ICS 35.080
CCS L 77

团　　体　　标　　准

T/CCUA 070—2026
T/BSCEA 007—2026

系统与软件工程 人工智能应用开发项目造价评估

System and software engineering — Artificial intelligence application development project cost evaluation

2026 - 10 - 20 发布　　2026 - 11 - 20 实施

中 国 计 算 机 用 户 协 会
北京软件造价评估技术创新联盟　　发布

目　　次

前言 II
1 范围 1
2 规范性引用文件 1
3 术语、定义和缩略语 1
3.1 术语和定义 1
3.2 缩略语 4
4 成本构成 4
4.1 总则 4
4.2 综合人力成本 5
4.3 直接非人力成本 7
5 评估过程 9
5.1 评估总体流程 9
5.2 确认项目范围 9
5.3 明确费用分类 10
5.4 选择评估方法 10
5.5 评估项目费用 10
6 综合人力成本评估 11
6.1 主体功能开发费 11
6.2 数据处理费 12
6.3 模型建设费 14
7 直接非人力成本评估 18
7.1 软硬件购置费 18
7.2 资源租赁费 19
7.3 人工智能环境集成费 21
7.4 数据资源与数据服务费 22
7.5 模型服务费 23
7.6 其他直接非人力成本 23
附录 A（资料性）典型任务包及对应岗位系数示例 25
附录 B（资料性）项目造价汇总表 27
附录 C（资料性）数据处理费评估参数及示例 28
附录 D（资料性）模型建设费评估参数及示例 31
附录 E（资料性）项目造价评估示例 36
参考文献 50

前　言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由中国计算机用户协会和北京软件造价评估技术创新联盟共同提出并归口。

本文件起草单位：中国计算机用户协会软件造价分会、北京软件造价评估技术创新联盟、北京中基数联科技有限公司、中国太平洋保险（集团）股份有限公司、中国华能集团有限公司工程技经中心、招商银行股份有限公司、中移九天人工智能科技（北京）有限公司、中国华电科工集团有限公司、中国移动通信集团浙江有限公司、中国科学院软件研究所、苏州市软件评测中心有限公司、北京赛迪时代信息产业股份有限公司、中国华电集团有限公司经济技术咨询分公司、江苏移动信息系统集成有限公司、中电信数智科技有限公司、国网经济技术研究院有限公司、北京维刚创新科学研究院、国电南京自动化股份有限公司、农银金融科技有限责任公司、北京华审金建国际工程项目管理有限公司、浙江省电子信息产品检验研究院、北京北咨信息工程咨询有限公司、北方实验室（沈阳）股份有限公司、吉林省电子信息产品检验研究院、江苏省软件产品检测中心、胥瓷信息科技（苏州）有限公司、江西省机电设备招标有限公司、中国地质调查局南京地质调查中心（华东地质科技创新中心）、中移物联网有限公司、中国移动（浙江）创新研究院有限公司、中国机械工业集团有限公司、中国雅江集团有限公司、华电电力科学研究院有限公司、江苏省战略与发展研究中心（江苏省信息中心）、中国进出口银行、工业和信息化部电子第五研究所、交通银行股份有限公司、山东省财政厅、中铁建数智科技（北京）有限公司、内蒙古电力（集团）有限责任公司数字研究分公司、丽水市数据局、山东省交通科学研究院、中建数字科技有限公司、山东省数智交通科技有限公司、华昆工程管理咨询有限公司、安徽安天利信工程管理股份有限公司、无锡市信息化协会、中安启成科技有限公司、宁波世明工程管理咨询有限公司、山东一直电子信息工程监理咨询有限公司、陕西卓信信息技术服务有限公司、建银工程咨询有限责任公司、中讯邮电咨询设计院有限公司、山东赛宝信息技术咨询有限公司、成都久信信息技术股份有限公司、湖南新星项目管理有限公司、京通建设管理有限公司、深圳市高亚弘信息咨询有限公司、北京扶风科技有限公司、河北千诚电子科技有限公司、云南金质信息技术服务有限公司。

本文件主要起草人：王海青、王青、代寒玲、余剑、李培圣、冯军红、戴锡平、孙喆、彭欣华、张春、付岩峰、陈旭攀、武海军、吴春雷、刘丹琦、徐少松、胡忠义、汪浩、周明明、张新业、程祎、胡宇、袁硕之、马邱哲、耿鑫州、刘珊、朱辰泽、王晓敏、刘娟、喻晓、任河、李琳、唐莹、邓阿北、薄慧、何丽丽、彭月、吴涵宇、马灿璇、冯诗正、白雅文、张晓东、邓子谦、林怡、宋宁宁、高岩、蒋小雨、田冬、王芳、石云昀、李晓明、隋舟、彭洁、施伟谊、邓海刚、毛世佳、游兴宇、缪梦蝶、王艳俊、许建超、赵冉、王磊磊、刘俊、王强、史强、董晓梁、由志远、赵奕霖、尹思达、王莉、邓姣凤、黄兆强、詹小容、赵欣、张世泽、张喃。

系统与软件工程 人工智能应用开发项目造价评估

1 范围

本文件确立了人工智能应用开发项目成本的构成，规定了人工智能应用开发项目的造价评估过程及方法。

本文件适用于各行业相关组织进行基于人工智能技术的应用开发项目的造价评估，包括人工智能算法建模和基于基础模型构建智能体的应用开发等，适用于相关单位开展基于人工智能技术的应用开发项目预算编制、招投标、结算及审计等活动时的造价评估。

本文件不适用于基础模型的预训练费、底层基础模型研发费、人工智能基础技术研究费、硬件开发费、人工智能应用的运行维护费的造价评估。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 36964-2018 软件工程 软件开发成本度量规范

T/BSCEA 002-2024 软件造价评估实施规程

T/BSCEA 005-2025、T/CCUA 060-2025 系统与软件工程 信息化项目 造价评估

3 术语、定义和缩略语

3.1 术语和定义

GB/T 41867-2022中界定的以及下列术语和定义适用于本文件。

3.1.1

人工智能应用开发 artificial intelligence application development

以人工智能技术和模型为基础，针对人类定义的给定目标，产生诸如内容、预测、推荐或决策等输出的一类应用软件开发活动。

3.1.2

智能体 artificial intelligence agent

感知并响应环境，主动采取行动以实现特定目标的实体。

[来源：ISO/IEC 22989-2022，3.1.1 有修改]

3.1.3

机器学习 machine learning

通过计算技术优化模型参数的过程，使模型的行为反映数据或经验。

[来源：GB/T 41867-2022，3.2.10]

3.1.4

深度学习 deep learning

通过训练具有许多隐藏层的神经网络来创建丰富层次表示的方法。

注：深度学习是机器学习的一个子集。

[来源：GB/T 41867-2022，3.2.27]

3.1.5

数据处理 data processing

对原始数据进行抽取、转换、加载的过程。

[来源：GB/T 37988-2019，3.13]

3.1.6

数据标注 data labeling

给数据样本指定目标变量和赋值的过程。

注：包括实体标注、意图分类、属性标引、关系标注等，为机器学习模型训练提供带有标签的训练样本。

[来源：GB/T 41867-2022，3.2.29]

3.1.7

数据预处理 data preprocessing

在数据输入机器学习模型之前，对原始数据进行清洗、格式化、归一化、编码等处理的过程，使其转换为算法能够理解和处理的格式。

3.1.8

预训练模型 pretrained model

预先在大型数据集上针对特定任务进行训练过的机器学习模型，可以针对不同但相关的任务重复使用或进行微调，为开发团队节省时间和计算资源。

3.1.9

模型优化 model optimization

提升模型执行速度、泛化能力，或改善利益相关方所关心的其他特性的方法。

注：如神经网络模型优化的方式包含剪枝、量化、调整参数、调整模型深度和宽度、增减特征或根据硬件平台具体

特性重新安排聚合算子等。

[来源：GB/T 41867-2022，3.2.19]

3.1.10

微调 fine-tuning

为提升人工智能模型的预测精确度，一种先以大型广泛领域数据集训练，再以小型专门领域数据集继续训练的附加训练技术。

注：常用于解决过拟合问题。

[来源：GB/T 41867-2022，3.2.31]

3.1.11

模型压缩 model compression

在尽量不损失模型精度的前提下，减小模型的体积、内存占用和推理延迟的技术。

注：包括剪枝、量化、知识蒸馏、低秩分解等方法，使模型能在资源受限的设备上部署。

3.1.12

模型评价 model evaluation

对人工智能或机器学习模型的性能进行系统性测量和分析的过程。

注：包括准确率、精确率、召回率、F1-score等指标衡量模型的准确性、效率、鲁棒性等质量特性。

3.1.13

模型部署 model deployment

将训练完成的机器学习模型整合到实际应用环境的过程。

注：包括模型打包、环境配置、服务发布、监控运维等环节。

3.1.14

智能体编排 agent orchestration

在统一系统中协调多个专用AI智能体，以高效实现共同目标的过程。

注：包括任务分配、工作流协调、数据共享和上下文管理，使多个智能体能够协作解决复杂问题。

3.1.15

检索增强生成 retrieval-augmented generation，RAG

一种通过从外部知识库中检索相关信息来增强和“锚定”大语言模型生成过程的技术，使模型能够基于检索到的上下文信息生成更准确、更新、可验证的答案。

3.1.16

提示词工程 prompt engineering

通过设计和优化输入给AI模型的提示词，引导模型产生准确、有用输出的一系列技术和方法。

注：包括指令设计、示例选择、输出格式规范等。

3.2 缩略语

下列缩略语适用于本文件。

API：应用程序接口（Application Programming Interface）
CV：计算机视觉（Computer Vision）
DaaS：数据即服务（Data as a Service）
FP：功能点（Function Point）
GPU：图形处理器（Graphics Processing Unit）
MaaS：模型即服务（Model as a Service）
NLP：自然语言处理（Natural Language Processing）
OCR：光学字符识别（Optical Character Recognition）
SR：语音识别（Speech Recognition）
TPU：张量处理器（Tensor Processing Unit）
WBS：工作分解结构（Work Breakdown Structure）

4 成本构成

4.1 总则

人工智能应用开发项目的开发成本是开发方为了满足特定业务需求而采用人工智能技术进行应用开发所付出的各种资源代价的总和，成本构成如图1所示，包括开发方的直接人力成本、直接非人力成本、间接人力成本、间接非人力成本，各类成本的定义按照GB/T 36964-2018的规定执行。图1虚框包含直接人力成本、间接人力成本和间接非人力成本三部分，其中间接人力成本、间接非人力成本通常分摊计入项目成本，与直接人力成本合在一起构成综合人力成本，按开发工作量和人力费率来计算，开发方的主要开发活动包括主体功能开发、数据处理和模型建设三部分，相应的综合人力成本包括主体功能开发费、数据处理费和模型建设费。直接非人力成本是开发方为此项目开发所付出的软硬件购置费、资源租赁费、人工智能环境集成费、数据资源与服务费、模型服务费和其他直接非人力成本。根据项目的具体情况，可能仅涉及其中的部分费用项，未发生的不计，将项目实际发生的各项费用计入相应的费用类别，避免重复列支。

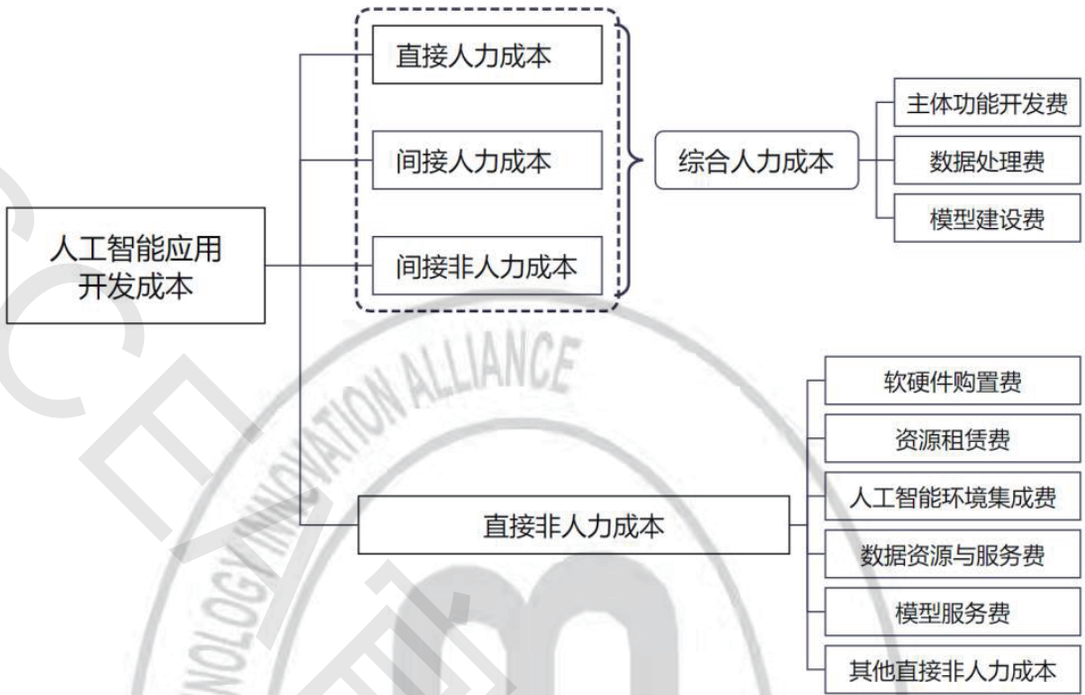


图 1　人工智能应用开发成本构成

上述成本包括人工智能应用开发项目整个开发过程所涉及的需求、设计、构建、测试、实施，以及相关的项目管理和支持等活动的成本。

注 1：本文件所述的人工智能应用开发项目造价是开发方付出的各类成本总和，仅指开发成本，按本文件 4.1 的范围说明，不包括硬件开发费，也不包括项目前期的规划、可研与方案设计、招标采购等费用，项目建设过程的监理费，人工智能应用开发完成后所需的安全、合规、伦理、测评、认证等第三方服务费用（以上费用归属于信息化项目建设其他费），此外也不包括运维成本。

注 2：直接非人力成本所涉及的各类费用仅限于人工智能应用开发期内发生的费用。应用交付后运行期内涉及的直接非人力成本，属于运维成本，不计入人工智能应用开发成本。

注 3：人工智能应用开发项目所交付的人工智能应用产品或服务应符合国家关于人工智能安全相关的政策法规和标准的规定，包括数据安全、内容安全、模型安全等，开发方在项目开发过程中为满足安全要求而产生的开发费用包含在开发成本中。

4.2　综合人力成本

4.2.1　主体功能开发费

主体功能开发费是为了满足用户人工智能业务需求，开展业务功能开发、人工智能平台（包括但不限于数据工具、模型工具、服务组件）功能开发和优化、人工智能应用集成等项目开发活动所产生的综合人力成本，包括开发方在项目开发过程中为实现这些软件功能和应用集成所付出的直接人力成本、间接人力成本和间接非人力成本。

人工智能应用开发项目的主体功能开发主要包括：

a）　面向用户的业务功能开发；

b) 对人工智能应用与数据采集系统（含采集终端）、业务系统、人工智能平台或第三方提供的 DaaS、MaaS 等系统进行应用集成开发；

c) 本项目需要开发或者交付的人工智能平台的功能（如数据采集、清洗、预处理、治理的功能组件、人工智能模型及参数、提示词、知识库的管理、模型训练、模型微调、模型评测、智能体编排、模型压缩、模型部署相关的功能组件）开发或优化。

根据项目的实际情况，主体功能开发可包含上述一项或多项内容，应根据项目实际发生情况和获取方式确定费用项和费用性质，正确地计取主体功能开发费，未发生的不计，不得重复计取。

4.2.2 数据处理费

4.2.2.1 数据处理费是指人工智能应用开发项目中，开发方为支撑人工智能模型建设而开展的数据工程活动所产生的费用中的综合人力成本，包括开发方为实施数据工程活动所付出的直接人力成本和间接成本。

