直观理解 RAG 弥补大模型短板的原理。

4. 伦理与合规风险排查

复盘整个测试过程，思考若将本次测试的 AI 能力应用到商用场景，可能存在哪些隐私泄露、版权侵权、内容违规风险，并列出对应的规避方案。



第2章

AI驱动的现代创新方法论

上一章厘清了人工智能的发展脉络、技术架构与伦理边界，对“AI 能做什么”作了系统性梳理。然而，掌握技术工具本身并不等同于具备创新能力。若空有技术视野却缺乏方法论的引导，再强大的 AI 也只能是被动响应的“问答工具”，而无法成为主动赋能的“创新引擎”。

当下生成式 AI、多模态、智能体等技术的普及，正在从方法、思维、实施流程三个维度重塑创新逻辑。



2.1 方法革新：传统创新模式的 AI 赋能改造

人类创新活动历史悠久，各行业均沉淀了一批经实践检验的经典创新方法。这些方法是创新工作的基本工具，即便在 AI 时代依然具备核心价值。但受限于人类自身的精力、知识储备、信息处理能力，传统创新方法在执行环节往往存在明显短板。

2.1.1 经典创新方法核心逻辑

创新方法种类繁多，既有偏向个人创意激发的理论，也有面向工业、产品、商业场景的系统性工具。

下面主要介绍沃拉斯创意四阶段模型、TRIZ 发明问题解决理论、联想类比法三类具有代表性的经典方法。

1. 沃拉斯创意四阶段模型

沃拉斯创意四阶段模型是心理学视角下的创意产生理论，也是个人开展创意构思最基础的框架，几乎适用于文案创作、产品设计、服务优化等所有轻量级创新场景。该模型将完整的创意生成过程划分为四个循序渐进的阶段，每个阶段有明确的任务与思维状态，如图 2-1 所示。

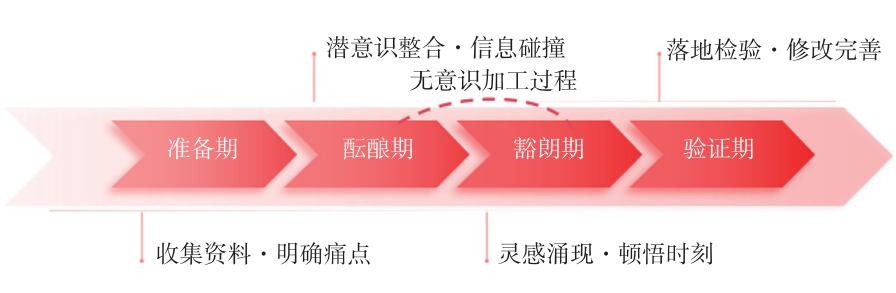


图 2-1　沃拉斯创意四阶段模型

(1)准备期(资料积累阶段)

准备期是创新的起点，核心工作是围绕目标问题收集资料、梳理现有信息、明确核心痛点，完成知识与素材的积累。

(2)酝酿期(潜意识加工阶段)

酝酿期是指暂时搁置对问题的直接求解，不再进行刻意思考，让无意识认知过程在后台对已有信息进行整合与重组。这一阶段看似“不思考”，实则大脑仍在隐性加工，为灵感的涌现积累条件。

(3)豁朗期(顿悟/灵感迸发阶段)

豁朗期是指经过长时间酝酿，全新的创意、解决方案会突然浮现。

(4)验证期(落地检验阶段)

验证期对灵感进行落地检验、修改完善，判断创意是否具备实用性，若存在缺陷则回溯前面的环节重新优化。

这套模型虽然完整解释了创意从无到有的过程，但在仅依靠人工落地沃拉斯模型时，信息收集、素材整合环节的效率会明显偏低。

拓展阅读

沃拉斯创意四阶段模型心理学实验与理论溯源

该模型由英国社会心理学家格雷厄姆·沃拉斯(Graham Wallas)于 1926 年在著作《思维的艺术》(*The Art of Thought*)中正式提出，是全球首个系统性梳理创造性思维全过程的心理学框架。沃拉斯并未依靠严格的实验室控制实验，而是通过对两类权威人物的创造自述进行整合归纳来构建理论：一是物理学家亥姆霍兹对科学灵感产生过程的长期记录，二是数学家庞加莱关于数学顿悟的经典随笔，二者均描述了“长期钻研→暂时搁置→灵感突发”的创作规律。

亥姆霍兹最早提出创造三阶段：持续钻研、放松搁置、灵感涌现；沃拉斯在此基础上新增验证期，形成完整四阶段框架，从而区分了“无意识灵感”与“有意识落地检验”两大思维模式。沃拉斯核心创新点在于提出边缘意识(fringe consciousness)概念，解释酝酿期大脑无意识信息重组的底层心理机制，为后世潜意识创意研究奠定基础。

2. TRIZ(发明问题解决理论)

如果说沃拉斯模型偏向个人灵感激发，那么 TRIZ 发明问题解决理论则是面向技术创新、工程创新的系统性工具，由研究者通过分析数百万份专利总结而来，核心目标是打破“盲目试错”的传统研发模式，按照固定规律解决技术矛盾。

TRIZ 的核心逻辑很简单，绝大多数技术难题、创新瓶颈，本质都是系统内部的矛盾冲突，比如产品想要体积变小，就可能导致性能下降；想要提升生产速度，就容易降低良品率。该方法会引导创新者主动识别矛盾类型，调用对应的创新原理、标准解法去破解问题。在工业设计、硬件研发、技术改造等场景中，TRIZ 能够大幅减少无效试错。

TRIZ 将创新中遇到的矛盾区分为两类(图 2-2)。

(1)技术矛盾

技术矛盾是指优化系统某一项参数，会造成另一项独立参数性能恶化，属于两个参数之间的冲突。例如，提升汽车行驶速度(优化参数)，会带来油耗上升(恶化参数)。针对技术矛盾，TRIZ 配套提供 39 个工程参数矛盾矩阵，可快速匹配适配的发明原理，是工程创新最常用工具。