4.2.2.2 数据处理费包含以下数据工程活动的综合人力成本：

a) 数据准备、整理、采集、清洗、转换、标注、预处理、数据质量控制、数据安全、存储与管理等数据工程活动；

b) 人工智能模型建设、模型优化过程中对数据进一步处理涉及的数据工程活动。

4.2.2.3 应根据项目实际数据处理活动发生情况和获取方式确定费用项和费用性质，正确地计取数据处理费，未发生的不计，不得重复计取。数据处理费计取说明如下：

a) 根据项目的具体情况，数据处理可只包含 4.2.2.2 数据工程活动中的部分活动，未发生的不计；

b) 由项目开发方实施数据工程活动而产生的综合人力成本，应计入数据处理费；

c) 如果部分数据工程活动不由项目开发方实施，而是通过购买数据或数据服务的方式获取，则相应的费用不得计入数据处理费，应计入直接非人力成本中的数据资源与服务费，不得重复计取；

d) 如果部分数据工程活动由项目开发方通过调用现有的人工智能平台中的数据处理工具实现，则应考虑由此带来的开发工作量及数据处理费的下降；

e) 如果项目开发方为完成本项目的数据处理而开发了专用的数据处理等工具，则相关工具作为交付给用户的功能（作为人工智能应用的部分功能，或者作为单独的工具），或者作为软件开发过程中的一次性转换功能，由此产生的费用应计入主体功能开发费，不计入数据处理费；

f) 评估数据处理费时，各类数据工程活动的人力费率，宜根据组织的历史项目数据、人力资源市场行情、行业基准数据以及各地人力资源和社会保障部门发布的相关岗位人力薪酬统计数据，结合组织情况综合确定。

4.2.3 模型建设费

4.2.3.1 模型建设费是指人工智能应用开发项目中开发方采用各种人工智能技术进行人工智能模型建设（包括新建和优化）活动所产生费用中的综合人力成本，包括开发方实施人工智能工程活动所付出的直接人力成本和间接成本。

4.2.3.2 人工智能模型建设工程活动包含但不限于以下活动：

a) 采用传统的机器学习和/或深度学习算法等技术构建人工智能模型的工程活动，包括模型训练、评测和模型优化、模型部署、模型安全保障等活动；

b) 基于就绪可用的基础模型（包括商用或开源的、通用或领域专用的、预训练的大模型或小模型等），采用各种人工智能技术如模型微调、提示词工程、知识库检索增强等，使用特定领域或应用场景的专业或私有数据集对基础模型进行进一步训练优化，以实现满足用户需求的具备特

定能力的定制模型的人工智能工程活动，包括模型搭建、模型训练、评测和模型优化、模型压缩与部署等。

4.2.3.3 应根据项目实际模型建设活动发生情况和获取方式确定费用项和费用性质，正确地计取模型建设费，未发生的不计，不得重复计取。模型建设费计取说明如下：

a) 根据项目的具体情况，模型建设可只包含 4.2.3.2 中的部分活动，未发生的不计；
b) 模型建设费不包含模型建设所需的数据工程活动的费用；
c) 由开发方实施人工智能工程活动而产生的综合人力成本，应计入模型建设费；
d) 如果部分人工智能工程活动不由项目开发方实施，而是通过外购第三方模型服务的方式获取，则相应的费用不计入模型建设费，应计入直接非人力成本中的模型服务费，不得重复计取；
e) 如果部分人工智能工程活动由开发方通过调用现有的人工智能平台中的工具和组件实现，则应考虑由此带来的工作量及模型建设费的下降；
f) 模型建设过程中如涉及人工智能平台的功能组件开发（如模型训练、模型评测、模型压缩与部署、模型监控、算子、算法管理、提示词管理、知识库管理、智能体管理等相关的功能组件或工具），以及人工智能应用与数据采集系统、业务系统、DaaS、MaaS 等系统进行应用集成开发，相应的开发费用应计入主体功能开发费，不计入模型建设费；
g) 模型建设费评估时，各类人工智能工程活动的人力费率，宜根据组织的历史项目数据、人力资源市场行情、行业基准数据及各地人力资源和社会保障部门发布的相关岗位人力薪酬统计数据，结合组织情况综合确定。

4.3 直接非人力成本

4.3.1 软硬件购置费

4.3.1.1 软硬件购置费是开发方为人工智能应用开发项目购置专用的软硬件的费用，主要包含下列费用：

a) 硬件设备购置费——购置人工智能应用开发所需的专用服务器、终端、网络设备、安全设备、各类外接设备等资源的费用；
b) 就绪可用软件购置费——购置人工智能应用开发所需的专用系统软件、支撑软件和应用软件等资源的费用，包括人工智能软件开发工具链的各类工具软件。

4.3.1.2 应根据项目开发所需软硬件的实际情况和获取方式确定费用项和费用性质，正确地计取软硬件购置费，未发生的不计，不得重复计取。软硬件购置费计取说明如下：

a) 本项目开发所需专用软硬件由开发方代业主方购置并用于项目开发，项目结束后所购置的软硬件作为资产移交给业主方，相应的购置费计入项目开发成本的直接非人力成本，全额纳入本项目预算中；项目决算核算开发成本时，应按财务规定将软硬件资产折旧分摊计入本项目的开发成本；
b) 如果所购置的软硬件并非本项目专用，则相应的购置费不计入本项目的直接非人力成本，本项目中开发方使用这些软硬件的相应费用包含在间接成本中，不再单独计算；
c) 如果开发方已有项目开发所需的软硬件，则本项目中开发方使用相应的软硬件的费用也不单独计算，包含在间接成本中；
d) 如果开发方通过资源租赁或购买服务方式获取项目所需的软硬件资源，则项目开发期内发生的相应费用应计入本项目直接非人力成本的资源租赁费；

e) 如果业主方提供项目开发所需的软硬件及基础环境，则本项目开发方付出的间接成本中需扣减此部分费用，可双方协商此部分费用的扣减方式，如按市场租赁价格扣除开发期内的相应费用，或者适当降低开发的人力费率等。

4.3.2 资源租赁费

4.3.2.1 资源租赁费指人工智能应用开发项目开发期内租赁第三方资源或购买服务所需的费用，主要包括但不限于下列费用：

a) 机房（柜）租赁费用——租赁机房设施及配套设备所发生的相关费用（含电费）；
b) 硬件租赁费用——租赁服务器（如人工智能应用所需的 GPU、TPU 算力服务器）、终端、网络设备、安全设备、各类外接设备等资源时发生的相关费用；
c) 软件租赁费——租赁系统软件、支撑软件（如人工智能应用软件开发工具链）和应用软件等资源时发生的相关费用；
d) 云服务租赁费——租赁云服务环境下的服务器、网络设备、安全设备、系统软件、支撑软件和应用软件等云资源的费用；
e) 通信链路租赁费——租赁互联网出口专线、数字专线、裸光纤等传输链路的费用。

4.3.2.2 应根据项目所需的各类资源实际情况和获取方式确定费用项和费用性质，正确地计取资源租赁费，未发生的不计，不得重复计取。资源租赁费计取说明如下：

a) 根据项目的实际情况，可只包含 4.3.2.1 中的部分费用项，未发生的不计；
b) 项目开发所需的软硬件等资源，可通过采购或租赁的方式获取，但不得重复计取；
c) 人工智能应用项目交付后，运行期所需的资源租赁费应计入运维成本，不包含在项目开发成本的资源租赁费中。

注：采取租赁方式或者购买服务等方式实施的项目，需充分考虑业务数据的安全性、合同期满后数据迁移成本等因素。

4.3.3 人工智能环境集成费

人工智能环境集成费是人工智能应用开发项目中为实现人工智能应用专用的设备互联互通、功能协同运作，提升整体管理效率，将项目中分散的人工智能专用硬件、软件进行整合，通过结构化设计和部署调试，形成统一协调、资源共享的人工智能应用开发与运行环境所需的费用。

应根据项目所需的人工智能应用开发与运行环境的实际情况及获取方式确定费用项和费用性质，正确地计取人工智能环境集成费，未发生的不计，不得重复计取。人工智能基础环境集成费计取说明如下：

a) 如专用的算力服务器、专用的网络设备、存储设备、资源管理与调度软件、专用开发工具等进行集成以构建人工智能应用开发及运行环境等，相应的费用应计入本项目直接非人力成本中的人工智能环境集成费；
b) 人工智能应用开发项目涉及所开发的应用软件与数据采集系统（含采集设备）、业务系统、人工智能平台或第三方 DaaS、MaaS 等系统进行应用集成开发所产生的费用，应计入项目的主体功能开发费，不计入人工智能环境集成费；
c) 如项目不涉及人工智能环境的集成，则不计取该项费用。

4.3.4 数据资源与服务费

数据资源与服务费是指人工智能应用开发期内通过购买或租赁等方式从第三方获取符合法规要求和业务需求的数据资源与数据服务所产生的费用，包括但不限于：数据购置费（数据购买或授权使用

费）、数据处理工具使用费、数据采集服务费、数据存储服务费、数据标注服务费、数据预处理服务费、数据分析服务费、数据治理服务费、数据安全服务费等。

应根据项目所需的数据资源与服务的实际情况和获取方式确定费用项和费用性质，正确地计取数据资源与服务费，未发生的不计，不得重复计取。

如果相关的数据服务是由项目开发方在开发期内自行完成相关的数据工程活动，则相关费用应计入项目综合人力成本中的数据处理费。

注：本条所述的数据资源与服务费与T/BSCEA 005-2025、T/CCUA 060-2025中所述的数据资源建设费所包含的内容有所不同，仅限于项目开发期内从第三方获取数据资源和DaaS的费用，但T/BSCEA 005-2025、T/CCUA 060-2025中数据资源建设相应费用项的评估方法同样适用于本条的外购数据资源与数据服务费。

4.3.5 模型服务费

模型服务费指人工智能应用开发期内通过购买授权、租赁服务或调用接口等方式获取第三方提供的MaaS所需的费用，主要包含下列费用：

a) 模型授权费——购买第三方模型服务所需的授权费；

b) 模型服务费——通过 API 调用方式使用第三方提供的模型训练、推理等模型服务所需的费用，一般按调用次数或词元使用量计量，通常采用按量计价、包月套餐或包月限量套餐+超出额度后按量计价等方式计算费用。

应根据项目所需的模型服务的实际发生情况和获取方式确定费用项和费用性质，正确地计取模型服务费，未发生的不计，不得重复计取。

4.3.6 其他直接非人力成本

其他直接非人力成本指人工智能应用开发项目开发期内所涉及的除4.3.1至4.3.5费用科目之外的直接非人力成本，如项目差旅费、技术协作费、项目开发场地租金等费用，应根据项目实际情况和获取方式确定费用项目和费用性质，正确地计取其他直接非人力成本，未发生的不计，不得重复计取。

注：本条所述的其他直接非人力成本，与T/BSCEA 005-2025、T/CCUA 060-2025中的其他费用存在不同，后者是信息化项目建设过程中其他第三方服务的费用，通常属于建设其他费。

5 评估过程

5.1 评估总体流程

人工智能应用开发项目造价评估的总体流程如图2所示。



图 2 人工智能应用开发项目造价评估总体流程

5.2 确认项目范围

明确人工智能应用开发项目的业务需求和建设目标，确定其中软件开发的范围、数据需求、模型建设需求，包括模型性能要求。可通过查阅文档资料（如项目立项书、建设方案、需求文档等）、调研访谈等方法收集并确认项目的开发工作内容，将人工智能应用开发活动分解为主体功能开发、数据处理、模型建设三部分，对其中的数据处理、模型建设按附录A示例进一步分解为任务包。

5.3 明确费用分类

在确认项目范围的情况下，应依据本文件第4章确定本项目应用开发的费用分类，包括人工智能应用开发活动涉及的主体功能开发费、数据处理费和模型建设费，以及人工智能应用开发所需的软硬件资源、基础环境及第三方DaaS、MaaS服务等直接非人力成本。

在具体评估过程中，可根据管理规定进行费用分类，未发生的不计，不得重复计取。

5.4 选择评估方法

人工智能应用开发项目造价评估时应根据项目特点、费用分类等要素选择合适的评估方法。常见的评估方法包括类比法、类推法、方程法及经验法等。方法的应用说明如下：

——经验法：通过工作任务分解的形式，以最小任务包或交付成果确定相应的费用。

——类比法：参考国家或地区公布的产品指导价；

——类推法：参考以往类似中标项目或项目合同案例中的有效价格（合同法）或通过市场询价方式确定市场行情价格（市场询价法）；

——方程法：建立参数模型，通过输入各项参数确定项目费用；或基于原有成本，确定与费用相关的比例系数，通过这些系数对基础数据进行调整和计算，从而评估出费用（比例系数法）；或以软件所具备的功能为基础，将软件功能分解为不同类型的功能点，通过对这些功能点的识别、分类、复杂度评估以及量化计算，最终评估出所需的费用（功能点法）；或建立组织的任务包分类标准，根据组织的历史项目数据分析确定每类标准任务包的基准工作量，以及任务包工作量的调整因子、相应开发岗位的人力费率，将人工智能应用开发活动分解为任务包并归类，典型任务包分解示例见附录 A，按每类标准任务包的基准工作量和调整因子计算人工智能应用开发活动涉及的各任务包的工作量，根据各类标准任务包的开发人力费率计算开发费用（基准工时法）。

人工智能应用开发项目造价评估时，根据项目具体情况确定项目涉及的费用分类，其中主体功能开发费应采用本文件6.1的方法进行评估，数据处理费应采用6.2的方法进行评估，模型建设费应采用6.3的方法进行评估，软件开发所涉及的直接非人力成本应采用第7章所述的各类费用评估方法分别进行评估。

5.5 评估项目费用

根据项目确定的费用类型和各类费用评估方法，对项目所涉及的各类费用分别进行评估，汇总得到项目造价。

采用不同的评估方法进行人工智能应用开发项目造价评估时，要求如下：

a) 在有类似的成功案例时，宜参考成功案例进行评估；
b) 应充分利用行业基准数据对各项费用进行评估；
c) 在费用评估过程中宜采用不同的方法分别评估并进行交叉验证，如果评估结果差异较大，宜采用专家评审的方法确定评估结果；
d) 评估结果宜为一个范围值；
e) 在评估过程中出具评估报告时，应按照 T/BSCEA 002-2024 执行；
f) 评估结果应形成人工智能应用开发项目造价汇总表，示例见附录 B。

6 综合人力成本评估

6.1 主体功能开发费

6.1.1 评估方法

主体功能开发费应依据GB/T 36964-2018进行评估。

6.1.2 评估过程

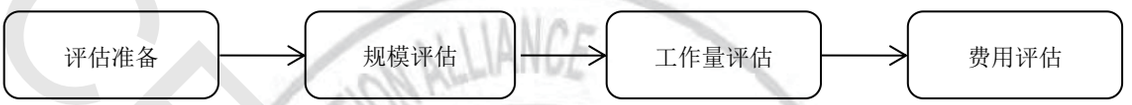


图 3 主体功能开发费的评估过程

主体功能开发费的评估过程分为四步，如图3所示：

a) 评估准备：明确人工智能应用项目主体功能开发的评估范围，宜根据本文件涵盖业务需求中的功能开发、人工智能应用软件与业务系统、人工智能平台或第三方 DaaS、MaaS 的系统进行应用集成、数据处理与模型建设中的人工智能平台功能开发这三部分，其中与人工智能技术直接相关的功能（如智能推理接口、模型调用、知识库查询等），可根据项目实际情况制定人工智能应用场景下特殊的功能点计数规则；

b) 规模评估：按照 GB/T 36964-2018 规定的功能点方法标准，对评估范围内的软件功能进行规模评估；

c) 工作量评估：根据规模评估的结果，按照 GB/T 36964-2018 规定的方法和行业基准数据评估工作量，根据项目的软件因素、开发因素等调整因子对工作量评估结果进行调整，确定工作量的合理范围；其中软件因素之一的应用类型调整因子宜按照一般业务处理类进行取值；

d) 费用评估：根据工作量评估结果评估主体功能开发费，人月费率取值宜参照软件行业基准数据中的软件开发人月费率，组织也可根据历史项目数据、结合人力资源市场行情、行业基准数据及各地人力资源和社会保障部门发布的相关岗位人力薪酬统计数据等综合分析确定。

6.1.3 评估说明

在评估人工智能应用的功能规模时，应重点关注与业务场景相关的用户功能、人工智能平台相关功能，以及人工智能应用与业务系统、人工智能平台及第三方DaaS、MaaS系统进行应用集成涉及的功能。主体功能费评估注意事项如下：

a) 应用类型调整：当人工智能应用开发项目按照本文件分为主体功能开发、数据处理与模型建设三部分分别评估时，主体功能开发部分按功能点方法和行业基准数据进行评估，不对软件的应用类型系数进行调整；当人工智能应用开发项目不单独评估数据处理、模型建设两部分的费用，而是整体作为软件开发项目按 GB/T 36964-2018 方法来评估费用时，宜依据项目采用的技术实现路径（如模型训练与微调、智能体编排、API 集成或系统融合开发）对软件的应用类型系数进行调整，以反映人工智能应用软件开发的复杂性；

b) 复用程度评估：人工智能应用开发中，应充分考虑对已有模型、算法、组件及工具的复用程度，在评估时标识可复用的部分并相应调整工作量；

c) 非功能评估：主体功能开发费包含了除数据处理和模型建设外的其他非功能需求开发的相关成本，相应的非功能需求的工作量和费用体现在评估模型的生产率与质量特性调整因子中；

d) 人工智能辅助开发的影响与调整：主体功能开发的工作量及费用宜根据项目采用人工智能进行软件开发的实际情况进行调整。

6.2 数据处理费

6.2.1 评估方法

数据处理费用涵盖由项目开发团队负责的传统人工智能应用、基于预训练模型的智能体建设中所需的数据处理活动而产生的综合人力成本。数据处理活动包括数据准备、数据整理、数据采集、数据清洗、数据标注、数据预处理、数据质量检测、数据安全等。

数据处理活动的工作量及费用可采用经验法、类推法、方程法进行评估，宜采用多种评估方法交叉验证。当多种方法的评估结果偏差在合理范围内时，宜取加权平均值作为评估结果；若偏差超出合理范围，应检查评估依据、参数、定额标准、比例取值等，并重新评估数据处理工作量及费用。

6.2.2 评估过程

人工智能应用开发项目的数据处理费的评估过程如图4所示。

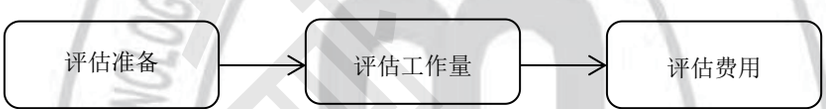


图 4 数据处理费的评估过程

评估过程如下：

a) 评估准备：依据人工智能模型建设需求，明确数据处理的工作要求，梳理数据类型、数据规模、质量验收标准、交付成果、交付周期等核心内容；对数据处理工作进行任务分解，对各任务包工作内容进行合理裁剪、适配选用；

b) 评估工作量：依据数据处理各任务包工作内容、数据规模、数据处理效率、基准工作量等，评估各任务包工作量，并根据数据种类、数据规模、数据质量、数据处理的复杂度等调整工作量；

c) 评估费用：根据各任务包工作量评估结果，对各任务包匹配对应岗位人力费率，汇总评估数据处理费。可根据数据处理费占人工智能应用开发综合人力成本的比例，验证数据处理费评估结果是否合理。

6.2.3 评估说明

6.2.3.1 评估准备

数据处理工作量及费用评估前，应依据数据处理工作要求进行任务包分解，明确各任务包工作内容、交付成果并评估数据规模。

数据规模评估时，宜综合考虑业务复杂度、模型参数量、模型性能目标、数据类型、数据质量水平等因素；宜在满足模型性能目标的基础上合理管控数据规模；可采用学习曲线分析等试验手段动态校核，确定合理的数据规模。不同类型模型所需数据规模示例见附录C.2。

6.2.3.2 评估工作量

6.2.3.2.1 评估单个任务包的工作量可采用经验法、类推法、方程法评估。方程法包含计件方程法、基准工时方程法等。其中计件方程法适用于数据规模易量化、数据加工效率稳定的场景，其余数据处理评估宜采用基准工时方程法。方法如下：

a）计件方程法

依据数据处理各任务包数据规模、数据加工效率评估各任务包工作量。

计算公式如下：

$$DPE_i = DS_i \times UPE_i \quad (1)$$

式中：

DPE_i——Data Processing Effort，数据处理第i个任务包工作量；

DS_i——Data Size，第i个任务包的待处理数据规模；

UPE_i——Unit Processing Efficiency，第i个任务包的单位数据加工效率。

单位数据加工效率宜采用行业或组织基准数据，也可通过项目小批量样本实测确定后再开展数据处理工作量评估。单位数据加工效率分为人工作业、自动化作业两类。

b）基准工时方程法

数据处理工作量评估时，开发团队可依据行业或组织基准工作量，选用适配的调整因子确定任务包工作量。

计算公式如下：

$$DPE_i = DBE_i \times AF \times DRDF \quad (2)$$

式中：

DPE_i——Data Processing Effort，数据处理第i个任务包的工作量；

DBE_i——Data Baseline Effort，数据处理第i个任务包的基准工作量，宜根据组织的历史数据分析结果确定标准任务包及基准工作量取值；

AF——Adjustment Factor，数据处理工作量调整因子；调整因子取值范围宜根据组织的历史数据分析结果，结合专家经验综合确定；

DRDF——Data-Processing Research & Development Factor，数据处理开发因素调整因子，通常由数据复用程度、数据处理工具化程度等因子相乘确定。

说明：

1）调整因子依据各任务包的工作量影响因素取值，典型影响因素见表1。数据处理基准工作量及调整因子示例见附录C.3。

表 1 数据处理各任务包工作量典型影响因素

编号	任务包名称	典型影响因素
WBS2.1	数据采集	数据类型、数据规模、采集方式、系统对接难度、业务复杂度等
WBS2.2	数据清洗	数据类型、数据质量、清洗复杂度、人工复核量等
WBS2.3	数据标注	数据类型、数据规模、标注复杂度、领域知识要求、质量要求等
WBS2.4	数据预处理	数据类型、特征维度、预处理复杂度、知识库文档量等

2）也可将以上数据处理任务包合并进行整体工作量评估，见附录C.4。

6.2.3.2.2 汇总数据处理所涉及的各个任务包的工作量评估结果，得到数据处理的总工作量，计算公式如下：

$$TDE = \sum_{i=1}^{n} DPE_i \quad (3)$$

式中：

TDE——Total Data Effort，数据处理总工作量。

6.2.3.3 评估费用

数据处理费用应依据各任务包工作量，结合对应岗位人力费率计算各任务包费用，最终汇总计算数据处理费用，计算公式如下：

$$TDC = \sum_{i=1}^{n} (DPE_i \times PF_i \times UHC) \quad \text{.....(4)}$$

式中：

TDC——Total Data Cost，数据处理综合人力成本；

PF_i——Position Factor，承担第i项数据处理活动的开发人员对应的人力费率的岗位系数；

UHC——Unified Human Cost，人工智能应用开发的综合人力费率；

评估费用时：

a) 数据处理岗位人力费率宜根据人力资源市场行情数据、行业基准数据及各地人力资源和社会保障部门发布的相关岗位人力薪酬统计数据、组织各类岗位的人力费率数据分析确定，取值示例见附录A；
b) 数据处理若采购第三方服务，购买服务费用应计入直接非人力成本，不得重复计算；
c) 数据处理采用自动化方式时需单独测算对应工具、脚本的开发或购置、租赁成本，工具、脚本开发费应计入6.1主体功能开发费，购置费、租赁费应计入直接非人力成本，不得重复计算；
d) 数据处理评估示例见附录C.4；
e) 可根据项目数据处理费用的占比来核验数据处理费用评估结果的合理性。

6.3 模型建设费

6.3.1 评估方法

人工智能应用开发项目的模型建设费可采用类推法或者方程法进行评估。

人工智能应用开发项目的建设方案清晰时，宜采用方程法评估模型建设的工作量和费用。

评估过程中宜采用不同的方法分别评估并进行交叉验证，如果评估结果差异较大，可采用专家评审的方法确定评估结果。

在评估工作量及费用时，宜给出评估结果的范围而不是单一的值。

6.3.2 评估过程

人工智能应用开发项目模型建设费的评估过程如图5所示。

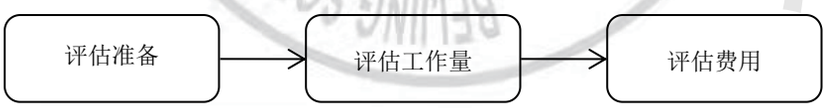


图5 模型建设费的评估过程

评估过程分三步：

a) 评估准备：根据项目需求和建设方案等材料，确定项目范围和目标，明确人工智能的能力边界，进行任务分解，明确模型建设活动对应的任务包，确定模型建设活动的评估范围，以及各任务包工作量的主要影响因素，如系统因素、开发因素、数据因素、算法因素等；

b) 评估工作量：根据项目任务分解结果，对模型建设活动所涉及的每个任务包分别评估工作量，工作量评估过程中应根据影响该任务包工作量的相关因素，如数据、算法、业务、开发等因素，对任务包的工作量评估结果进行调整，汇总各任务包的工作量得到模型建设的总工作量；

c) 评估费用：根据模型建设活动的各任务包工作量评估结果，按承担各任务包开发的人员岗位对应的人力费率评估各任务包的费用，汇总后得到模型建设活动的费用。

6.3.3 评估说明

6.3.3.1 评估准备

6.3.3.1.1 宜参考行业的通常做法，收集组织的人工智能应用开发历史项目信息，对人工智能应用主要开发活动建立统一的任务包分解结构，每个任务包交付一个明确的工作产品，人工智能模型建设常见的典型任务包示例见附录A，包括但不限于：传统算法模型构建、基于预训练模型的模型微调、提示词工程、RAG知识库构建、智能体编排等，组织宜根据实际情况自定义本组织的任务包分解标准。

6.3.3.1.2 宜根据组织人工智能应用开发项目历史数据进行分析，结合行业基准数据确定各典型任务包的基准工作量的合理范围，以及任务包工作量的主要影响因素，包括局部影响因素和全局影响因素，建立单个任务包的工作量评估方程，以及模型建设活动的工作量评估方程。

6.3.3.1.3 模型建设工作量的主要全局和局部影响因素包括但不限于下列内容。

a) 人工智能应用系统的业务能力级别要求：应将需方对人工智能应用的业务能力（如智能语音、自然语言理解、智能图像视频分析等）要求进行分级，作为全局影响因素对模型建设的总工作量进行调整，宜根据历史项目数据分析结果，结合专家经验确定各级别的取值。

b) 质量因素：宜将影响人工智能应用开发工作量的主要质量因素作为调整因子，对各任务包的工作量评估结果及总工作量评估结果进行调整，质量因素包括但不限于功能的正确性、鲁棒性等质量特性，可将各主要的质量影响因素进行分级，如低、中、高，各级别的取值宜根据数据分析结果，结合专家经验确定；组织宜按照ISO/IEC 25059-2023中的质量特性，根据组织的历史数据分析结果自定义质量因素调整因子及其取值。

c) 开发相关的影响因素，包括影响总工作量的全局影响因素，以及仅影响部分任务包的局部影响因素，包括但不限于：

1) 开发团队因素：如开发团队的经验、规模和人员能力等，取值应参考软件行业基准数据中的团队系数取值；

2) 技术与工具成熟度因素：开发所采用的主要人工智能技术、主要人工智能开发工具链的成熟度等；可将技术与工具的成熟度进行分级，如低、中、高，各级别的取值宜根据数据分析结果，结合专家经验确定；

3) 任务包的自研比例：项目团队在典型任务包的开发过程中，可复用已有的成果物，或者通过采购第三方提供的服务来完成任务包的开发工作，应明确任务包的自研比例，进行分级，如不涉及、少量、大量、全量，各级别的取值宜根据数据分析结果，结合专家经验确定；应根据项目的实际情况分别确定各任务包的自研比例；

4) 任务包的工具支持程度：项目团队在任务包的开发过程中，可利用人工智能平台已有的工具来帮助项目团队提升工作效率，降低工作量，应明确工具对任务包的支持程度，进行分级，如无、低、中、高，各级别的取值宜根据数据分析结果，结合专家经验确定；应根据项目的实际情况分别确定各任务包的工具支持程度；

5) 数据因素：数据因素除对数据处理的工作量有影响，对模型建设的工作量也有影响，模型建设活动应考虑的数据因素包括但不限于数据的质量、数据量等；可将各主要的影响因素进行分级，如低、中、高，各级别的取值宜根据数据分析结果，结合专家经验确定；

6) 算法因素：影响模型建设各类标准任务包工作量的算法因素包括但不限于：机器学习算法的种类、数量、算法复杂度、基础模型的参数规模、微调方式、提示词总数、单条提示词包含的规则数、知识库文档数量、知识图谱建设等级、智能体数量、智能体编排的流程分支与节点数量、工具调用数量、模型压缩方式、模型部署方式等；每个标准任务包工作量的算法影响因素不同，应分别识别；可将各主要的影响因素进行分级，如低、中、高，各级别的取值宜根据数据分析结果，结合专家经验确定。

d) 根据项目建设方案进行任务分解，识别模型建设所涉及的各任务包，以及工作量主要影响因素的具体情况，确定模型建设活动的评估范围，注意不应与 6.1 主体功能开发及 6.2 数据处理的工作范围重叠而导致部分工作量重复计算。

6.3.3.2 评估工作量

6.3.3.2.1 评估单个任务包的工作量时，根据所属的标准任务包类别的基准工作量、该任务包工作量影响因子的具体情况，以及该任务包的自研比例和工具支持程度来评估工作量。

示例：假设基于基准数据建立的单个任务包工作量回归方程如下：

$$TE_i = TBE_i \times TF_i \times SR_i \times TS_i \times SF \times MRDF \quad (5)$$

式中：

TE_i——Task Effort，第i个任务包的工作量；

TBE_i——Task Baseline Effort，第i个任务包的基准工作量，宜根据组织的历史数据分析结果确定标准任务包及基准工作量取值；

TF_i——Task Factor，第i个任务包工作量的局部调整因子，根据任务包的不同，可包含多个不同的局部影响因素，宜根据组织的历史数据分析结果，结合专家经验综合确定任务包局部调整因子的取值范围；

SR_i——Self-developed Ratio，第i个任务包的自研比例，即任务包开发工作量中扣除外包负责以及复用已有成果物后实际由项目团队开发的工作比例，可分为：不涉及、少量、大量、全量，宜根据组织的历史数据分析结果，结合专家经验确定任务包自研比例的取值；

TS_i——Tools Supporting，现有开发工具或平台功能组件对第i个任务包的支持程度，工具支持程度越高，则人工工作量越省，可分为：无、低、中、高，宜根据组织的历史数据分析结果，结合专家经验确定任务包工具支持度的取值；

SF——System Factor，人工智能应用的全局性系统因素调整因子，包含多个系统因素的综合影响结果，包括但不限于如下因素：

SF_1——人工智能应用的业务能力级别要求，分为1～5级共5个级别，宜根据组织的历史数据分析结果，结合专家经验确定业务能力调整因子的取值范围；

SF_2——人工智能应用的质量因素调整因子，包含但不限于功能正确性（准确率）、鲁棒性、性能、可靠性等，宜根据组织的历史数据分析结果将各主要质量因素对工作量的影响程度进行分级，结合专家经验确定质量因素调整因子综合影响结果的取值范围；人工智能系统的各质量特性因素参见ISO/IEC 25059-2023的说明；