(2)物理矛盾

物理矛盾是指针对同一个系统参数同时存在两种完全相反的设计要求。例如，手机屏幕，既需要大屏满足观看需求，又需要小尺寸方便随身携带。

针对不同类型的矛盾，TRIZ 提供了对应的 40 个发明原理和 76 个标准解作为求解工具。

但 TRIZ 使用门槛较高，需要使用者掌握大量专业知识与专利经验，人工梳理矛盾、检索解法的流程繁琐，学习和使用成本偏高。

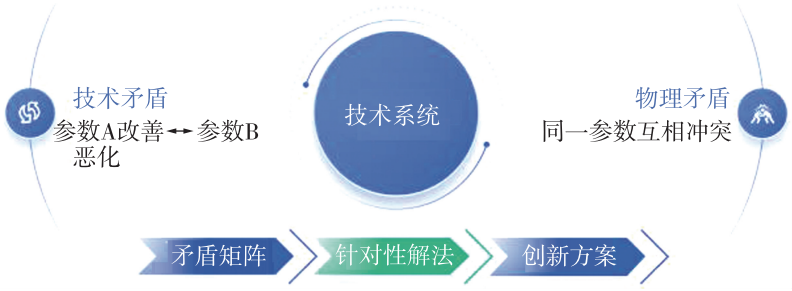


图 2-2　TRIZ 矛盾识别与求解逻辑

3. 联想类比法

联想类比法是门槛最低、应用最灵活的创新方法，广泛用于商业策划、文创设计、模式创新等领域。它的核心逻辑是举一反三和跨界借鉴。将 A 领域成熟的模式、结构、功能、创意，迁移到 B 领域中，通过对比和联想产生新想法，如图 2-3 所示。例如，把共享单车“分时租赁”的模式迁移到共享充电宝；将外卖平台的“实时配送追踪”模式迁移到校园快递代取服务中，就是典型的类比创新。这种方法上手简单，但很容易陷入思维局限，人类的认知边界决定了可联想的范围，多数人只能联想到熟悉的行业，很难实现大范围和跨学科的深度联想。

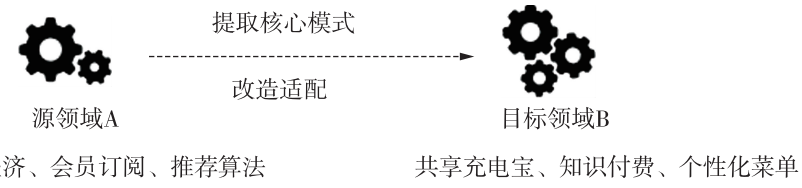


图 2-3　联想类比法跨界迁移示意

综合来看，三类经典创新方法各有侧重，覆盖了从个人创意到技术研发的全场景，但人工执行效率低、信息覆盖面窄、重复工作多、跨界联想受限，是大部分传统创新方法共同的痛点，这也为 AI 赋能留下了巨大空间。

2.1.2　AI 赋能传统创新的优化路径

AI 并非要取代这些传统方法，而是作为高效协作工具嵌入创新全流程，弥补人类在信息处理、海量检索、重复劳动和跨界联想等方面的不足，实现“人主导方向，AI 负责执行与拓展”的人机协同模式。

结合上述三类经典方法，下面将分场景讲解具体的优化路径。

1. 针对沃拉斯创意四阶段模型

AI 可以对四个环节逐一进行效率升级。

在准备期，传统模式需要人工查阅文献、整理资料、调研行业现状，耗时长久。借助大模型搭配 RAG 知识库，可快速检索、汇总行业公开资料、竞品信息与素材，大幅压缩资料收集周期。

在酝酿期，AI 可以基于现有素材进行多维度发散思考，批量生成不同方向的初步想法，弥补人类潜意识思考的局限性。

在豁朗期，AI 可以对涌现的灵感进行补充、延伸，拓展创意的细节。

在验证期，AI 能够模拟场景推演创意落地效果，快速排查明显的漏洞，减少线下验证的成本。

整体来看，AI 让创意产生的周期大幅缩短，从“慢节奏灵感等待”转变为“高效率创意迭代”。

2. 针对 TRIZ

AI 的赋能集中在矛盾识别、解法检索和方案模拟这三大环节。

TRIZ 的传统使用方法需要人工分析技术系统、梳理矛盾点、对照理论查找解法，对使用者的专业能力要求极高。依托行业大模型与工业知识库，输入产品参数和故障现象后，AI 可以辅助梳理系统参数冲突，匹配 TRIZ 创新原理与历史专利参考案例，搭配仿真工具，可以对提出的技术方案进行虚拟测试，不用制作实体样品就能预判可行性。这一优化让原本门槛较高的技术创新方法变得更容易被用户上手使用，也大幅降低了技术试错的成本。

3. 针对联想类比法

AI 重要价值之一就是打破人类的认知孤岛。

人类的知识范围有限，跨领域联想往往局限在相邻行业，而大模型学习了全学科、全行业的海量知识，能够基于指定场景，主动完成跨学科、跨产业的联想匹配。例如，若需为校园文创做

创新设计，人工只能联想到文具、饰品等周边产品，AI 则可以联动文旅、数字藏品、互动游戏、虚拟 IP 等数十个领域给出类比思路，极大拓宽创意边界。同时，AI 还能对每一个联想方向做出解读，说明借鉴点和改造方式，让类比创意不再停留在空想阶段。

传统创新方法与 AI 赋能后的变化见表 2-1。

表 2-1　传统创新方法与 AI 赋能后的变化

创新方法	传统模式核心痛点	AI 赋能优化环节	赋能后核心变化
沃拉斯创意四阶段模型	资料收集慢、灵感发散有限、验证周期长	全流程素材整理、发散构思、虚拟验证	创意周期缩短，想法维度更丰富
TRIZ	矛盾识别难、解法检索繁琐、专业门槛高	自动识别技术矛盾、匹配解法、方案仿真	降低使用门槛，减少技术试错次数
联想类比法	联想范围窄、跨界创意不足、落地解读弱	跨领域联想推荐、创意落地分析	拓宽创意边界，类比方案更具落地性

通过以上对比分析可以总结出 AI 赋能传统创新的四条通用路径，这也是贯穿所有创新场景的核心规律。

(1) 替代重复劳动，释放高阶能力

将资料整理、文案撰写、数据统计、格式排版等机械工作交给 AI，让人类聚焦创意、决策、价值判断。

(2) 依托海量知识，突破信息边界

依托 AI 的海量知识储备，突破个人视野限制，实现跨领域、跨行业的信息联动。

(3) 虚拟推演测试，降低试错成本

通过虚拟推演和模拟测试，用数字化方式替代线下实体试错，节约时间、物料与资金成本。

(4) 压缩迭代周期，加速优化频率

传统创新一轮迭代往往需要数周，AI 辅助下可以实现“想法—方案—修改”的快速循环，支持高频次迭代优化。

以上四条路径构成 AI 人机协同创新的内在机理，即 AI 负责低价值、高重复性的海量信息处理工作，人类专注价值判断、方向决策、创意审美等高阶创新工作，二者优势互补，弥补单一主体的能力短板。