系统因素SF由各个子项按公式计算得到：$SF = SF_1 \times SF_2$；

MRDF——Model Research & Development Factor，模型建设全局性开发因素调整因子，是与人工智能应用自身属性无关，但与开发的团队、技术及工具等相关的因素，包括但不限于如下因素：

$MRDF_1$——模型建设开发团队经验，包括人工智能应用开发经验、业务经验，取值与中国软件行业基准数据中的团队经验调整因子保持一致；

$MRDF_2$——本项目所采用的主要人工智能技术与主要开发工具的成熟度，可将其对工作量的综合影响分级并确定调整系数，组织宜根据历史数据分析结果，结合专家经验自定义调整因子及取值范围。

模型建设开发因素MRDF由各个子项按公式计算得到：$MRDF = MRDF_1 \times MRDF_2$ 。

6.3.3.2.2 模型建设典型任务包及工作量的主要局部影响因素，包括但不限于：

a) 基于传统的机器学习算法构建某类用途的算法模型，如预测规划类任务包、NLP 类任务包、SR 类任务包、OCR 识别类任务包、CV 类任务包等，主要的工作量影响因素包括但不限于：模型类型、算法的种类、数量、算法复杂度、算法参数调整量、算法开发方式（如开源或已有算法、开源算法微调、自研算法）、训练样本数据量、数据质量等；

b) 模型搭建，基于预训练模型搭建人工智能模型，主要的工作量影响因素包括但不限于：基础模型的参数规模量级、模型调度涉及的基础模型数量、业务场景的复杂度、数据质量等；

c) 模型微调，主要工作量影响因素包括但不限于：微调方式、数据质量等；

d) 知识库检索增强，主要工作量影响因素包括但不限于：知识库文档数量、知识图谱建设复杂度等；

e) 提示词工程，主要工作量影响因素包括但不限于：提示词总量、单条提示词包含的规则数量等。

6.3.3.2.3 模型建设任务包工作量评估时：

a) 模型建设的典型任务包、主要的全局及局部影响因素及取值见附录 D，组织可根据自身的历史数据分析结果，结合专家经验对典型任务包、各类影响因素及取值进行自定义；

b) 模型建设活动的工作量评估结果，不含数据处理及主体功能开发的工作量，仅含本项目涉及的模型建设标准任务包的活动工作量，不涉及的任务包不计算；

c) 模型建设中如果同种任务包存在多个，如构建多个智能体，则每个任务包分别评估工作量，并根据成果物的复用情况调整自研比例；

d) 对于每个独立的人工智能模型或智能体，其所对应的数据处理活动应分别按 6.2 所述方法评估工作量。

6.3.3.2.4 传统人工智能模型的判定及计量规则如下所述。

a) 同时满足以下条件，可计为 1 个独立的传统人工智能模型：
   1) 识别或预测目标对象不同；
   2) 业务管控规则、判定逻辑不同；
   3) 需要单独采集数据、单独标注、单独训练。

b) 同一识别任务仅部署在多站点，只算 1 个模型。

c) 同一软件内包含多项独立识别任务，按任务数量计算模型数量。

d) 共用数据集、共用算法框架、仅变更部署点位，不重复计算模型。

6.3.3.2.5 独立智能体的判定及计量规则如下所述。

a) 同时满足以下条件，可判定为 1 个独立的人工智能智能体：
   1) 具备独立业务入口、独立交互界面或独立 API；
   2) 拥有独立知识库、提示词、编排逻辑与业务规则；
   3) 可独立部署、独立启停、独立交付、独立完成业务目标。

b) 同一闭环业务场景无论内部包含多少流程节点、智能体编排环节，均计为 1 个智能体。

c) 内部子任务、多角色协同、条件分支不单独计为独立智能体。

d) 仅通过调整因子体现复杂度，不重复计量。

6.3.3.2.6 汇总模型建设所涉及的各个任务包的工作量评估结果，得到模型建设的总工作量，计算公式如下：

$$TME = \sum_{i=1}^{n} TE_i \quad (6)$$

式中：

TME——Total Model Efforts，模型建设总工作量；

TE_i——Task Effort，第i个任务包的工作量。

6.3.3.3 评估费用

评估费用前，应根据组织的人工智能应用开发历史项目各类岗位的人力费率数据，结合人力资源市场行情、行业基准数据和各地人力资源和社会保障部门发布的相关岗位人力薪酬统计数据等进行分析，综合确定人工智能应用开发的综合人力费率，以及各主要开发岗位对应的人力费率的岗位系数。

评估费用时，根据模型建设所涉及的各项任务包的工作量评估结果，确定各项任务包对应的开发人员岗位，按综合人力费率和岗位系数评估模型建设的费用。

计算公式如下：

$$TMC = \sum_{i=1}^{n} (TE_i \times PF_i \times UHC) \quad (7)$$

式中：

TMC——Total Model Costs，模型建设综合人力成本；

TE_i——Task Effort，第i个任务包的工作量；

PF_i——Position Factor，承担第i项模型建设活动的开发人员对应的人力费率的岗位系数；

UHC——Unified Human Cost，人工智能应用开发的综合人力费率。

评估费用时：

a) 人工智能应用开发的综合人力费率、岗位系数，宜参考组织历史数据分析结果、行业基准数据、各地人力资源和社会保障部门发布的人力薪酬统计数据综合确定，取值参见附录 A，完整的项目评估示例参见附录 E；

b) 模型建设费是开发团队承担模型建设的各项任务包产生的综合人力成本，不含数据处理及主体功能开发的相关费用；

c) 如模型建设活动中的部分任务包通过采购第三方提供的模型服务实现，相关的模型服务费用计入直接非人力成本，项目开发团队因此减少的工作量和费用，应从模型建设活动工作量和费用中扣除，不得重复计算。

7 直接非人力成本评估

7.1 软硬件购置费

7.1.1 评估方法

软硬件购置费评估可采用类比法或类推法。

评估硬件设备中的算力设备时应明确算力设备为裸机或整机（包含硬件、基础系统和驱动），采用类比法、类推法中的一种或多种方法综合确定购置费。

评估就绪可用软件中人工智能应用开发环境所需的支撑软件时，应明确购置软件使用方式（永久、年或用户数）和购置软件费用是否包含系统集成费用，采用类比法、类推法中的一种或多种方法综合确定购置费。

其他软硬件购置费评估方法应按 T/BSCEA 005-2025、T/CCUA 060-2025 执行。

7.1.2 评估过程

软硬件购置费的评估过程如图6所示。



图 6 软硬件购置费的评估过程

其中：

a) 信息获取：了解产品品牌、型号、伴随服务、质保等基本信息；
b) 询价准备：梳理采购清单中的重点参数，按同类型设备分割清单，制作询价单；
c) 价格收集：收集与整理报价单、成交记录；
d) 价格确定：根据收集结果确定合理的价格。

7.1.3 评估说明

评估软硬件购置费用时：

a) 评估算力设备时，宜明确算力设备所需电力和散热。其他注意事项可参照信息化项目造价评估相应标准。询价时，需获取不少于三家单位的正式报价，可在剔除异常数据后选取中值作为参考价格；
b) 市场询价结果需形成询价记录单；
c) 本项目开发所需专用软硬件由开发方代业主方购置并用于项目开发，项目结束后所购置的软硬件作为资产移交给业主方，相应的购置费计入项目开发成本的直接非人力成本，全额列入项目预算中；项目决算进行开发成本核算时，应按财务规定将软硬件资产折旧分摊计入本项目的开发成本；非本项目专用的软硬件购置费，不计入本项目开发成本的直接非人力成本。

7.2 资源租赁费

7.2.1 评估方法

7.2.1.1 资源租赁费用计算公式如下：

$$\mathrm{RRF}=\sum_{i=1}^{n}(RUP_i\times N_i\times LT_i) \quad (8)$$

式中：

RRF——资源租赁费用；

RUP——资源租赁单价；

N——数量；

LT——租赁期。

资源租赁数量及租赁期可参考设计方案确定。

7.2.1.2 资源租赁单价评估时，应首先明确资源的技术参数、配置、品牌、型号、地域、售后服务要求等相关参数，再采用类比法、类推法、方程法中的一种或多种方法综合确定租赁单价。

a) 采用类比法评估资源租赁单价时，参考国家或地区公布的产品指导价，将所需租赁资源的核心属性与产品指导价中的基准数据进行比对，进而得到资源租赁单价；
b) 采用类推法评估资源租赁单价时，参考以往类似中标项目或项目合同案例中的有效价格或通过市场询价的方式，得到资源租赁单价；
c) 采用方程法评估资源租赁单价时，需根据资源购置单价，租赁方式所产生的资金成本，以及租赁期内的运维费用，以及资源折旧（分摊）年限来计算，计算公式如下：

$$RUP = （UP + CC）/ DL + OMUP \quad (9)$$

式中：
RUP——资源租赁单价；
UP——资源购置单价；
CC——资金成本，按所在地方的政策规定计算；
DL——折旧（分摊）年限；
OMUP——租赁期内的资源运维单价，需扣除免费服务期后计算资源运维单价。
资源购置单价的评估应符合7.1的相关规定。

7.2.2 评估过程

基本过程见下图：

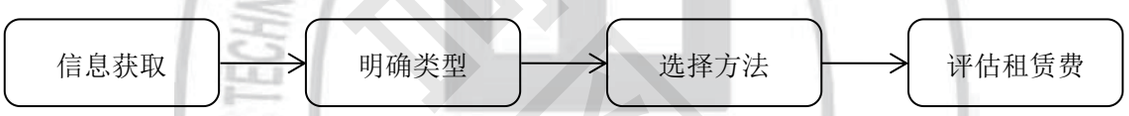


图 7 资源租赁费用的评估过程

其中：
a) 信息获取：梳理所需租赁资源清单的重点参数（如品牌型号、参数要求、数量、租赁期、地域、用途、售后服务等），按类型分割清单，制作询价单；
b) 明确类型：评估租赁费用前，应先确定资源租赁费用的类型。可通过查阅文档资料（如项目说明书、项目建设方案、需求文档等）、调研访谈等方法收集并确认资源租赁的类型；
c) 选择方法：评估资源租赁费用时，应根据项目特征、费用分类等要素采用类比法、类推法、方程法中的一种或多种方法；
d) 评估租赁费：根据确定的评估方法评估资源租赁单价，宜采用多种方法进行交叉验证。

7.2.3 评估说明

评估资源租赁费用时：
a) 参考产品指导价时，宜以项目所在地公布信息为首要参考文件；
b) 参考以往类似中标项目价格时，宜优先选择项目所在单位以往类似中标项目价格；
c) 中标项目或项目合同案例宜为近一年内成交项目；
d) 市场询价时，需获取不少于三家单位的正式报价，可在剔除异常数据后选取中值作为参考价格；
e) 开源、利旧、统一采购或免费提供的就绪可用软件产品，以及利旧、统一采购或免费提供的硬件设备产品，不再单独计取租赁费用；
f) 市场询价结果需形成询价记录单；
g) 评估结果需形成机房（柜）租赁费用明细表、软硬件租赁费用明细表、云服务（通信链路）租赁费用明细表。

7.3 人工智能环境集成费

7.3.1 评估方法

人工智能环境集成实施包括人工智能专用的硬件系统集成、软件系统集成和安全系统集成。

人工智能环境集成费评估宜采用比例系数法和经验法。

a) 比例系数法：参考各地或各组织信息化项目预算管理指导文件，以项目集成对象相关的设备、就绪可用软件购置费等为基数，乘以相应费率和调整系数计取费用；

b) 经验法：借助行业专家经验，定性定量结合，评估集成所需投入的人力成本和工具费用。将集成任务拆分成多个可管理、可度量的子任务，如硬件部署、软件调试、数据整合等，逐项计算人力工作量（含方案设计及测试验证）、工具使用等费用，然后进行汇总。

7.3.2 评估过程

7.3.2.1 概述

基本过程见下图：

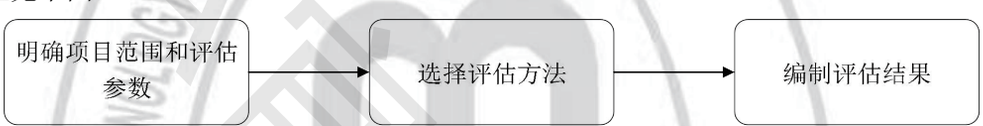


图 8 人工智能环境集成费的评估过程

7.3.2.2 明确项目范围和评估参数

7.3.2.2.1 界定项目人工智能环境集成的范围，收集计算基数，确定集成费率：

a) 首先确定人工智能环境集成的具体内容（硬件集成、软件集成等），再收集相关设备、就绪可用软件购置费作为计算基数；

b) 按 T/BSCEA 005—2025、T/CCUA 060-2025 中系统集成费的费率及调整因子说明，确定本项目集成对象的集成费率。

7.3.2.2.2 分析项目特性，确认调整系数：

a) 项目规模（如设备数量、软件模块数、集成点数等）；

b) 技术复杂度（如是否涉及跨平台集成、高可用性架构、安全等级要求等）；

c) 行业合规特性（如金融、医疗、交通、军队等行业的特殊监管要求）；

d) 实施区域（本地、跨区域或特殊实施环境）；

e) 系统集成服务能力要求（如实施团队或个人需具备相应的技术资格条件）。

7.3.2.3 选择评估方法

7.3.2.3.1 采用方程法（比例系数）时，参考行业或所属地区发布的系统集成服务费用定额标准，注意应按集成对象性质规定分档计取。若无明确定额标准，可参考市场平均水平或同类项目案例。

计算公式如下：

$$\mathrm{TC} = \sum_{i=1}^{n} (SC_i \times R_i \times S_i \times F_i) \quad (10)$$

式中：

TC——人工智能环境集成实施总费用；

SC——集成对象（硬件、软件）各子项的购置费基数；

R——各子项对应的取费费率；
S——各子项投资规模调整系数；
F——各子项特性影响因子。

7.3.2.3.2 评估人工智能环境集成实施费用时：

a) 需求澄清：应确保评估前明确集成对象范围（是否包含硬件采购、就绪可用软件、定制化软件开发、培训等）；
b) 风险调整：对高风险因素（如技术不确定性、合规要求严格）的评估，可根据具体情况设置调整因子；
c) 相关调整系数取值按 T/BSCEA 005—2025、T/CCUA 060-2025 计取；
d) 采用经验法时，应科学合理分解工作任务，描述任务要求、分析项目特性，确认调整系数，定义人员级别单价数据，分步计算各项工作量及费用，汇总得到评估结果。

注：集成对象及集成要求明确时适用方程（比例系数）法，复杂度高但可做任务拆解的项目适用经验法。

7.3.2.4 编制评估结果

根据项目人工智能环境集成范围相关参数和所选的评估方法，评估集成实施费用，编制评估结果。

7.3.3 评估说明

评估人工智能环境集成费时，人工智能应用开发项目涉及所开发的软件与数据采集系统（含采集设备）、业务系统、人工智能支撑平台或外部数据服务、模型服务等系统的应用集成费用，以及人工智能应用开发项目所需进行的数据采集、数据清洗、标注、格式转换和治理等数据处理费用，应计入主体功能开发费，不计入人工智能环境集成费。

7.4 数据资源与数据服务费

7.4.1 评估方法

数据资源与数据服务费评估可采用类推法进行市场询价。

7.4.2 评估过程

基本过程如下图所示。



图 9 数据资源与数据服务费的评估过程

其中：

a) 信息获取：了解数据的具体需求，包括数据购置（数据购买或授权使用）、数据处理工具、数据采集服务、数据存储服务、数据标注服务、数据预处理服务、数据分析服务、数据治理服务、数据安全服务等服务要求；
b) 询价准备：梳理各类数据资源的服务要求，制作询价单；
c) 价格收集：收集与整理报价单、成交记录；
d) 价格确定：根据收集结果确定合理的价格。

7.4.3 评估说明

评估数据资源与数据服务费时：

a) 不包含开发方在开发期内自行完成相关的数据工程活动产生的费用；
b) 市场询价结果需形成询价记录单。

7.5 模型服务费

7.5.1 评估方法

模型服务费用评估可采用类推法进行市场询价。

7.5.2 评估过程

基本过程见下图：

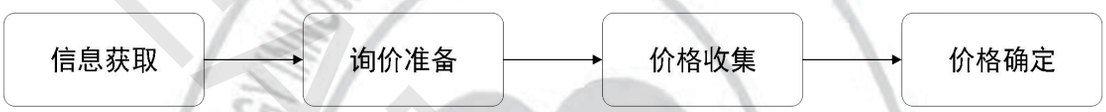


图 10 模型服务费用的评估过程

其中：

a) 信息获取：了解所需购买授权、租赁服务或调用接口模型服务所需的服务内容、服务方式（按量计价、包月套餐或包月限量套餐+超出额度后按量计价等方式）等基本信息；
b) 询价准备：梳理采购清单中的重点参数，制作询价单；
c) 价格收集：收集与整理报价单、成交记录；
d) 价格确定：根据收集结果确定合理的价格。

7.5.3 评估说明

评估模型服务费用时：

a) 评估模型授权费用时，应明确授权模式、授权边界和授权周期；
b) 询价时，需获取不少于三家单位的正式报价，可在剔除异常数据后选取中值作为参考价格。市场询价结果需形成询价记录单。

7.6 其他直接非人力成本

7.6.1 评估方法

其他直接非人力成本的评估可采用经验法、类推法和方程法。

7.6.2 评估过程

基本过程见下图：

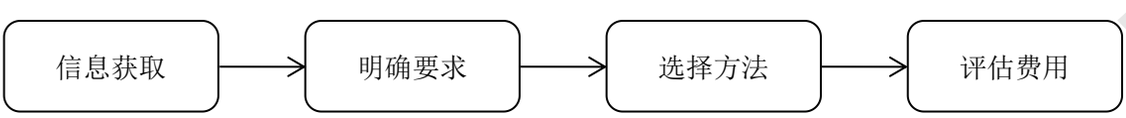


图 11 其他直接非人力成本的评估过程

其中：

a） 信息获取：了解其他直接非人力成本对应的目标产品或服务等基本信息；
b） 明确要求：梳理使用其他直接非人力成本的详细产品或服务要求，并同委托方确认；
c） 选择方法：选择合适的评估方法；
d） 评估费用：评估其他直接非人力成本。

7.6.3 评估说明

评估其他直接非人力成本时：

a） 应明确使用其他直接非人力成本的产品或服务；
b） 采用经验法时，应科学合理分解工作任务，描述任务要求、分析项目特性，确认调整系数，定义人员级别单价数据，分步计算各项工作量及费用，汇总得到评估结果；
c） 采用类推法进行市场询价时，宜获取不少于三家单位的正式报价，可在剔除异常数据后选取中值作为参考价格；
d） 采用类推法参考以往类似中标项目价格时，宜优先参考项目所在单位以往类似中标项目价格；
e） 采用方程法时，可参考行业或所属地区发布的费用定额标准，若无明确定额标准，可参考市场平均水平或同类项目案例。

BSCEA 北软院 BEIJING SOFTWARE COST EVALUATION TECHNOLOGY ASSOCIATION 非经授权不得使用

附 录 A
（资料性）
典型任务包及对应岗位系数示例

人工智能应用开发典型任务包及对应岗位系数取值示例见表 A.1。

表 A.1 人工智能应用开发典型任务包及对应岗位系数取值示例

任务类别	典型任务包示例	主要工作内容	岗位	岗位要求	岗位系数取值示例	评估说明
1. 主体功能开发	1.1 主体功能开发	常规人工智能应用软件开发活动，包括需求、设计、构建、测试、部署等软件工程活动，开发范围包含业务功能开发、应用系统集成开发、人工智能平台功能开发等。	国家职业编码： 4-04-05-01，计算机程序设计员	参见《国家职业技能标准 计算机程序设计员（2022 年版）》，本文件所述的岗位人员要求为该标准中的二级（技师）。	1.0	按 6.1 所述采用功能点法评估
2. 数据处理	2.1 数据采集	依据人工智能模型建设需求，通过业务系统对接、离线数据抽取、现场采集、归集多模态原始数据，完成数据源合规核验、台账登记、授权归档，按需采集增量数据。	国家职业编码： 4-04-05-05，人工智能训练师	参见《国家职业技能标准 人工智能训练师（2021 版）》，本文件所述的岗位人员要求为该标准中的二级（技师）。	0.9～1.0	按 6.2 所述采用方程法评估
	2.2 数据清洗	对原始数据执行数据去重、纠错、格式对齐、脱敏与异常数据人工复核修正，按需对迭代更新数据二次清洗，形成满足质量要求的合规数据集。				
	2.3 数据标注	开展多模态数据分类、目标检测、语义理解等真值标注，构建问答对、图文对等训练样本，经质检后划分训练、评测数据集，按需补充优化样本。				
	2.4 数据预处理	对标注数据集开展降噪、归一化、格式转换、非结构化解析等处理，完成特征提取筛选、样本均衡与特殊样本适配，合理划分训练、测试数据集，使数据符合模型训练接入规范；完成知识库文档的整理。				
3. 传统人工智能模型开发	3.1 传统人工智能模型开发	采用机器学习算法等技术进行人工智能模型的分析、研究、开发，并对人工智能系统进行设计、优化、管理和应用。	国家职业编码：2-02-10-09，人工智能工程技术人员	参见《国家职业技能标准 人工智能工程技术人员（2021 版）》，本文件所述的岗位人员要求为该标准的中级。	1.2～1.3	按 6.3 所述采用方程法评估
4.智能体开发	4.1 业务分析与方案设计	进行业务场景分析，明确业务目标，开展技术方案设计与论证。	国家职业编码：4-04-05-13，生成式人工智能系统应用员 （其中的智能体开发员）	相应的国家职业技能标准尚在制定中，本文件所述的岗位人员要求为该职业中的二级（技师）。	1.4～1.5	
	4.2 模型选型与搭建	选型基础模型，搭建与适配模型训练与推理基础环境，确定训练策略，设定评估指标体系。				

表A.1（续）

任务类别	典型任务包示例	主要工作内容	岗位	岗位要求	岗位系数取值示例	评估说明
4.智能体开发	4.3 模型微调	采用业务领域数据对基础模型进行训练和参数调优，提升特定业务场景下模型输出准确性。	国家职业编码：4-04-05-13，生成式人工智能系统应用员 （其中的智能体开发员）	相应的国家职业技能标准尚在制定中，本文件所述的岗位人员要求为该职业中的二级（技师）。	1.4～1.5	按6.3所述采用方程法评估
	4.4 RAG知识库构建	通过对业务领域知识文档进行切片、向量化处理后构建RAG知识库，形成知识图谱，提供检索支撑能力。				
	4.5 提示词工程	通过设计适配业务、结合检索知识的提示指令，提升模型的输出应答质量，并通过上下文工程管理信息流，负责对信息的供给、筛选与管理，构建动态、可扩展的认知环境，支撑复杂任务如多轮对话等。				
	4.6 智能体编排	面向复杂业务场景，将具备独立能力的业务智能体、工具调用、RAG知识库检索、业务规则节点，按照既定业务逻辑进行流程组装与统一调度，协调多个专业化的智能体，高效、可靠地协同完成复杂任务。				
	4.7 模型压缩与部署	模型压缩是采用模型量化、剪枝、知识蒸馏等技术，在可接受的精度损失范围内，最大限度地减少模型的计算量、内存占用和存储体积，让模型瘦身，以适应特定的硬件和资源约束；模型部署是将压缩后的模型部署到稳定、安全、可观测的生产环境，进行测试与上线发布。				

说明：

1) 本表中的任务包是本文件第6章人工智能应用开发综合人力成本构成的三部分（主体功能开发、数据处理和模型建设）所含的典型任务包示例，用于人工智能应用开发中各部分的工作量及费用评估，并非人工智能应用开发项目的详细WBS，部分任务未列出，如测试、项目管理等活动，评估时，将这些未列出的活动的工作量分摊到各典型任务包，形成评估的基础任务包，以便按第6章所述方法进行评估。各组织需根据实际情况定义本组织的基础任务包标准。

2) 人工智能应用开发的基准人力费率宜参考中国软件行业基准数据，不同岗位的人力费率差异通过岗位系数来体现。

3) 上述岗位的开发人员能独立承担相应岗位的开发工作，其能力不低于相应职级标准的中等职级要求；各组织宜根据组织和项目的实际情况综合确定岗位要求及岗位系数取值。

4) 实际项目中存在同一个开发人员承担多个岗位的职责（即一人多岗），其岗位系数宜按所承担的多个岗位中的最高岗位系数取值。

附 录 B
（资料性）
项目造价汇总表

人工智能应用开发项目造价汇总表见表 B.1。

表 B.1 人工智能应用开发项目造价汇总表

序号	费用名称	金额（单位：元）—下限	金额（单位：元）—中值	金额（单位：元）—上限	备注
一	**综合人力成本（含直接人力成本和间接成本）**				
1.1	主体功能开发费				
1.2	数据处理费				
1.3	模型建设费				
小计（一 = 1.1 + 1.2 + 1.3）					
二	**直接非人力成本**				
2.1	软硬件购置费				
2.2	资源租赁费				
2.3	人工智能环境集成费				
2.4	数据资源与数据服务费				
2.5	模型服务费				
2.6	其他直接非人力成本				
小计（二 = 2.1 + 2.2 + 2.3 + 2.4 + 2.5 + 2.6）					
总计（= 一 + 二）					

附 录 C
（资料性）
数据处理费评估参数及示例

C.1 概述

本附录给出人工智能应用开发的数据处理典型任务包及基准工作量，各类调整因子取值示例，相关参数取值基于本文件部分编写单位的实际项目数据分析结果确定，并进行了小范围的验证。由于项目样本数量有限，相关参数取值仅供本文件读者应用参考。鉴于人工智能技术发展迅速，实际应用中宜根据组织历史数据进行校正。此外，组织宜不断积累数据，定期测算和更新基准工作量及调整因子取值。

C.2 不同类型模型所需数据规模示例

不同类型模型所需数据规模示例见下表C.1。表中罗列常见人工智能模型建设所需数据规模示例。鉴于人工智能技术发展迅速，相关数据规模宜在实际应用中动态调整。

表C.1 不同类型模型所需数据规模示例

模型类型	任务类型	数据规模示例
传统人工智能模型	传统机器学习	训练样本量宜不少于模型特征数量的 10 倍。
	深度学习图像分类	多类别任务每个类别宜不少于 1000 张标注图像；二分类任务正样本不宜少于 100 个。
	深度学习目标检测	每个目标类别宜不少于 2000 个标注实例。
	视频逐帧标注	单类场景宜不少于 10000 帧。
智能体	模型微调	每类任务场景宜准备 500～2000 条完整执行样本。
	智能体 RAG 知识库构建	单领域知识片段不宜少于 2000 条；原始文档数量常规业务领域宜 50～500 份，垂直专业领域宜 300～2000 份。

C.3 数据处理基准工作量及调整因子示例

人工智能模型建设数据处理基准工作量及各类调整因子取值示例见表C.2。鉴于人工智能技术发展迅速，实际应用中宜根据组织数据进行校正，持续积累数据,定期测算和更新基准工作量及调整因子取值。

表C.2 数据处理基准工作量及调整因子示例

序号	典型任务包	基准工作量示例（人天）	影响因素示例	影响因素示例说明	影响因素分级示例	调整因子计算及取值示例	备注
1	传统人工智能的数据处理	11～19	数据类型 $DF_{1.1}$	简单：文本； 一般：图像音频； 复杂：视频、多模态融合；	简单：-2 一般：0 复杂：2	$AF_1 = 1.0 + (DF_{1.1} + DF_{1.2} + DF_{1.3}) \times 0.03$ 取值范围：0.79～1.21	
			数据规模 $DF_{1.2}$	小规模：样本量＜5000 条； 中规模：样本量在 5000～5 万条； 大规模：样本量>5 万条；	小规模：-3 中规模：0 大规模：3		

表C.2（续）

序号	典型任务包	基准工作量示例（人天）	影响因素示例	影响因素示例说明	影响因素分级示例	调整因子计算及取值示例	备注
1	传统人工智能的数据处理	11～19	数据质量 $DF_{1.3}$	高：低质量数据占比<5%； 中：低质量数据占比在 5%～20%之间； 低：低质量数据占比超过 20%；	高：-2 中：0 低：2	AF_1 = 1.0 + ($DF_{1.1}$ + $DF_{1.2}$ + $DF_{1.3}$) × 0.03 取值范围：0.79～1.21	
2	智能体建设的数据处理	33～53	数据类型 $DF_{2.1}$	简单：文本； 一般：图像音频； 复杂：视频、多模态融合；	简单：-2 一般：0 复杂：2	AF_2 = 1.0 + ($DF_{2.1}$ + $DF_{2.2}$ + $DF_{2.3}$) × 0.05 取值范围：0.65～1.35	
			数据规模 $DF_{2.2}$	小规模：样本量＜5000 条；知识库文档≤50 份； 中规模：样本量在 5000～5 万条；知识库文档 51～200 份； 大规模：样本量>5 万条；知识库＞200 份；	小规模：-3 中规模：0 大规模：3		
			数据质量 $DF_{2.3}$	高质量：低质量数据占比<5%； 中：低质量数据占比在 5%～20%之间； 低质量：低质量数据占比超过 20%；	高质量：-2 中：0 低质量：2		
3	——	——	数据复用程度 DR	无复用：全部数据新增加工； 部分复用：可复用存量数据占比≤50%； 较高复用：可复用存量数据占比＞50%； 完全复用：全部复用存量数据；	无复用：1 部分复用：0.67 较高复用：0.33 完全复用：0	DR 取值范围：0～1.0	DRDF = DR × DDT
4	——	——	数据处理工具化程度 DDT	无：无工具支持数据处理，全部人工处理； 低：工具支持部分不超过 50%，其他需人工； 中：工具支持超过 50%； 高：全部采用工具。	无：1.0 低：0.67 中：0.33 高：0.17	DDT 取值范围：0.17～1.0	

C.4 数据处理评估示例

C.4.1 计件方程法示例

受各任务包处理方式、工具、人工操作要求差异影响，不同场景下单位数据加工效率会存在明显波动。为保障工作量与费用测算结果合理可靠，模型建设实施单位可选取少量代表性样本开展数据处理效率测试，依托实测得到的单位数据加工效率，计算数据处理工作量。

示例1：某多模态融合文本标注任务包数据规模为3000条，企业基准为单条标注标准耗时0.2人时/条；1人月为174人时。

数据标注任务包工作量＝3000×0.2÷174 ≈3.45人月

项目组抽取50条样本试验标注，累计耗时约15人时，单条标注耗时为0.3人时/条，重新评估如下：

任务包工作量＝3000×0.3÷174 ≈5.17人月

C.4.2 基准工时方程法示例

示例2：某智能体应用数据处理工作量评估，该智能体应用数据处理的数据类型为文本、数据规模为6000条数据、数据质量为中，无数据复用、工具支持不超过50%；1人月为21.75人天。采用基准工时方程法评估数据处理工作量如下：

根据智能体特征选取数据处理基准工作量43人天，各影响因素取值：数据类型$DF_{2.1}$为-2、数据规模$DF_{2.2}$为0、数据质量$DF_{2.3}$为0，根据以上影响因素计算调整因子AF_2：