2.2　思维升级：AI 时代创新核心思维逻辑

创新方法是“工具”，而创新思维是“内核”。当 AI 全面融入创新流程后，我们不能再沿用纯人工时代的固有思维。如果依然用传统思路使用 AI，只把 AI 当成简单的“问答工具”，无法发挥

其最大价值。想要做好 AI 创新，首先要完成思维层面的升级。

2.2.1 跨域迁移与需求洞察思维

跨域迁移思维与需求洞察，是一组相辅相成的核心思维方式。跨域迁移负责从广阔的知识疆域中寻找创新灵感与解决方案，需求洞察负责精准识别创新的目标与方向。二者结合，才能避免“创意新颖但脱离实际”的问题。

1. 跨域迁移思维

跨域迁移思维是在传统联想类比法基础上延伸而来，适配 AI 能力的高阶思维。它的核心定义是打破学科、行业、场景、技术之间的壁垒，将 A 领域成熟的技术、模式、流程、创意，经过改造后迁移到 B 领域，结合 B 领域的特点形成全新创新成果。在没有 AI 的时代，跨域迁移高度依赖个人阅历，只有深耕多个行业的从业者才能做到。而 AI 的出现，让普通学习者也能轻松实现大范围跨域。

跨域迁移分为技术跨域、模式跨域和创意跨域这三种常见类型。技术跨域聚焦底层算法和硬件设备；模式跨域聚焦商业运营和服务流程；创意跨域聚焦视觉、文化和内容设计。

(1) 技术跨域

技术跨域是把某一领域的成熟技术迁移到新的场景中。例如，工业领域的机器视觉检测技术，原本用于工厂质检，经过轻量化改造后，迁移到校园门禁考勤、实验室设备状态识别等场景，这就是典型的技术跨域。

(2) 模式跨域

模式跨域是将商业运营和服务模式跨界复用。例如，线上直播带货的模式，从电商领域迁移到文旅行业，打造“直播逛景区、直播卖农特产”的新形态。又如，会员订阅制从视频平台迁移到知识付费领域，形成了知识专栏订阅模式。

(3) 创意跨域

创意跨域多见于文创、设计领域。例如，将非物质文化遗产的传统纹样、工艺创意，迁移到数字绘画、虚拟形象设计中，让古老文化在现代数字载体中获得新生。

用好跨域迁移思维，不能简单照搬，必须遵循“借鉴—改造—适配”三步流程，而 AI 可以全程提供辅助，如图 2-4 所示。

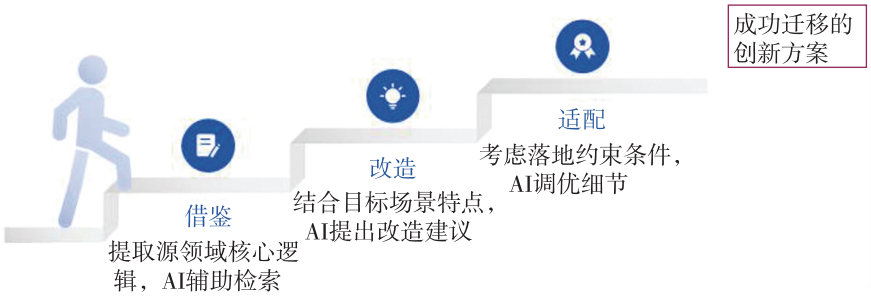


图 2-4 跨域迁移三步流程

①借鉴：借助 AI 检索目标领域或其他领域的成熟案例，提取核心逻辑与成功要素。

②改造：结合目标场景的特点，让 AI 分析原方案的优缺点，提出针对性的改造建议。

③适配：针对落地环境的约束条件，如用户习惯、技术基础、成本预算，做细节调整，以确保迁移后的方案贴合实际。

例如，一个学生团队想要为学校食堂设计一款 AI 点餐助手。如果仅从餐饮行业内部寻找灵感，思路可能局限于“扫码点餐”“排队叫号”等已有模式。但如果运用跨域迁移思维，他们可以借鉴“音乐 App 的个性化推荐算法”，根据用户的点餐历史、口味偏好、营养需求，为每位学生生成个性化的每日菜单推荐。这就是将推荐系统技术从娱乐领域迁移到餐饮领域的典型案例。

2. 需求挖掘思维

需求洞察的本质是穿透表面现象，找到用户真实、潜在、尚未被满足的诉求。多数创新初学者易将用户表面诉求等同于真实需求，最终造成创新项目偏离核心目标。例如，用户说“我想要一款更漂亮的 App 界面”，表面需求是美化用户界面，但深层需求可能是“系统操作逻辑不清晰，我希望使用起来更简便”。如果只针对表面需求去优化界面美观度，而忽略了底层操作逻辑的问题，创新就很难真正解决用户的痛点。

结合 AI 的分析能力，可以将需求洞察分为显性需求、隐性需求和前瞻性需求三个层次，如图 2-5 所示。创新需求挖掘遵循“先显性、再隐性、最后预判前瞻性需求”的递进顺序；显性需求是基础，隐性需求是创新核心突破口，前瞻性需求是长期差异化竞争力来源。

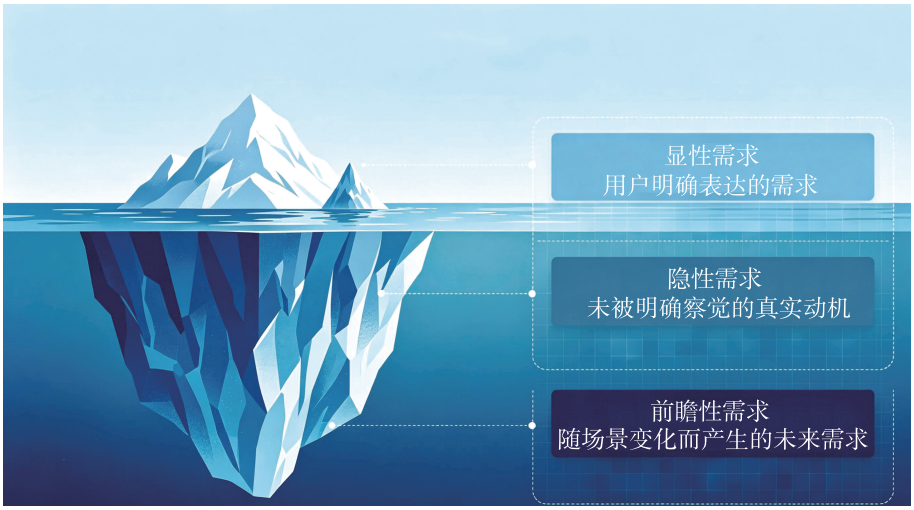


图 2-5　需求洞察三层模型

(1) 显性需求

显性需求即用户明确提出的诉求。AI 可以通过整理用户留言、问卷调查、评论区反馈等文本数据，快速汇总出一份显性需求清单。例如，学生在课程反馈中写道“希望作业提交后能立刻知道哪里错了”，这就是一个明确的显性需求。