$AF_2 = 1.0 + (-2 + 0 + 0) \times 0.05 = 0.9$

其中，数据复用程度因子DR为1，工具化程度因子DDT为0.67。

该智能体应用数据处理总工作量为：

数据处理总工作量＝43×0.9×1×0.67/21.75 ≈1.19人月

附 录 D
（资料性）
模型建设费评估参数及示例

D.1 概述

本附录给出人工智能模型建设的典型任务包及基准工作量，各类调整因子取值示例，相关参数取值基于本文件部分编写单位的实际项目数据分析结果确定，并进行了小范围的验证。由于项目样本数量有限，相关参数取值仅供本文件读者应用参考。鉴于人工智能技术发展日新月异，实际应用中宜根据组织历史数据进行校正。此外，组织宜不断积累数据，定期测算和更新基准工作量及调整因子取值。

D.2 模型建设典型任务包基准工作量及调整因子示例

模型建设典型任务包基准工作量及调整因子示例见表D.1。

表D.1 模型建设典型任务包基准工作量及调整因子示例

序号	典型任务包	基准工作量示例（人天）	影响因素示例	影响因素说明示例	影响因素分级示例	调整因子计算及取值示例	备注
1	传统机器学习算法模型构建	14～24	模型类型 $AF_{1.1}$	简单：简单预测类； 较简单：复杂预测类、文本类自然语言处理； 一般：图像类 CV； 较复杂：视频类 CV； 复杂：多模态融合；	简单：-2 较简单：-1 一般：0 较复杂：1 复杂：3	TF_1 = 1.0 + ($AF_{1.1}$ + $AF_{1.2}$ + $DF_{1.1}$ + $DF_{1.2}$) × 0.03 取值范围：0.76～1.27	
			模型参数规模 $AF_{1.2}$	小规模：<0.1B； 中规模：0.1～1B； 大规模：>1B；	小规模：-1 中规模：0 大规模：1		
			数据规模 $DF_{1.1}$	小规模：样本量＜5000 条； 中规模：样本量在 5000～5 万条； 大规模：样本量>5 万条；	小规模：-1 中规模：0 大规模：1		
			数据质量 $DF_{1.2}$	高：数据错误率<5%； 中：数据错误率在 5%～20%（含上、下限）之间； 低：数据错误率>20%；	高：-4 中：0 低：4		
2	模型选型与搭建	13～19	基础模型参数量 $AF_{2.1}$	轻量：≤7B； 中等：7～34B； 重量：＞34B；	轻量：-1 中等：0 重量：1	TF_2 = 1.0 + ($AF_{2.1}$ + $AF_{2.2}$ + $AF_{2.3}$) × 0.05 取值范围：0.8～1.2	
			基础模型调度数量 $AF_{2.2}$	少：1 个； 一般：2～3 个； 多：>3 个；	少：-1 一般：0 多：1		
			部署隔离要求 $AF_{2.3}$	低：无隔离； 中：基础业务隔离； 高：物理隔离；	低：-2 中：0 高：2		

表D.1（续）

序号	典型任务包	基准工作量示例（人天）	影响因素示例	影响因素示例说明	影响因素分级示例	调整因子计算及取值示例	备注
3	模型微调	25～30	微调方式 $AF_{3.1}$	部分：LoRA 微调； 全量：SFT 微调；	部分：0 全量：1	$TF_3 = 1.0 + (AF_{3.1} + AF_{3.2} + DF_{3.1}) \times 0.015$ 取值范围：0.93～1.09	
			数据质量 $DF_{3.1}$	高：数据错误率<5%； 中：数据错误率在 5%～20%（含上下限）之间； 低：数据错误率>20%；	高：-4 中：0 低：4		
			参数对比组数量 $AF_{3.2}$	少：≤1 组； 中等：2～3 组； 多：≥4 组；	少：-1 中等：0 多：1		
4	知识库检索增强	19～25	知识库文档数量 $DF_{4.1}$	少：≤50 份； 中等：51～200 份； 多：>200 份；	少：-1 中等：0 多：1	$TF_4 = 1.0 + (AF_{4.1} + DF_{4.1}) \times 0.03$ 取值范围：0.88～1.12	
			知识图谱建设要求 $AF_{4.1}$	低：不建图谱； 中：简易实体关联图谱； 高：完整本体+推理图谱；	低：-3； 中：0； 高：3		
5	提示词工程	23～75	提示词总数 $AF_{5.1}$	少：≤20 条； 中等：21～100 条； 多：>100 条；	少：-1 中等：0 多：1	$TF_5 = 1.0 + (AF_{5.1} + AF_{5.2}) \times 0.26$ 取值范围：0.48～1.52	
			单条提示词规则数量 $AF_{5.2}$	少：规则≤3 条、异常≤3 个； 中等：规则 2～3 条、异常 4～10 个； 多：规则 3～5 条、异常＞10 个；	少：-1 中等：0 多：1		
6	智能体编排	20～74	独立智能体数量 $AF_{6.1}$	少：1 个； 中等：2～3 个； 多：>3 个；	少：-1 中等：0 多：1	$TF_6 = 1.0 + (AF_{6.1} + AF_{6.2} + AF_{6.3}) \times 0.09$ 取值范围：0.46～1.54	
			API/工具调用数量 $AF_{6.2}$	少：≤3 个； 中等：4～10 个； 多：>10 个；	少：-1 中等：0 多：1		
			流程分支节点数 $AF_{6.3}$	少：≤3 个； 中等：4～10 个； 多：>10 个；	少：-1 中等：0 多：1		
7	——	——	第 i 个任务包的自研比例 SR_i	不涉及：不涉及自研，全部为外包或利用已有成果物； 少量：自研部分<50%，其他为外包或利用已有成果物； 大量：自研部分≥50%，但有少量外包或利用已有成果物； 全量：全部为自研；	不涉及：0 少量：0.33 大量：0.67 全量：1.0	取值范围：0～1.0	仅用于模型建设的任务包，各任务包分别计算自研比例。

表D.1（续）

序号	典型任务包	基准工作量示例（人天）	影响因素示例	影响因素示例说明	影响因素分级示例	调整因子计算及取值示例	备注
8	——	——	第 i 个任务包的工具（或人工智能平台）支持程度 TS_i	无：无工具支持模型建设，全部人工开发； 低：工具支持部分<50%，其他需人工开发； 中：工具支持≥50%，少量需人工开发； 高：全部采用工具，开发人员仅需调用和配置相关参数。	无：1.0 低：0.67 中：0.33 高：0.17	取值范围：0.17～1.0	仅用于模型建设的任务包，各任务包分别计算工具支持程度。

模型建设全局调整因子示例见表D.2。

表D.2 模型建设全局调整因子示例

序号	调整因子分类编号	影响因素示例	影响因素示例说明	影响因素取值示例	调整因子取值示例	说明
1	系统能力等级 SF_1	人工智能业务能力等级要求 SLF	人工智能业务能力等级要求分为 5 级： 1 级（基础级）：行业及格水平； 2 级（常规级）：行业中等水平； 3 级（专业级）：行业良好水平； 4 级（优秀级）：行业优秀水平； 5 级（领先级）：行业领先水平；	1 级：1 2 级：2 3 级：3 4 级：4 5 级：5	$SF_1 = 1.0 + (SLF-1) \times 0.25$ 取值范围：1.0～2.0	$SF = SF_1 \times SF_2$
2	系统质量特性 SF_2	人工智能应用的系统质量因素要求	$SF_{2.1}$：功能正确性 低：<80%； 一般：80%～90%（含下限）； 高：90%～97%（含下限）； 极高：≥97%，	低： -4 一般：0 高： 4 极高：10	$SF_2 = 1.0 + (SF_{2.1} + SF_{2.2} + SF_{2.3} + SF_{2.4}) \times 0.05$ 取值范围：0.5～1.8	
			$SF_{2.2}$：鲁棒性 低：常见噪声与变体输入下，输出基本稳定； 中：多数非标准输入下，输出稳定，具备基础对抗防御能力；有明确要求，需做常规检测加固； 高：极端对抗输入下，输出仍稳定一致，具备主动防御与恢复能力；需系统性投入，包括专项测试、红队演练及持续监控机制；	低：-3 中： 0 高： 3		

表 D.2（续）

序号	调整因子分类编号	影响因素示例	影响因素示例说明	影响因素取值示例	调整因子取值示例	说明
2	系统质量特性SF_2	人工智能应用的系统质量因素要求	$SF_{2.3}$：性能 低：异步/批量处理级，响应时间>2s，如长文本生成，复杂代码编写补全、批量数据处理等应用场景； 中：常规服务级，响应时间在200ms～2s（端到端，短文本，响应时间的 P90 值<2s），如知识问答、智能客服、文本生成等应用场景； 高：实时交互级，响应时间≤200ms，同声传译、实时对话、实时控制等应用场景；	低：-1 中： 0 高： 1	SF_2 = 1.0 + ($SF_{2.1}$ + $SF_{2.2}$ + $SF_{2.3}$ + $SF_{2.4}$) × 0.05 取值范围：0.5～1.8	$SF = SF_1 \times SF_2$
			$SF_{2.4}$：可靠性，发生故障/幻觉的影响程度及恢复的难易程度。 低：无明示或特别要求； 中：带来一定不便或经济损失，修复较容易； 高：带来重大经济损失或有生命危害，修复难度高；	低：-2 中： 0 高： 2		
3	开发团队经验 $MRDF_1$	开发团队的人工智能应用开发、业务开发等综合经验	开发团队综合经验： 低：无经验； 中：仅业务开发经验或人工智能应用开发经验； 高：业务和人工智能应用开发均有经验；	低：1.2 中：1.0 高：0.8	取值范围：0.8～1.2	$MRDF = MRDF_1 \times MRDF_2$
4	技术与工具成熟度 $MRDF_2$	人工智能技术与应用开发工具的成熟度	技术与工具成熟度 低：任务包所使用的主要人工智能技术、工具应用均不足 6 个月； 中：任务包所使用的主要技术或工具应用在 6 个月到 1 年之间； 高：任务包所使用的主要技术与工具均广泛应用 1 年以上。	低：1.0 中：0.9 高：0.8	取值范围：0.8～1.0	

D.3 模型建设典型任务包工作量评估示例

某人工智能应用开发项目，其中模型建设涉及提示词工程任务，假设基准工作量为40人天，主要影响因素包括提示词总数、单条提示词规则数量，预计提示词大约50～60条，每条规则在3～5条，无可复用的提示词成果物，全部由团队自行开发，人工智能平台有提示词辅助支持工具。系统业务能力级别要求为3级，功能正确性要求为高，鲁棒性要求为高，性能要求为中，可靠性要求为中，团队有AI开发经验，技术与工具成熟度为高，根据以上说明评估提示词工程的工作量如下所述。

a）确定提示词工程任务包的局部调整系数如下：

1） 提示词总数$AF_{5.1}$等级为中，分值为 0，单条提示词规则数量$AF_{5.2}$等级为高，分值为 1；

2） 提示词工程的局部调整因子 $TF_{prompt} = 1 + (AF_{5.1} + AF_{5.2}) \times 0.25$ = 1+（0+1）×0.25 = 1.25；

3） 提示词工程全部由团队开发，无可复用成果物，故自研比例为全量，SR_{prompt}取值为 1.0；

4） 人工智能平台有提示词有辅助支持工具，工具支持度为低， TS_{prompt}取值为 0.67。

b）确定全局调整系数如下：

1） 业务能力等级要求SF_1等级为 3，$SF_1 = 1 + (3\text{-}1) \times 0.25$ = 1.5；

2） 质量因素中，功能正确性要求为高，分值为 4，鲁棒性要求为高，分值为 3，性能要求为中，分值为 0，可靠性要求为中，分值为 0；

3） 质量因素调整因子 $SF_2 = 1 + (SF_{2.1} + SF_{2.2} + SF_{2.3} + SF_{2.4}) \times 0.05$ = 1+(4+3+0+0) × 0.05 =1.35；

4） 团队有 AI 应用开发经验，调整因子$MRDF_1$取值为 1.0；

5） 技术与工具成熟度高，调整因子$MRDF_2$取值为 0.8。

c）计算提示词工程的工作量

$TE_{prompt} = TBE_{prompt} \times TF_{prompt} \times SR_{prompt} \times TS_{prompt} \times SF_1 \times SF_2 \times MRDF_1 \times MRDF_2$

= 40 × 1.25 × 1.0 × 0.67 × 1.5 × 1.35 × 1.0 × 0.8

≈ 54.3 人天。

附 录 E
（资料性）
项目造价评估示例

E.1 人工智能应用开发项目信息

E.1.1 项目背景及目标

位于北京的某电子商务企业为顺应信息化、智能化发展趋势，解决用户办理业务的痛点，基于客户服务的要求，启动建设智能问答系统开发项目。该项目于 2022 年立项，项目需求为：

a） 新业务功能开发，在现有业务功能基础上新开发智能问答功能；
b） 知识库建设，对现有客户服务相关的业务数据进行处理，以满足知识问答的要求；
c） 接入基础大模型，在此基础上构建知识问答智能体，实现智能问答服务；
d） 应用集成，对接客户服务系统，以支持用户与人工座席进行对话，全面满足客户服务需求；
e） 开发用于数据处理的通用工具，提高数据处理效率和准确性，减轻数据处理人员的工作量。

E.1.2 人工智能应用开发需求

E.1.2.1 主体功能开发需求

E.1.2.1.1 智能问答

智能问答界面包括如下基础问答交互、问答辅助、转人工等功能。

a） 基础问答交互功能：
   1） 文本智能问答：支持用户输入自然语言文字发起业务咨询，系统依托知识库识别用户业务诉求，快速输出标准化、通俗化业务解答，覆盖客户日常业务咨询痛点；
   2） 多轮对话交互：具备上下文记忆能力，可承接用户连续追问，理解对话语境，实现连贯的多轮问答交互。
b） 问答辅助功能：
   1） 答案反馈评价：用户可对问答结果进行满意度评价，包括满意、一般、不满意，同时支持自定义文字留言反馈问题，为知识库优化、算法迭代提供依据；
   2） 对话记录管理：自动留存用户全部咨询对话记录，支持按时间、业务关键词检索，可单条删除或批量清空。
c） 转人工：
   1） 一键转人工服务：对接客户服务系统，用户在对话界面可一键触发转人工操作，并跳转至客户服务系统的人工座席页面，对话由客服人员接管。无需跨系统跳转、无需重新登录，实现一体化用户服务集成；
   2） 人机对话无缝接续：文本智能问答转人工后，人工座席可完整调取用户全程对话记录、咨询问题、业务疑问、操作场景等信息，用户无需重复复述问题，实现人工专业服务的无缝衔接。

E.1.2.1.2 对话管理

对话管理包括欢迎语、推荐问题管理、兜底话术配置等功能，具体需求如下：

a） 欢迎语管理

管理员可以通过页面增删改查欢迎语的文案内容，并配置个性化规则，如按不同时段、不同渠道展

示不同版本。

b） 推荐问题管理

管理员可通过页面增删改查固定推荐问题列表，也可手动调整系统自动统计的“热点问题”的展示顺序。

c） 兜底话术配置

当系统无法回答时，管理员可以通过页面编辑默认的拟人化拒绝话术，并配置转人工的引导入口链接。

E.1.2.1.3 知识库管理

知识库管理包括创建知识库、知识导入、在线测试、段落管理、知识更新、知识图谱手动关联、知识审核与发布、版本回滚等功能，具体需求如下：

a） 创建知识库

管理员通过可视化界面新建知识库，配置基础信息（名称、描述、所属分类）；设定文档段落切分策略；选择索引方式（向量索引、全文索引或混合索引），并指定模型，完成后提交创建。

b） 知识导入

将已梳理的文档类知识（Word/PDF/Excel/HTML 等）及问答对知识，通过界面批量或单文件导入，异常文件提供错误日志下载。系统自动执行格式解析与清洗，文本规范化，内容分块，向量化嵌入，调用模型将文本片段转化为稠密向量，索引存储并将向量与原始文本存入向量数据库，同时构建倒排索引供全文检索。

c） 在线测试

知识导入完成后，提供问答测试面板，管理员输入测试问题，可实时查看检索召回的分段列表及相似度得分、重排序后的结果，以及大模型最终生成答案。通过可视化比对，验证分段质量与检索效果，辅助后续调优决策。

d） 段落管理

提供知识段落列表视图，管理员可按文档、标签、状态等维度筛选查看。支持对已切分的段落进行编辑、合并、拆分、禁用等操作，优化语义边界。

e） 知识更新

支持修改知识库基础信息、切分策略、索引方式；编辑已有知识条目内容，操作流程与创建一致。支持删除知识库或单条知识，需弹窗二次确认；删除后，所有关联该知识库的问答应用将无法检索该库内容，确保数据一致性。提供多条件组合筛选（关键词、知识库名称、文档类型、创建时间、状态等），以调阅知识库全部内容列表，并提供导出功能。

f） 图谱关系手动关联

管理员可通过图形化界面，以拖拽方式手动建立知识节点间的语义关系，支持关系类型自定义（因果、时序、包含、关联），补全自动抽取的不足，增强复杂推理能力。

g） 知识审核与发布

对新增或修改的知识条目，实行审核发布机制。管理员提交知识后进入待审核状态，审核员在页面逐条确认内容准确性、合规性，审核通过后，管理员方可执行发布操作，正式生效并同步更新检索索引。审核不通过则退回修改并附注原因。

h） 版本回滚

系统自动记录知识库每次发布的历史版本（含文档内容、问答对、图谱关系、切分配置等）。当新版本知识导致问答效果下降时，管理员可通过版本列表一键回滚至任一历史版本，保障业务连续性。

E.1.2.1.4 检索调优

支持文档分段设定规则、知识点人工干预操作、自动与手动去重等功能，具体需求如下：

a) 文档分段规则设定：管理员需在页面上手动切换“自动分段”与“自定义分段”，若选择自定义，也可设置分段标识符、最大长度和重叠字数等具体数值；

b) 知识点管理（知识干预操作）：支持手工新增、编辑、复制知识点，通过人工干预提升关键知识的命中效率，以人工方式直接干预和补充模型可能漏召回的短板内容；

c) 去重任务触发：系统可自动去重，管理员也可在页面点击“执行去重”按钮，对系统推荐合并的重复项进行人工确认并合并。

E.1.2.1.5 索引管理

向量检索调优，索引重建触发：当知识库新增或大量修改内容后，管理员需在后台页面手动点击“重建/更新索引”按钮，确保新增内容能被及时检索到。

E.1.2.1.6 提示词管理

提示词编写与修改：管理员需在后台专门的提示词管理页面，直接编写或修改问答通用指令、推荐指令、改写指令的模板内容，保存后实时生效。若有多套提示词模板，管理员可启用不同版本，并进行 AB 对比测试。

E.1.2.1.7 应用集成需求

应用集成需求如下：

a） 与客服系统对接

对于用户指定转人工或系统判定需人工介入的问题，智能体需将用户问答记录、会话上下文等数据推送至客服系统，并支持前端页面跳转至客服系统人工服务入口，由座席人员接管对话。

b） 基础大模型接入

智能体通过标准 API 调用公司技术中台的基础大模型，作为推理生成能力的底层支撑。集成包含认证鉴权、请求响应格式规范、超时与重试策略等。当模型服务不可用或响应异常时，自动执行重试或切换至备用模型，降级后返回兜底回复，保障服务稳定性。

E.1.2.1.8 人工智能相关功能组件开发

为了减轻数据处理人员的工作量，提高数据处理效率和准确性，专门开发了面向数据处理的通用工具，具体需求如下：

a） 数据导入

用户可导入本地单个或多个文件等待处理数据，工具自动识别文件格式（CSV、Excel、JSON、XML、TXT、PDF、图片、音频等），在导入过程中，用户可指定一个或多个关键字段，工具基于本地缓存（索引）检测重复数据，自动跳过已存在的记录，防止重复入库。界面默认展示前 10 个文件，用户可点击其中任意一个文件的名称，查看文件详情。

b） 文本专项清洗

针对文本清洗，用户可选择文本列，可分别执行重复句子去重、无效话术过滤、错别字智能修正、敏感词替换（用户自定义敏感词及替换词）功能操作，文本清洗执行完毕后，展示文本清洗前后的比对结果。

c） 自定义清洗规则

允许用户通过正则表达式或 Python 脚本编写自定义清洗逻辑，并保存为个人规则，后续可反复使用。用户可修改、删除或查看自定义的规则。

d） 批量格式转换

针对不同格式的数据，支持大规模批量文件格式转换，提高转换效率。用户可选择文件夹/文件列表、源格式和目标格式后，即可执行批量转换。格式转换完成后的数据可批量导出。

E.1.2.2 数据处理需求

E.1.2.2.1 主题界定与知识划分

支持业务域界定，输出知识体系文档等功能，具体需求如下：

a） 业务域界定

明确目标问答场景与服务范围（如产品咨询、售后服务等），定义知识边界与覆盖粒度，确保后续采集与梳理聚焦在业务范围内。

b） 知识体系框架输出

输出知识体系框架文档，包含核心业务概念、实体类型（事项、材料、条件等）、属性定义及典型问题分类体系，作为后续数据采集、标注以及知识库创建、推荐问题配置的依据。

E.1.2.2.2 数据采集与源评估

支持多数据源采集、数据源评估等功能，具体需求如下：

a） 多源数据获取

从内部系统、业务部门、公开渠道等采集商品数据、订单与物流数据、用户数据、市场动态数据、历史工单等各类资料。

b） 数据源评估

评估各数据源的时效性（是否最新版本）、权威性（来源是否可靠）、完整性（内容是否齐全）及合规性（是否可被用于知识库构建），剔除过时、重复或无效的资料，形成待加工数据源清单。

E.1.2.2.3 信息梳理与结构化抽象

进行信息梳理与结构化抽象，支持对资料进行深度阅读与提炼、多维度分类、知识骨架抽象等功能，具体需求如下：

a） 深度阅读与提炼

通读采集到的资料，提炼其中的关键概念、业务流程、办理规则、材料要求、条件约束等信息。

b） 多维度分类

提炼出的信息按业务层级（业务线/经营域→服务场景→具体意图）、业务类别（交易执行类、信息查询类、规则告知类、投诉维权类）、文档类型（业务规则/SOP、自助操作指南、历史工单/案例库）、适用场景（售前导购、订单履约中、售后保障、大促营销活动）等维度进行分类整理。

c） 知识骨架抽象

采用思维导图、分类表格、知识图谱草图等工具，将分类后的信息抽象为知识骨架，形成知识体系的初步结构。

E.1.2.2.4 知识建模与关系构建

支持知识建模，构建知识图谱，具体需求如下：

a） 实体关系设计

设计知识图谱或语义网络的 Schema，明确实体间的业务关系类型（因果、时序、包含、关联等），如“事项→所需材料”、“事项→办理部门”、“前置条件→事项”等。

b） 知识拓扑输出

定义主干知识与分支细节的层次结构，形成可推理的知识拓扑，作为后续知识录入、图谱关系手动关联和检索策略优化的蓝图。

E.1.2.2.5 数据加工与标注

对采集的数据进行清洗、转换、标注和预处理等数据处理，具体需求如下：

a） 数据清洗

对采集的原始数据进行去重、纠错、格式统一、缺失值补全等处理，确保数据准确完整。

b） 内容转换

统一业务术语（如“身份证”统一为“居民身份证”）、标准化日期与单位格式、进行文本归一化处理。

c） 内容标注

按照标注规范对知识内容进行标注，包括实体标注（商品名称、品牌/店铺名称、订单号、SKU 属性、物流公司等）、意图分类（售前咨询类、售后处理类、物流查询类、投诉维权类、营销活动类等）、属性标引（发货时效、退换货周期、保修期限、满减规则、适用尺码/颜色等）、关系标注（下单与发货间的时序关系、退货与退款间的因果关系、差评与售后单间的关联关系等）。

d） 预处理

对文本进行合理分句、业务模板化处理、问题同义扩展（如“怎么办”“如何办理”等不同问法）、多轮对话上下文切片设计，为后续向量化与检索提供良好输入。

E.1.2.2.6 知识库交付与验收

产出结构化业务知识库，包含标准问答对、知识三元组、业务规则集等成果物。设定质量验收指标（知识覆盖率、准确率、一致性、冗余度），通过抽样验证与专家评审双重确认，确保知识库满足模型训练与推理的质量要求后正式交付。

E.1.2.3 知识问答智能体构建

E.1.2.3.1 对话交互与智能体推理

知识问答智能体可进行对话交互，自动识别用户意图并给出问题答案，支持记忆与多轮对话，支持安全与异常处理，具体需求如下：

a） 会话初始化与引导

系统根据渠道、时段、用户属性等维度动态渲染个性化欢迎语并推送，同步下发预设固定问题与高频热点组成的推荐问题列表。

b） 智能体推理与执行

系统自动识别用户意图并判断处理方式：常规咨询触发 RAG 检索增强生成（含查询改写、多路召回、重排序、上下文融合及推理生成）；复杂问题触发智能体多步推理，支持任务拆解、执行路径规划、工具调用（知识库检索、业务系统查询等）及自我反思，根据中间结果动态调整规划。简单事项采用检索结果直接抽取式回答，复杂事项经多步推理与多片段融合后输出结构化答案。

c） 记忆与多轮对话管理

构建短期记忆（当前会话的对话历史、已执行步骤）与长期记忆（用户历史交互中的关键信息与偏好），支持跨会话复用。结合记忆机制实现上下文理解、指代消解与省略补全，支撑多轮深度交互。

d） 安全与异常处理

建立双向审核机制，对用户输入与模型输出同步进行涉黄、涉暴、涉恐、违法等敏感内容检测与过滤。当问题超出业务范围、检索置信度低于阈值或任务执行失败时，自动柔性降级，采用兜底话术并引导转人工。

E.1.2.3.2 知识库工程构建

构建客服知识库，支持知识采集与梳理、知识录入与索引、质量审核与版本关联、增量更新与图谱补全等，具体需求如下：

a） 知识采集与梳理

多源获取商品数据、订单与物流数据、用户数据、市场动态数据、历史工单等文档资料，完成内容提取与格式归一化。基于 NLP 技术进行实体识别与关系抽取，梳理业务关联，构建知识图谱基础结构。

b） 知识录入与索引

将加工后的知识导入系统，形成覆盖文档库、问答对库、知识图谱的结构化体系。调用模型将文本片段转化为稠密向量，同步构建向量索引、倒排索引及图谱索引。

c） 质量审核与版本管理

内置质量校验机制，对知识的完整性、准确性进行自动化检测。新增或修改的知识提交后进入待审核状态，审核通过后发布生效并同步更新索引。系统自动记录每次发布的历史版本，支持一键回滚至任一历史版本。

d） 增量更新与图谱补全

建立增量更新机制，自动识别知识变更并同步索引，确保知识时效性。图谱关系除自动抽取外，支持外部输入语义关系（因果、时序、包含、关联），补全自动抽取不足，增强复杂推理能力。

E.1.2.3.3 模型与智能体评估调优

支持按策略进行文档分段，检索策略切换，检索参数配置、自动向量检索调优等，具体需求如下：

a） 文档分段策略

支持自动智能分段与自定义规则分段两种模式，根据预设参数（分句标识符、最大长度、重叠占比）执行分块。

b） 检索调优

支持全文检索、语义检索、混合检索三种策略切换。引入重排序机制，对召回内容按相关性重新排序，打破原始相似度得分顺序。支持将外部输入的标准知识点纳入检索范围，提升关键知识召回命中率。

c） 检索参数配置

支持按预设 Top-K 值控制召回数量，权衡回答全面性与推理效率。按设定的相似度阈值过滤检索结果，仅召回高于阈值的内容，平衡查准率与查全率。对语义相同或答案重复的问答对自动检测并去重。

d） 向量检索专项调优

支持按配置选择检索策略（关键词/语义/混合），按预设 Top-K 值执行向量召回，按相似度阈值过滤结果。知识库更新后自动对新增内容建立向量索引，核心参数变更时支持全量索引重建。

E.1.2.3.4 提示词与智能体配置

支持智能体行为规划策略，工具描述提示词、问答与推荐提示词模板设置与版本切换等功能，具体需求如下：

a） 智能体行为与规划提示词

基于预设的智能体角色定位、行为准则、回答风格及拒答口径，控制智能体整体行为与输出风格。基于规划策略模板，引导智能体进行任务拆解、推理步骤规划与反思，提升复杂问题处理能力。

b） 工具描述提示词

基于预设的工具功能说明、参数格式与使用场景描述，帮助智能体理解各工具的用途与调用方式，实现准确的工具选择与参数填充。

c） 问答与推荐提示词

基于预设的回答风格与格式规范，控制最终答案输出形式。根据对话上下文轮数动态调整推荐策略，基于改写模板自动执行历史对话参考轮数调整、指代消解与省略补全。

d） 模板版本管理

支持多套提示词模板并存，可通过配置切换启用不同版本。

E.1.2.4 智能体部署与上线

E.1.2.4.1 环境准备

根据业务需求与预估并发量，规划推理服务及配套中间件（向量数据库、检索引擎、缓存服务等）的资源配置，并配置弹性扩缩容能力，完成网络策略、身份认证、传输加密等安全设置。通过容器化封装确保开发、测试与生产环境一致。

E.1.2.4.2 智能体服务部署

对基础模型进行推理优化，部署智能体服务并封装为标准 API 接口供调用。同步将已构建的知识索引部署至生产环境，确保与智能体版本匹配。统一管理模型参数、工具配置、记忆存储等运行参数，生产环境敏感信息加密存储。

E.1.2.4.3 上线验证

在预发布环境完成功能与性能测试，覆盖意图识别、工具调用准确性、多轮对话、审核拒答等核心流程。采用灰度发布，先对少量用户开放，确认服务稳定后逐步扩大至全量用户。支持不同版本间对比测试，确定最优方案后正式上线。

E.1.2.4.4 监控与运维

持续监控系统运行状态、资源使用及服务响应性能。跟踪问答准确率、工具调用成功率、用户反馈，记录全链路日志并配置告警规则。根据负载自动扩缩容，保留近期稳定版本支持快速回滚，知识库更新自动同步至检索服务。

E.1.2.4.5 上线验收

逐项确认硬件配置、智能体服务、知识索引、核心问答与工具调用能力、性能指标、安全策略、监控告警及回滚能力均满足要求，验收通过后正式交付。

E.2 评估过程

E.2.1 主体功能开发费评估

E.2.1.1 主体功能开发规模

根据项目的主体功能开发需求，按 GB/T 36964-2018 采用功能点方法进行人工智能应用软件规模评估。软件主体功能规模评估明细见下表 E.1。

表 E.1 软件主体功能规模评估明细表

编号	子系统	一级模块	二级模块	功能点计数项名称	类别	UFP	重用程度	修改类型	US	备注
1	智能问答系统	智能问答界面	基础问答交互功能	文本问答	EO	5	低	新增	5.00	
2			问答辅助功能	满意度评价	EI	4	低	新增	4.00	
3				满意度信息	ILF	10	低	新增	10.00	
4				留言反馈	EI	4	低	新增	4.00	
5				对话记录查询	EQ	4	低	新增	4.00	
6				对话记录删除	EI	4	低	新增	4.00	
7				对话信息	ILF	10	低	新增	10.00	
8			转人工	一键转人工	EI	4	低	新增	4.00	
9		对话管理	欢迎语管理	欢迎语信息	ILF	10	低	新增	10.00	
10				欢迎语新增	EI	4	低	新增	4.00	
11				欢迎语删除	EI	4	低	新增	4.00	
12				欢迎语编辑	EI	4	低	新增	4.00	
13				欢迎语查询	EQ	4	低	新增	4.00	
14				欢迎语规则配置	EI	4	低	新增	4.00	
15				欢迎语规则查询	EI	4	低	新增	4.00	
16				欢迎语规则	ILF	10	低	新增	10.00	
17			推荐问题管理	推荐问题信息	ILF	10	低	新增	10.00	
18				推荐问题新增	EI	4	低	新增	4.00	
19				推荐问题编辑	EI	4	低	新增	4.00	
20				推荐问题查询	EQ	4	低	新增	4.00	
21				推荐问题删除	EI	4	低	新增	4.00	
22				热点问题展示	EQ	4	低	新增		
23				热点问题展示顺序调整	EI	4	低	新增	4.00	
24			兜底话术配置	兜底话术配置	EI	4	低	新增	4.00	
25		知识库管理	创建知识库	知识库文件	ILF	10	低	新增	10.00	
26				知识库创建	EI	4	低	新增	4.00	
27			知识导入	知识导入	EI	4	高	新增	1.33	