(2) 隐性需求

隐性需求是用户自身尚未察觉或无法清晰表达的需求。AI 可以基于用户的行为数据和使用场景进行分析，挖掘用户潜在的痛点。例如，某款学习 App 的用户并没有直接抱怨“记笔记太麻烦”，但 AI 通过分析用户的操作日志发现，超过 60% 的用户在浏览课程视频时会频繁暂停并切换到备忘录，这一行为暗示用户存在“希望记笔记更方便”的隐性需求，进而催生出“AI 自动生成课程笔记摘要”的创意功能。

(3) 前瞻性需求

结合行业发展趋势和技术迭代方向，预判用户未来可能产生的需求，这是高阶创新的核心来源。AI 可以整合行业趋势报告、技术动态、政策导向等信息，辅助我们做出预判。例如，随着教育部逐步推进“人工智能通识课程”纳入高校必修课，可以预判，未来两年内高校教师将产生“快速生成 AI 课程教案”“AI 教学案例库建设”等前瞻性需求。

在 AI 创新实践中，标准的组合逻辑是先用需求洞察思维锁定场景痛点与创新目标，再用跨域迁移思维从更广泛领域寻找解决方案。双向结合，让创意既有落地价值，又有新颖性。

案例分析：假设一个学生团队想要为校园流浪猫救助站设计一个创新方案。

①运用需求洞察思维。他们发现救助站的核心痛点是志愿者人手不足，无法 24 小时监控流浪猫的健康状况；领养信息传播范围有限，很多健康流浪猫长期无人领养。

②运用跨域迁移思维。他们将“智能家居摄像头+异常行为告警”的技术模式迁移到救助站场景，部署 AI 摄像头，自动识别流浪猫进食异常、活动减少等健康预警信号；同时借鉴“社交电商”的传播模式，为每只流浪猫生成专属的“领养故事页”，通过朋友圈裂变传播扩大领养信息的触达范围。

这个案例展示了两种思维的有机融合，洞察找到了“真问题”，迁移找到了“好答案”。

2.2.2 快速试错与迭代创新思维

传统创新模式中，试错的成本很高，无论是制作实体产品还是搭建线下服务体系，或是开发一套系统，每一次错误都意味着时间、资金、人力的损耗，因此大家普遍倾向于“一次性做到完美”，前期反复论证，迟迟不敢落地实施，创新节奏十分缓慢。而 AI 凭借低成本模拟、快速生成和即时修改的能力，彻底改变了试错模式，也形成了快速试错与迭代创新这一核心思维，目前广泛应用于互联网产品、AI 应用、短视频、数字内容等创新领域。

1. 快速试错

首先需要澄清一个常见的认知误区，很多人会将“快速试错”等同于“盲目试错”，认为就是随意尝试和不计后果，这是典型的理解偏差。AI 时代的快速试错，是指在明确核心目标的前提下，利用数字化工具以极低的成本，在可控范围内验证方案的可行性，主动发现问题和暴露缺陷，它是有规划、有目标的探索，与毫无方向的盲目尝试有着本质区别。

AI 为快速试错提供了两大核心支撑。

(1) 近乎零实体物料成本

无论是产品设计方案、对话机器人原型、短视频脚本，还是小型应用程序的初版，都可以在数字空间内完成构建和测试，不需要制作实体物料。一个想法从产生到验证，可能只需要几十分钟的 AI 对话和几次点击，而不需要耗费太多实物资源。

(2) 并行试错能力

人类在同一时间只能测试一套方案，需要修改和测试完成后，才能开始下一套。而 AI 可以同时生成多套不同方向的方案，同步开展测试，大幅提升试错效率。例如，为一个新产品设计宣传标语，人工可能需要一天想三条，再花一天测试效果，而 AI 可以在几分钟内生成几十条不同风格的标语，并辅助分析每条标语的适用场景和目标受众。

2. 迭代创新

快速试错本身不是目的，它的价值在于为后续的迭代优化提供方向和依据。迭代创新，就是“测试问题→分析反馈→优化修改→再次测试”的循环往复过程，如图 2-6 所示。

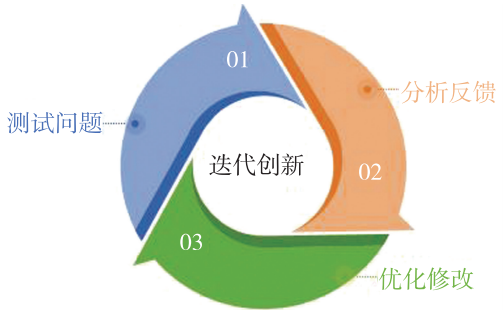


图 2-6　“迭代创新”循环流程

传统创新的迭代周期很长，一次迭代可能需要数月时间，包括收集反馈、开会讨论、重新设计、重新生产和再次测试。而 AI 驱动下的迭代可以做到“小步快跑、高频更新”，每一轮迭代只针对上一轮发现的局部问题进行优化，不推翻整体框架，逐步完善产品或方案。这种思维不再追求“一步到位”，而是坦然接受“初期版本不完美”，优先推出最小可行版本(Minimum Viable Product，MVP)，再根据真实反馈持续优化。

迭代创新的核心原则可以概括为以下三点。

①短周期：每一轮迭代的时间尽可能压缩，理想状态下以天为单位，而非以周或月为单位。

②小幅度：每次只改动一个或少数几个变量，便于精准定位问题原因。

③强反馈：每一轮迭代后必须有明确的数据或用户反馈作为下一轮改进的依据，避免闭门造车。

3. 落地流程

下面以开发一款面向大学生的“AI 学习助手”为例，完整演示“快速试错+迭代创新”的落地流程。

第一步：搭建最小原型

利用大模型的 API 接口，快速制作出最基础的功能，包括知识点问答和概念解释。不追求界面美观，不追求功能全面，甚至可以不开发独立的 App，而是先在微信小程序或网页对话框中实现核心对话能力，整体原型搭建可在 1 周内完成。