表 E.1（续）

编号	子系统	一级模块	二级模块	功能点计数项名称	类别	UFP	重用程度	修改类型	US	备注
28				错误日志下载	EQ	4	低	新增	4.00	
29			在线测试	问答测试	EO	5	低	新增	5.00	
30				测试结果信息	ILF	10	低	新增	10.00	
31			段落管理	知识段落信息	ILF	10	低	新增	10.00	
32				知识段落查看	EQ	4	低	新增	4.00	
33				知识段落编辑	EI	4	低	新增	4.00	
34				知识段落合并	EI	4	低	新增	4.00	
35				知识段落拆分	EI	4	低	新增	4.00	
36				知识段落禁用	EI	4	低	新增	4.00	
37			知识更新	知识库修改	EI	4	低	新增	4.00	
38				知识删除	EI	4	低	新增	4.00	
39				知识库列表查询	EQ	4	低	新增	4.00	
40				知识导出	EQ	4	高	新增	1.33	
41			图谱关系手动关联	知识图谱信息	ILF	10	低	新增	10.00	
42				图谱关系手动关联	EI	4	低	新增	4.00	
43				关系类型自定义	EI	4	低	新增	4.00	
44			知识审核与发布	知识审核	EI	4	低	新增	4.00	
45				知识条目发布	EI	4	低	新增	4.00	
46			版本回滚	版本回滚	EI	4	低	新增	4.00	
47		检索调优		文档分段规则	ILF	10	低	新增	10.00	
48				文档分段规则设置	EI	4	低	新增	4.00	
49				文档分段规则查询	EQ	4	低	新增	4.00	
50				文档分段规则切换	EI	4	低	新增	4.00	
51				知识点新增	EI	4	低	新增	4.00	
52				知识点展示	EI	4	低	新增	4.00	
53				知识点编辑	EI	4	低	新增	4.00	
54				知识点复制	EI	4	低	新增	4.00	
55				知识点信息	ILF	10	低	新增	10.00	
56				知识去重	EI	4	低	新增	4.00	
57			索引管理	索引重建	EI	4	低	新增	4.00	
58			提示词管理	提示词新建	EI	4	低	新增	4.00	
59				提示词修改	EI	4	低	新增	4.00	
60				提示词展示	EI	4	低	新增	4.00	
61				提示词版本启用	EI	4	低	新增	4.00	

表 E.1（续）

编号	子系统	一级模块	二级模块	功能点计数项名称	类别	UFP	重用程度	修改类型	US	备注
62				提示词比对测试	EO	5	低	新增	5.00	
63				提示词信息	ILF	10	低	新增	10.00	
64		应用集成需求	与客服系统对接	对话记录推送						非独立业务功能
65			基础大模型接入	对话信息						功能重复，已计数
66		人工智能相关功能组件开发	数据导入	数据导入	EI	4	低	新增	4.00	
67				文件列表展示	EQ	4	低	新增	4.00	
68				文件详情查看	EQ	4	低	新增	4.00	
69			文本专项清洗	重复句子去重	EI	4	低	新增	4.00	
70				无效话术过滤	EI	4	低	新增	4.00	
71				错别字智能修正	EI	4	低	新增	4.00	
72				敏感词替换	EI	4	低	新增	4.00	
73				文本清洗结果展示	EQ	4	低	新增	4.00	
74			自定义清洗规则	清洗规则	ILF	10	低	新增	10.00	
75				清洗规则新增	EI	4	低	新增	4.00	
76				清洗规则修改	EI	4	低	新增	4.00	
77				清洗规则查询	EQ	4	低	新增	4.00	
78				清洗规则删除	EI	4	低	新增	4.00	
79			批量格式转换	文件批量格式转换	EI	4	低	新增	4.00	
80				文件批量导出	EQ	4	低	新增	4.00	
合计						393			383.67	

E.2.1.2 主体功能开发费

根据项目的主体功能开发需求及其规模评估结果，计算主体功能开发工作量及费用，主体功能开发费评估结果见下表 E.2。

表 E.2 主体功能开发费

主体功能开发费				
调整因子类别			取值	取值说明
1.计算未调整功能点数 US（功能点） US =UFP*重用程度调整系数*修改类型调整系数，其中： UFP=10*ILF+7*EIF+4*EI+5*EO+4*EQ（或 35*ILF+15*EIF）			383.67	
2.设定规模变更调整因子 CF			1.21	立项阶段
3.计算调整后的功能点 S（功能点） S=US*CF			464.24	
4.基准数据（生产率） PDR （功能点耗时率）		下限（人时/功能点）	5.61	全行业基准数据中位数(CSBMK®-202210)
		中值（人时/功能点）	7.01	
		上限（人时/功能点）	8.41	
5.计算未调整的工作量（人天） UE=PDR*S/8		下限	325.43	
		中值	406.79	
		上限	488.14	
6.设定调整因子	软件因素调整因子（SWF）	应用类型调整因子	1.00	业务处理
		非功能性特征调整因子	1.00	无特别要求
		完整性级别	1.00	没有明确的完整性级别或等级为 C/D
	开发因素调整因子（RDF）	开发语言调整因子	1.00	JAVA、C++、C#及其他同级别语言/平台
		开发团队背景调整因子	0.80	为本行业开发过类似的项目
7.计算调整后的工作量 AE_1（人天） AE_1=UE*SWF*RDF; 其中：SWF= AT*QR，RDF=SL*DT		下限	260.34	
		中值	325.43	
		上限	390.52	
8.AE_1 对应的平均人力成本费率 F_1（万元/人月） （人月单价）			3.2343	2022 年北京地区人月单价，不包含直接非人力成本（CSBMK@-202210）
9.测算软件主体功能开发费 P_1（万元） $P_1=AE_1/21.75*F_1$ （不包含直接非人力成本 DNC）		下限	38.71	
		中值	48.39	
		上限	58.07	

根据上表评估结果，主体功能开发费中值为 48.39 万元。

E.2.2 数据处理费评估

根据项目的数据处理需求，本项目涉及的模型类型为智能体，采用基准工时方程法评估数据处理费，任务难度为中等。本项目的人工智能应用开发的综合人力费率为3.2343万元/人月，数据处理人力费率的岗位系数取中值0.95，基准工作量参考附录C取中值43人天。

任务难度为中等，参考附录C.3，调整因子$DF_{2.1}$、$DF_{2.2}$、$DF_{2.3}$均取中值。

数据处理工作量调整因子AF_2的计算结果如下：

$AF_2=1.0+(DF_{2.1}+DF_{2.2}+DF_{2.3})\times0.05=1.0+(0+0+0)\times0.05=1.0$

无数据复用，数据复用程度因子DR取值为1.0。

项目团队人工处理的数据量小于50%，数据处理工具化程度因子DDT取值为0.33。

数据处理费计算如下：

数据处理费用=43×1.0×1.0×0.33÷21.75×0.95×3.2343≈2.00万元

说明：1人月等于21.75人天。

根据以上评估过程，数据处理费评估结果中值为 2.00 万元。

E.2.3 模型建设费评估

E.2.3.1 概述

根据项目知识问答智能体构建需求，本项目涉及的模型类型为智能体，采用方程法评估智能应用模型建设费，任务难度为中等，模型建设所涉及的任务包为对话交互式智能体模型搭建、检索增强知识库构建、模型微调、提示词工程。本项目的人工智能应用开发的综合人力费率为 3.2343 万元/人月。模型建设人力费率的岗位系数取中值 1.45。

E.2.3.2 评估各任务包的工作量

E.2.3.2.1 对话交互式智能体模型搭建任务包

任务难度为中等，参考附录 D，相关参数取值说明如下：

a) 任务包的基准工作量（TBE）=16 人天；
b) 已有部分可复用的成果物，任务包的自研比例为大量，SR 取 0.67；
c) 平台无工具支持，工具支持度 TS 取 1.0；
d) 业务能力级别要求SF_1取 1.5；
e) 质量因素调整因子SF_2取 1.15；
f) 开发团队经验$MRDF_1$取 1.0；
g) 技术与工具成熟度$MRDF_2$取 0.9；
h) 调整因子$AF_{2.1}$、$AF_{2.2}$、$AF_{2.3}$均取中值。

局部调整因子TF_2的计算结果如下：

$TF_2 = 1.0 + (AF_{2.1} + AF_{2.2} + AF_{2.3}) \times 0.05=1.0+(0+0+0)\times0.05=1.0$

根据 6.3 的评估方法，模型搭建任务包的工作量计算如下：

模型搭建任务包的工作量=16×1.0×0.67×1.0×1.5×1.15×1.0×0.9 ≈16.64 人天

E.2.3.2.2 检索增强知识库构建任务包

任务难度为中等，参考附录 D，相关参数取值说明如下：

a) 任务包的基准工作量（TBE）=22 人天；

b） 平台无可复用的成果物，全部为自研，SR 取 1.0；
c） 平台有部分知识库支持工具，工具支持度 TS 取 0.67；
d） 业务能力级别要求SF_1取 1.5；
e） 质量因素调整因子SF_2取 1.15；
f） 开发团队经验$MRDF_1$取 1.0；
g） 技术与工具成熟度$MRDF_2$取 0.9；
h） 调整因子$AF_{4.1}$、$DF_{4.1}$均取中值。
局部调整因子TF_4的计算结果如下：
$TF_4 = 1.0 + (AF_{4.1} + DF_{4.1}) \times 0.03 = 1.0+(0+0)\times 0.03 = 1.0$
根据 6.3 节的评估方法，检索增强知识库构建任务包的工作量计算结果如下：
检索增强知识库构建任务包的工作量=22×1.0×1.0×0.67×1.5×1.15×1.0×0.9≈22.88 人天

E.2.3.2.3 模型微调任务包

任务难度为中等，参考附录 D，相关参数取值说明如下：
a） 任务包的基准工作量（TBE）=27.5 人天；
b） 平台无可复用的成果物，全部为自研，SR 取 1.0；
c） 平台有部分模型微调支持工具，工具支持度 TS 取 0.67；
d） 业务能力级别要求SF_1取 1.5；
e） 质量因素调整因子SF_2取 1.15；
f） 开发团队经验$MRDF_1$取 1.0；
g） 技术与工具成熟度$MRDF_2$取 0.9；
h） 调整因子$AF_{3.1}$、$AF_{3.2}$、$DF_{3.1}$均取中值。
局部调整因子TF_3的计算结果如下：
$TF_3 = 1.0 + (AF_{3.1} + AF_{3.2} + DF_{3.1}) \times 0.015=1.0+(0+0+0)\times 0.015=1.0$
根据 6.3.1 的评估方法，模型微调任务包的工作量计算结果如下：
模型微调任务包的工作量=27.5×1.0×1.0×0.67×1.5×1.15×1.0×0.9≈28.60 人天

E.2.3.2.4 提示词工程任务包

任务难度为中等，参考附录 D，相关参数取值说明如下：
a） 任务包的基准工作量（TBE）=49 人天；
b） 平台无可复用的成果物，全部为自研，SR 取 1.0；
c） 平台有部分知识库支持工具，工具支持度 TS 取 0.67；
d） 业务能力级别要求SF_1取 1.5；
e） 质量因素调整因子SF_2取 1.15；
f） 开发团队经验$MRDF_1$取 1.0；
g） 技术与工具成熟度$MRDF_2$取 0.9；
h） 调整因子$AF_{5.1}$、$AF_{5.2}$均取中值。
局部调整因子TF_5的计算结果如下：
$TF_5 = 1.0 + (AF_{5.1} + AF_{5.2}) \times 0.26=1.0+(0+0)\times 0.26=1.0$
根据 6.3.1 的评估方法，提示词工程任务包的工作量计算结果如下：
提示词工程任务包的工作量=49×1.0×1.0×0.67×1.5×1.15×1.0×0.9≈50.97 人天

E.2.3.3 评估模型建设的总工作量

知识问答智能体构建的总工作量（TME）=16.64+22.88+28.60+50.97=119.09 人天

E.2.3.4 评估模型建设费

模型建设按附录 A 岗位系数取值为 1.45
模型建设费=119.09÷21.75×3.2343×1.45≈25.68 万元
说明：

a) 智能应用模型建设所需基础大模型是从公司技术中台接入的，因此不涉及传统机器学习算法模型任务包及其调整因子TF_1；

b) 本项目中的知识问答智能体构建过程无需智能体编排，因此不涉及智能体编排任务包及其调整因子TF_6。

根据以上评估过程，模型建设费的评估结果中值为 25.68 万元。

E.2.4 直接非人力成本评估

本项目需求说明中无直接非人力成本的相关说明，暂不考虑直接非人力成本。

E.2.5 人工智能应用开发项目造价评估结果汇总

汇总上述主体功能开发费、数据处理费、模型建设费、直接非人力成本评估结果，得到本项目的人工智能应用开发总成本，评估结果见表 E.3。

表 E.3 人工智能应用开发项目造价评估结果

序号	费用名称	金额（单位：万元）—下限	金额（单位：万元）—中值	金额（单位：万元）—上限	备注
一	**综合人力成本（含直接人力成本和间接成本）**				
1.1	主体功能开发费	38.71	48.39	58.07	
1.2	数据处理费	1.60	2.00	2.40	
1.3	模型建设费	20.54	25.68	30.82	
小计（一 = 1.1 + 1.2 + 1.3）		60.85	76.07	91.29	
二	**直接非人力成本**				
2.1	软硬件购置费	0.00	0.00	0.00	
2.2	资源租赁费	0.00	0.00	0.00	
2.3	人工智能环境集成费	0.00	0.00	0.00	
2.4	数据资源与数据服务费	0.00	0.00	0.00	
2.5	模型服务费	0.00	0.00	0.00	
2.6	其他直接非人力成本	0.00	0.00	0.00	
小计（二 = 2.1 + 2.2 + 2.3 + 2.4 + 2.5 + 2.6）		0.00	0.00	0.00	
总计（= 一 + 二）		60.85	76.07	91.29	

综上，本项目人工智能应用开发项目造价的合理范围介于 60.85 万元到 91.29 万元之间，76.07 万元是该项目的合理价格。

参 考 文 献

[1] GB/T 5271.6-2000 信息技术 词汇 第6部分：数据的准备与处理
[2] GB/T 37988-2019 信息安全技术 数据安全能力成熟度模型
[3] GB/T 41867-2022 信息技术 人工智能 术语
[4] GB/T 42588-2023 系统与软件工程 功能规模测量 NESMA方法
[5] GB/T 45288.1-2025 人工智能 大模型 第1部分：通用要求
[6] 国家职业技术技能标准（职业编码：2-02-10-09） 人工智能工程技术人员（2021年版）
[7] 国家职业技能标准（职业编码： 4-04-05-01） 计算机程序设计员（2022年版）
[8] 国家职业技能标准（职业编码： 4-04-05-05） 人工智能训练师（2021年版）
[9] 国家职业编码：4-04-05-13 生成式人工智能系统应用员（含智能体开发员）
[10] ISO/IEC 22989-2022 Information technology — Artificial intelligence—Artificial intelligence concepts and terminology
[11] ISO/IEC 25059-2023 Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems

BSCE官网展示专用

BSCE官网展示专用