第二步：小范围试错

邀请 10 ~ 15 名同学作为种子用户，在真实的作业和复习场景中使用这款学习助手。同时，借

助 AI 自动分析对话日志，识别出以下问题。

- 某些专业术语的解释不够准确；
- 对于数学公式类的题目，模型输出格式混乱；
- 学生经常询问“有没有这道题的类似题练习”，但当前版本没有这个功能。

第三步：第一轮迭代

针对试错中发现的问题，逐一优化。

- 补充专业知识库，对易错术语添加权威来源的释义；
- 优化提示词，要求模型在输出数学公式时使用 LaTeX 格式；
- 新增“相似题目推荐”功能，利用 RAG 技术从题库中检索并推送相关练习题。

本轮迭代用时 3 天。

第四步：再次测试、再次迭代

将优化后的版本再次提供给同一批种子用户测试，收集新一轮反馈。这一次，用户可能会提出新的问题，如“希望能标记已掌握的题目”“希望能看到自己的学习进度曲线”。于是进入第二轮迭代，新增学习进度追踪功能。

如此循环往复，每一轮试错和迭代都可以在 3 ~ 5 天之内完成。1 个月后，这款学习助手已经从最初仅有问答功能的粗糙原型，逐步进化为具备知识点讲解、错题整理、练习推荐、学习进度可视化等功能的相对完善的工具。

总之，快速试错与迭代创新思维的核心精神可以概括为一句话：先做出来，再做对，再做好。



2.3 实施评估：AI 辅助创新需求与可行性分析

在掌握创新方法与创新思维，形成初步创意之后，不宜直接进入开发应用环节。创新是否具备应用条件、是否具有实际价值以及是否存在潜在风险，都需要经过系统评估。需求梳理是明确创新“做什么”，可行性分析是判断创新“能不能做、值不值得做”。AI 可以辅助我们完成研制工作的主要环节，使分析过程更全面、数据更客观，避免仅凭主观判断做出决策。

2.3.1 场景需求挖掘与痛点梳理

创意的根基是真实场景与真实痛点。脱离场景的创新只是空中楼阁，缺乏痛点支撑的方案难以被用户接受。场景需求挖掘与痛点梳理，就是将模糊的想法转化为清晰、可落地的需求清单，这是创新项目从构思走向实操的关键一步。

结合 AI 工具，可以把这个工作拆解为场景定位、用户分层、痛点拆解、需求转化四个连贯步骤，如图 2-7 所示。

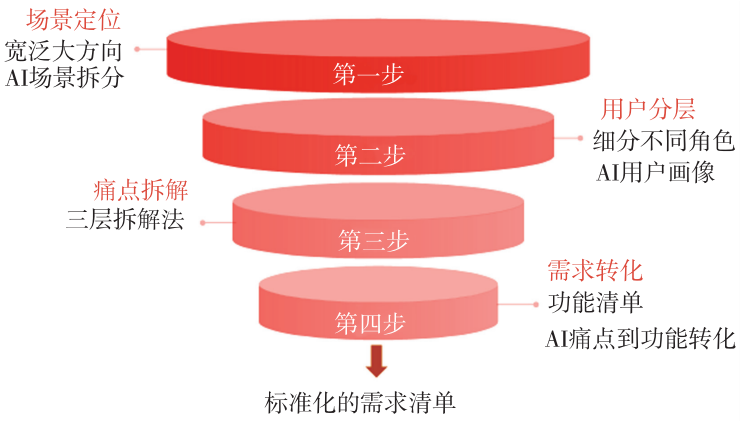


图 2-7　场景需求挖掘四步流程

1. 精准场景定位

所谓场景，就是创新成果所处的具体环境、使用时机与使用人群。例如，“AI 教育创新”是一个宽泛方向，而“中小学课后作业辅导场景”“高校实验室仪器教学场景”“成人职业技能培训场景”就是具体场景。场景越具体，创新的方向就越清晰。

传统模式下，需要人工调研行业资料划分场景，耗时且容易遗漏；借助 AI，可以输入大方向后，让 AI 按照使用人群、使用环境、使用时长、使用频率等维度，拆分出若干细分场景，并标注每个场景的特点与规模，帮助我们快速筛选出适合切入的目标场景。

例如：输入“校园食堂场景”，AI 可以拆分出以下细分场景。

- 早餐高峰时段的学生快速点餐场景；
- 午餐时段的菜品推荐与营养搭配场景；
- 晚餐后的剩菜统计与浪费分析场景；
- 教职工的小范围聚餐预订场景。

2. 用户分层

同一个场景中，不同身份、不同角色的用户，其诉求往往截然不同。以“校园 AI 图书馆助手”为例，用户至少可以分为三类：

- 借阅学生：关注如何快速找到书籍、办理借还手续、获取文献资料；
- 图书馆管理员：关注图书盘点效率、书架整理、违规行为管理；
- 图书馆运维人员：关注设备故障报修、系统维护、数据统计。

如果不进行用户分层，而是试图用一个方案满足所有人的需求，结果往往是哪个群体都没服务好。AI 可以基于场景描述，自动划分用户群体，并总结每类用户的核心身份、日常行为和主要接触点，为后续痛点分析奠定基础。

3. 核心痛点拆解

痛点拆解是整个环节的核心。痛点是指用户在场景中遇到的麻烦、低效问题或未被满足的诉求。建议采用“三层拆解法”，逐层深入，如图 2-8 所示。

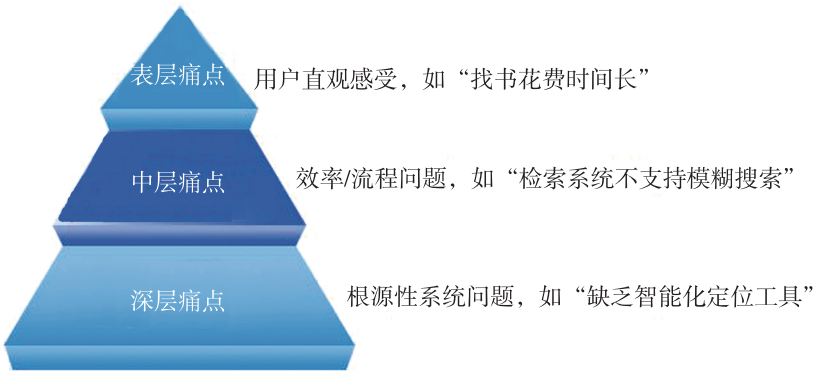


图 2-8　痛点三层拆解结构

(1) 表层痛点

用户直观感受到的行为麻烦。例如，“在图书馆找一本书需要花费很长时间”。

(2) 中层痛点

表层痛点背后的效率或流程问题。例如，“图书馆的图书分类检索系统老旧，检索效率低，且不支持模糊搜索”。

(3) 深层痛点

造成中层痛点的根源性问题。例如，“图书馆缺乏智能化的图书定位与导航工具，完全依赖人工排架和读者自行寻找”。

用户可以将场景描述和用户信息输入 AI，让 AI 结合行业通用问题，批量梳理出三级痛点，并按照“影响范围”和“严重程度”两个维度进行排序，优先解决高优先级痛点。

4. 需求汇总与转化

梳理完痛点之后，需要将“用户痛点”转化为“产品功能需求”。简单来说，就是针对每一个痛点，明确对应出解决方案(表 2-2)。

表 2-2　需求汇总与转化

痛点层级	痛点描述	对应的功能需求
表层	找书花费时间长	搭建 AI 智能导览功能，支持自然语言查书并生成室内导航路线
中层	检索系统老旧，不支持模糊搜索	接入 RAG 知识库，实现语义级别的图书检索
深层	缺乏智能化定位工具	部署蓝牙信标或 UWB 室内定位系统，结合小程序实现图书实时定位

最终整理成标准化的需求清单，明确每一项需求的功能描述、目标用户、使用方式和优先级。AI 可以辅助完成从痛点到需求的转化工作，同时规范清单格式，方便后续开发使用。

整套流程层层递进，从宏观场景聚焦到具体功能。配合 AI 工具，可以大幅提升梳理效率，同时保证需求的全面性和系统性。

2.3.2　项目多维可行性研判方法

完成场景需求与用户痛点的系统性梳理后，仅靠创意与需求不足以支撑商业落地。无论是什

么项目，在正式启动之前都必须开展完整、客观的多维可行性研判。研判的核心目的，是提前识别项目全生命周期中的技术、成本、市场与合规风险，避免投入大量时间和精力后才发现项目无法推进。

四维研判存在固定先后优先级，第一步研判技术落地可行性；第二步核查数据、内容、知识产权合规风险；第三步测算全周期投入成本；最后评估业务应用价值，若前置维度存在致命缺陷，可直接终止项目，无须开展后续分析。可行性研判可以分为技术可行性、投入与运营成本可行性、应用与业务可行性、合规与风险可行性四个核心维度。这四个维度相互制约、缺一不可，任何一项存在重大缺陷，项目都需要调整方案甚至暂缓推进。

1. 技术可行性研判



技术可行性研判

技术可行性研判是所有 AI 项目立项的前置门槛。企业判断技术可行性，核心不只是“能不能做出功能”，更是评估现有技术体系、团队能力、配套基础设施能否长期稳定支撑业务运行。

技术可行性研判通常分为以下三个层次：

(1)技术需求拆解。

梳理项目的全部功能需求，拆解对应的底层 AI 技术模块，区分两类技术。

①通用成熟技术：如通用对话大模型 API 调用、轻量化 RAG 知识库搭建、基础图像识别等。这些技术已大规模商用，落地难度低，供应链成熟。

②前沿待突破技术：如高精度工业小样本模型训练、实时多智能体协同系统、特定领域的垂直大模型微调等。这些技术存在算力需求高、算法适配难、长期稳定性不确定等因素。

(2)现有条件评估。

评估项目团队已有的基础条件，包括可使用的算力资源(如 GPU 服务器、云平台代金券)、团队成员的技术储备(是否有人熟悉大模型 API 调用、前后端开发)、可采购的第三方服务和开源框架(如智谱 AI API、LangChain 框架、Ollama 本地部署等)。

(3)潜在风险预判。

预判项目长期运行可能遇到的技术隐患，例如，高峰期并发请求的处理能力、模型更新迭代的难度、跨系统数据打通的复杂度、开源框架的社区活跃度和维护持续性等。

在实际项目中，用户可以借助 AI 辅助完成技术研判工作。输入完整的业务需求和功能清单，AI 可以自动拆解所需技术栈，匹配市面上成熟的商用产品和开源工具，标注各类技术的落地难度和行业参考案例，同时梳理出潜在的技术瓶颈和替代方案。这可以大幅缩短技术调研周期，形成客观的技术评估说明。

2. 投入与全周期成本可行性研判

对于有落地意向的项目，成本测算是必不可少的环节。成本测算不能只计算一次性开发费用，必须覆盖项目从启动到持续运维的全生命周期开销，以此判断项目的投入产出比是否合理。

完整的成本构成分为以下四大板块。

(1)一次性前期投入

包括模型 API 授权费用、软件采购、定制开发、硬件算力租赁、数据采集与加工、人员外包等启动阶段的支出。

(2)周期性运营成本

包括云服务器计费、API 调用年费、月度数据更新费用、内容审核人力成本等常态化开销。

(3)迭代优化成本

AI 模型需要持续微调、知识库需要定期更新、功能需要不断升级，对应每年固定的优化预算。

(4)隐性风险成本

如数据泄露后的应急处理和赔付、版权纠纷的赔偿、系统故障导致的停工损失等应急预留成本。

完成成本统计后，还需要对比项目预期的收益或价值。对于参赛或课程项目，收益可能不是金钱，而是用户量、使用频次、社会影响力等指标。学生可以借助 AI 快速搭建标准化的成本测算模板，根据项目规模和预期的使用人数，自动测算不同方案下的全生命周期总成本，对比轻量化方案与完整方案的差异，输出清晰的成本对比分析。

3. 应用与业务可行性研判

该维度核心判断的是：该 AI 创新方案能否创造真实的应用价值，在同类方案中是否具备独特的竞争力。这一研判分为两种情况：

(1)面向外部用户的应用场景(如创业项目、公益项目)

- 目标用户群体的规模和增长趋势如何？
- 市场上是否存在类似的已有产品或方案？它们的功能、体验和定价是怎样的？
- 本方案的差异化优势和核心竞争力是什么？
- 目标用户是否愿意使用(或付费使用)？使用意愿有多强烈？

(2)面向组织内部的应用场景(如学校的课程设计、校园服务优化)

- 方案能否提升现有流程的效率？能否降低人力消耗？
- 方案能否减少差错率？能否支撑更大规模的服务？
- 方案能否与现有的管理制度、工作流程和组织架构兼容？

AI 可以帮助使用者完成行业信息的整合与分析，包括批量抓取行业报告、搜集同类产品的公开资料、归纳当前市场未被满足的需求缺口，同时模拟不同运营模式下的应用效果，客观论证项目的业务价值。

4. 合规与长期风险可行性研判

这是 AI 项目区别于传统信息化项目的核心评审项，也是极易被使用者忽视。随着《生成式人工智能服务管理暂行办法》《中华人民共和国个人信息保护法》等法规落地实施，AI 项目在数据采集、模型训练、内容生成到对外服务的全流程都面临硬性监管要求。一旦出现合规漏洞，项目不

仅可能无法落地，还可能带来法律风险。

合规研判主要聚焦以下五大核心板块。

(1) 数据合规

核查项目采集、存储、调用的用户信息、行业文档是否获得合法授权，是否存在过度采集、私自爬取涉密版权数据行为，数据存储、脱敏、销毁流程是否符合监管标准。

(2) 生成内容合规

评估 AI 输出的文本、图像、音频等内容是否建立了审核机制。能否有效规避虚假信息、歧视性言论、不良导向内容的风险？针对中小学、未成年人校园场景，需额外落实青少年模式、内容分级过滤、隐私信息最小化采集、家长授权机制等专项安全措施。

(3) 生成内容标注合规

根据《生成式人工智能服务管理暂行办法》要求，对外公开提供 AI 生成文本、图像、音视频内容时，必须对 AI 生成内容进行显著标识，区分人工创作与机器生成内容，避免使用者混淆虚假信息与真实内容；面向竞赛、校园公开服务的 AI 项目需提前设计内容自动标注功能。

(4) 知识产权合规

梳理项目所使用的训练数据集、开源模型、第三方素材的版权协议。区分项目使用性质，例如，校园课程实训、公益校内服务属于非商用场景；创新创业参赛落地、对外付费服务属于商用场景，两类场景对开源模型、数据集、素材的版权授权要求完全不同，不可混用。

(5) 业务安全风险

对于涉及医疗建议、金融咨询、安全预警等高风险场景的项目，是否落实了“人在回路(Human-in-the-Loop，指 AI 仅做辅助分析，最终决策由人工完成)”机制？AI 是否被错误地设置为承担核心决策责任？是否建立了人工复核流程？

使用者可以借助 AI 完成合规风险自查：输入完整的项目流程、数据使用范围和内容生成场景，AI 可以逐条匹配现行法律法规的要求，列出全部潜在的合规风险点，并配套标准化的整改建议，形成一份完整的合规风险评估报告。

5. 综合研判结论

完成四大维度逐项分析后，需要整合全部研判结论形成综合评估意见。一般来说，综合结论分为以下三类。

- 项目可行：四个维度均无明显缺陷，可以按照现有方案启动项目。
- 局部待优化：某个维度存在一定风险，需要调整技术选型、控制成本规模或完善合规流程后再推进。
- 整体不可行：存在无法解决的技术瓶颈、过高的运营成本或重大的合规风险，建议放弃当前方案，更换创新方向。

使用者应将最终的研判结论整理成标准化的项目可行性评估报告，作为项目启动前的决策依据。

拓展阅读

国内大学生 AI 创新典型竞赛落地案例

以下案例均来自全国大学生顶级双创赛事，完整运用跨域迁移、需求分层、快速迭代、四维可行性研判全套创新方法。

案例 1：羽视智算——轻量化端侧视觉大模型

1. 痛点洞察(三层需求拆解)

表层：移动端、无人机、摄像头云端视觉大模型延迟高、流量消耗大；

中层：现有轻量化模型精度大幅衰减；

深层：缺少多模型协同蒸馏轻量化算法，端侧算力资源有限。

2. 跨域迁移创新思路

借鉴教师多模型协同教学的教育逻辑，迁移至模型蒸馏领域，设计多教师协同轻量化算法；将工业设备分层检测的迭代优化流程，迁移至模型多轮迭代训练。

3. AI 赋能落地与迭代流程

依托大模型 API 快速生成轻量化算法原型，小批量设备并行测试，每周一轮迭代；通过 AI 仿真模拟不同硬件算力下的运行效果，大幅减少实体设备测试成本。

4. 可行性研判亮点

技术：选用开源轻量化模型降低算力门槛。

合规：模型训练数据集全部采用开源标注数据，无隐私采集行为。

成本：依托高校免费 GPU 完成训练，轻量化部署无高额云端费用。

业务：适配智慧城市、消费电子双场景，商业化落地清晰。

案例 2：AI 智药引擎——抗癌药物分子智能筛选平台

1. 创新思维运用

跨域迁移：将 NLP 文本特征提取技术从大语言模型迁移至药物分子结构解析。

快速迭代：每周生成一批候选分子，AI 模拟分子结合效果，快速淘汰无效方案。

2. 落地研判核心优势

解决药企分子筛选周期长、成本高的行业痛点；区分校园实训非商用与企业合作商用两套版权方案，提前完成知识产权合规梳理；输出标准化软件著作权，已对接两家药企落地试用。

案例 3：万像翊新——8K+XR 历史影像 AI 修复重建

1. 需求分层拆解

显性需求：老旧历史胶片模糊、褪色、破损。

隐性需求：博物馆、档案馆缺少低成本数字化修复工具。

前瞻性需求：文旅数字活化、虚拟展陈的 AI 影像重建需求。

2. 经典创新方法结合

使用沃拉斯四阶段模型：收集海量老影像素材(准备期)→AI 批量生成多套修复算法思路

拓展阅读

（酝酿期）→融合多模态修复模型（豁朗期）→线下博物馆实景测试迭代（验证期）；同时借助 AI 完成 TRIZ 技术矛盾自动识别，平衡修复精度与渲染速度。

案例 4：心之声 AI 心电诊断

1. 核心创新逻辑

跨域迁移：图像分类大模型技术迁移至心电波形识别；分层需求：基层医院心电医师资源不足（显性）、偏远地区无专业诊断人员（隐性）、居家慢病远程监测（前瞻）。

2. 合规重点落地

医疗高风险场景严格落实“人在回路”，AI 仅输出辅助诊断建议，最终诊断由医师人工复核；患者心电数据本地脱敏存储，严格遵循《中华人民共和国个人信息保护法》，面向未成年人增设数据采集家长授权流程。



2.4 本章实训：场景化创新问题定义实践

实训目标

（1）熟练运用场景定位、用户分层、痛点拆解的方法，完成真实场景的需求梳理。

（2）掌握多维可行性研判流程，能够结合 AI 工具评估 AI 创新项目的落地性。

（3）践行跨域迁移、快速迭代等创新思维，形成完整、规范的创新问题定义方案。

（4）强化合规意识，在问题定义阶段提前排查 AI 项目的伦理、版权、数据风险。

实训环境与素材

（1）基础工具：主流通用大模型（文心一言、通义千问、DeepSeek、讯飞星火等），用于辅助调研、分析、梳理内容。

（2）参考素材：校园场景、小微企业办公、社区服务、线上学习四大通用参考场景（学生也可自主选择生活、学习、工作中的真实场景）。

实训步骤

1. 场景选定与定位

从给定参考场景或自主挖掘场景中，选定一个具体细分场景，明确场景环境、使用人群、日常运作模式。借助 AI 补充该场景的行业现状、普遍存在的问题，完成场景描述，要求内容具体、不宽泛。

2. 用户分层与痛点拆解

对场景内的用户进行分层，针对每一类用户，按照“表层痛点—中层痛点—深层痛点”的结构

拆解问题；结合 AI 拓展痛点维度，补充人工思考遗漏的问题，并按照优先级对痛点排序。

3. 痛点转需求，初步构思创新方案

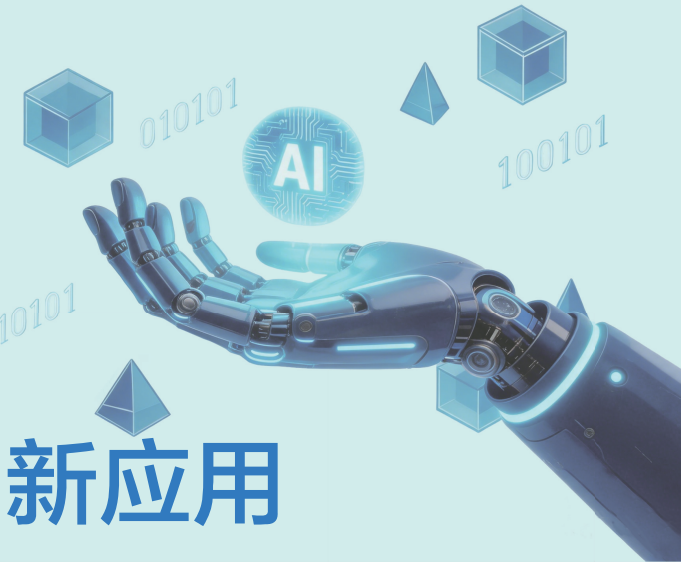
将高优先级痛点转化为可落地的功能需求，结合跨域迁移思维，借鉴其他领域的成熟方案，构思基于 AI 的创新解决思路，简要说明用到的 AI 技术(对话大模型、RAG、视觉技术等)。

4. 多维可行性研判

依次完成技术、成本、应用、合规四个维度的分析，借助 AI 排查风险、优化方案，最终给出综合研判结论。

5. 文档整合与复盘

将场景描述、用户分层、痛点清单、需求清单、创新思路、可行性研判内容整合为一份完整的场景化创新问题定义报告；小组内部复盘本次创新用到了哪些本章所学的创新思维与方法，方案还有哪些可优化空间。



第3章 通用对话大模型创新应用

对话是人类传递信息、完成协作最基础的交互方式，企业客服、办公助理、行业咨询、内容创作、方案撰写等绝大多数职场工作，都离不开问答与沟通场景。早年企业搭建的任务型对话系统存在明显短板，无法适配复杂、开放的业务需求；而通用大语言模型凭借强大的自然语言理解、逻辑推理、长文本生成能力，重构人机对话逻辑，成为各行各业数字化创新的基础工具。



3.1 问题导入：传统任务型对话系统的应用瓶颈

在人工智能发展历程中，让机器理解人类自然语言并给出有效应答，始终是自然语言处理的（NLP）核心研究目标。大模型普及前，对话系统已大规模应用于客服热线、智能音箱、银行语音导航等场景。其中，以完成特定业务为目标的系统，统称为任务型对话系统（Task－Oriented Dialogue System），其设计目标是在预先定义的业务范围内完成相应任务，例如话费查询、机票预订、网点地址咨询。

以银行客服电话为例，用户常常会遭遇如下体验：语音菜单层层嵌套，需要按好几个数字键才能找到人工服务；用户说出的需求稍显复杂，系统就反复回答“对不起，我没有听清，请重新输入”。上述现象集中反映了传统任务型对话系统的典型局限性。

从技术架构层面分析，传统任务型对话系统采用的是“规则主导、模板填充”的架构模式。开发者需要预先定义用户可能表达的各类意图，比如“查询余额”“转账汇款”“挂失卡片”，并为每种意图编写相应的应答模板和业务处理逻辑。这种方式的优点在于流程可控、稳定性高，但其缺陷同样突出，一旦用户的表达方式脱离了预设的规则范围，系统就变得无所适从。换言之，传统系统只能理解“有限的几种问法”，而人类的自然语言表达是千变万化的。

这种局限性在日常使用中造成了明显的摩擦感。用户需要学习“怎么跟机器说话”，而不是机

器去适应人的表达习惯。此外，传统系统缺乏上下文记忆能力，在一个多轮对话中，用户可能需要反复提供已经说过的信息，因为系统无法将前后对话联系起来。这也解释了为何大量用户宁愿排队等待人工服务，也不愿与机器客服交互。



3.2 探索分析：大模型场景价值与创新空间

大模型技术的出现，从根本上拓宽了对话系统的能力边界。以 GPT 系列、文心一言、通义千问、豆包、DeepSeek 等为代表的大语言模型，不再依赖预设的规则模板，而是通过对海量文本数据的学习，掌握了理解自然语言、生成流畅回复的通用能力。

通用对话大模型打破固定流程、标准化话术限制，具备开放式语义理解、多步骤逻辑推理、超长多轮对话、自主结构化内容生成能力，从而拓展出大量全新数字化创新场景。

3.2.1 传统任务型对话系统的技术架构缺陷

理解大模型所带来的变革，不能仅停留在用户体验层面的描述，而应深入系统内部，剖析传统任务型对话系统在技术架构层面的根本性局限。

传统任务型对话系统普遍采用模块化流水线架构，典型结构如图 3-1 所示。其中，从自然语言理解(NLU)到自然语言生成(NLG)的四个模块构成了语言理解与决策的核心链路，具体功能如下。



图 3-1 模块化流水线架构

①语音识别(ASR)。将用户的语音输入转换为文本，是整个对话链路的入口。例如，用户对着手机说“我想订一张明天去北京的机票”，ASR 模块将其转定为文字，供后续的 NLU 模块处理。

②自然语言理解(NLU)。识别用户意图并提取关键信息(Slot，即“槽位”)。例如用户说“我想订一张明天去北京的机票”，NLU 模块将其解析为意图=“订机票”，槽位={日期=“明天”，目的地=“北京”}。

③对话状态追踪(DST)。维护当前对话的进度，记录已收集了哪些槽位信息、还缺少哪些信息。

④对话策略学习(DPL)。根据当前状态决定下一步行动，比如“询问缺失的日期”或“调用订