ICS 35.080
CCS 77

T/CCUA

团　　体　　标　　准

T/CCUA XXX—XXXX
T/BSCEA XXX—XXXX

系统与软件工程 软件开发造价评估平台 能力要求

Systems and software engineering -Software develpment cost evaluation platform - Capability requirements

（征求意见稿）

（本稿完成日期：2025-08-24）

XXXX－XX－XX 发布　　XXXX－XX－XX 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 III
引言 IV
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 缩略语 2
5 能力模型框架 2
6 平台能力等级 3
6.1 L1 级 3
6.2 L2 级 3
6.3 L3 级 3
6.4 L4 级 3
7 L1 级能力要求 4
7.1 核心功能能力域 4
7.2 管理与支撑功能能力域 4
7.3 非功能能力域 5
8 L2 级能力要求 5
8.1 核心功能能力域 5
8.2 管理与支撑功能能力域 6
8.3 非功能能力域 6
9 L3 级能力要求 显著降低评估人员工作量 7
9.1 核心功能能力域 7
9.2 管理与支撑功能能力域 8
9.3 非功能能力域 9
10 L4 级能力要求 人调整能达到 9
10.1 核心功能能力域 9
10.2 管理与支撑功能能力域 10
10.3 非功能能力域 11
附录 A （资料性） 平台能力等级主要差异 12
附录 B （资料性） 典型应用场景 15
B.1 L1 级应用场景 15
B.2 L2 级应用场景 15
B.3 L3 级应用场景 15
B.4 L4 级应用场景 15

参考文献 17

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：

本文件主要起草人：

引　言

2018年《政府投资条例》(国务院令第712号)、《中央预算内直接投资项目管理办法》(国家发展和改革委员会令2014年第7号)、《国家发展改革委投资咨询评估管理办法》(发改投资规〔2022〕632号)，以及北京、山东、河北、山西、湖北、广西等省、市、自治区级的政务信息化项目预算编制指南均提出要求开展软件造价评估活动。

我国自2013年开始陆续发布了SJ/T11463—2013《软件研发成本度量规范》、GB/T 36964—2018《软件工程 软件开发成本度量规范》、GB/T 28827.7—2022《信息技术服务 运行维护 第7部分：成本度量规范》等一系列标准，主要解决了软件项目规模度量以及软件项目建设和运维成本度量的标准化问题。

政府信息化项目和大型国央企信息化项目预算编制、结算等软件造价评估场景需要大量软件造价工程师提供服务。

快速增长的市场需求和软件造价人工评估面临如下问题：

- 软件造价人工评估效率低，评估项目周期长
- 软件造价人工评估质量参差不齐
- 优秀的软件造价工程师培养周期长，无法快速填补市场的缺口
- 各评估委托单位用于软件造价评估的成本较高

一些软件造价评估机构及厂商投入研发力量研发软件造价评估平台，提高软件造价评估工作的效率，降低评估成本。

面对市场上各种软件开发造价评估平台，用户在选择平台时，均面临如下问题：

- 评估结果是否符合国家标准
- 评估结果是否准确
- 是否能适配特定行业的特殊要求
- 上传的需求文档是否安全，保护隐私
- 是否随着软件发展，评估结果更合理
- 是否能集成用户现有流程和系统
- 是否支持私有化部署
- 是否适配国产化改造

同样，评测机构需要一份比对在评测软件开发造价评估平台时同样，软件开发造价评估平台开发厂商也需要一个作为对标要求。

为指导软件开发造价评估平台能力建设，提升软件开发造价评估平台服务质量，评估软件开发造价评估平台能力，提供相关各方选择合适的软件开发造价评估平台依据，特制订本文件。

软件与软件工程 软件开发造价评估平台 能力要求

1 范围

本文件规定了软件开发造价评估平台（以下简称“平台”）能力要求的模型框架，确定了平台能力等级划分及每个能力域的能力项要求。

本文件适用于：

a） 开发厂商对标平台能力要求建设与持续改进平台；

b） 第三方依据本文件实施平台能力等级评估工作；

c） 用户依据本文件选择平台。

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 4754—2017 国民经济行业分类

GB/T 36964—2018 软件工程 软件开发成本度量规范

GB/T 42449—2023 系统与软件工程 功能规模测量 IFPUG方法

GB/T 42588—2023 系统与软件工程 功能规模测量 NESMA方法

GB/T 45288.1—2025 人工智能 大模型 第1部份 通用要求

T/CCUA 005—2024 软件造价评估实施规程

3 术语和定义

下列术语和定义适用于本文件。

3.1

大模型 large-scale mode

大模型即大规模深度学习模型。基于大量数据训练得到，具有复杂计算架构，能处理复杂任务，且具备一定泛化性的深度学习模型。

注：大模型的参数量由其功能和模态决定，一般不低于 1 亿。大模型训练使用的数据总量受参数量的影响，达到收敛的大模型的参数量的对数与其训练数据总量的对数成正比。

[GB/T 45288.1—2025，定义3.1]

3.2

等级 level

根据各能力项达成程度确定的平台综合能力水平。

3.3

国产化 localization

平台由自主可控的组件、大模型和算法构建，且运行在自主可控的软硬件环境中。

3.4

行业 industry

从事相同性质的经济活动的所有单位的集合。

[GB/T 4754—2017,定义2.1]

3.5

基本功能组件 base functional component

功能规模测量方法中为了测量目标而定义和使用的用户功能需求的基本单位。

注:NESMAFPA 方法中的基本功能组件有内部逻辑文件、外部逻辑文件、外部输入、外部输出和外部查询。通过这五种基本功能组件，功能用户需求的功能规模被测量。

[GB/T 42588—2023,定义3.1.5]

3.6

能力项 capability item

为满足特定预期效果所应具备的某项信息处理能力。

3.7

能力域 capability domain

处理某类特定问题的一组相关能力项的集合。

3.8

评估 evaluate

符合GB/T 36964—2018软件开发成本度量规范要求度量软件开发规模、工作量和成本的活动。

3.9

软件开发造价评估平台 software development cost evaluation platform

评估产生软件产品的一整套活动所需付出的各种资源代价总和的平台。

3.10

审核 review

检查评估结果或报告是否符合规范要求的活动。

3.11

微调 fine-tuning

为提升机器学习模型预测准确性,使用专门领域数据在大模型上继续训练的过程。

注1:专门领域数据一般是特定场景的生产数据或合成数据。

注2:常用的微调方法包括提示词微调、全参微调、参数高效微调等。

[GB/T 45288.1—2025,定义3.4]

4 缩略语

下列缩略语适用于本文件。

CSBMK®：中国软件基准数据（China Software Benchmarking）

FP：功能点（Function Point）

5 能力模型框架

能力模型框架包括能力域和能力项，能力域由能力项组成。平台能力模型框架见图 1。其中：

---能力域包括核心功能能力域、管理与支撑功能能力域和非功能能力域；

---核心能力域包括造价评估过程合规能力项、评估输入解析能力项、软件规模评估能力项、软件工作量成本评估能力项、造价评估过程管理能力项和学习进化能力项；

---管理与支撑功能能力域包括系统集成能力项和系统管理功能能力项；

---非功能能力域包括安全能力项、国产化能力项和性能能力项。

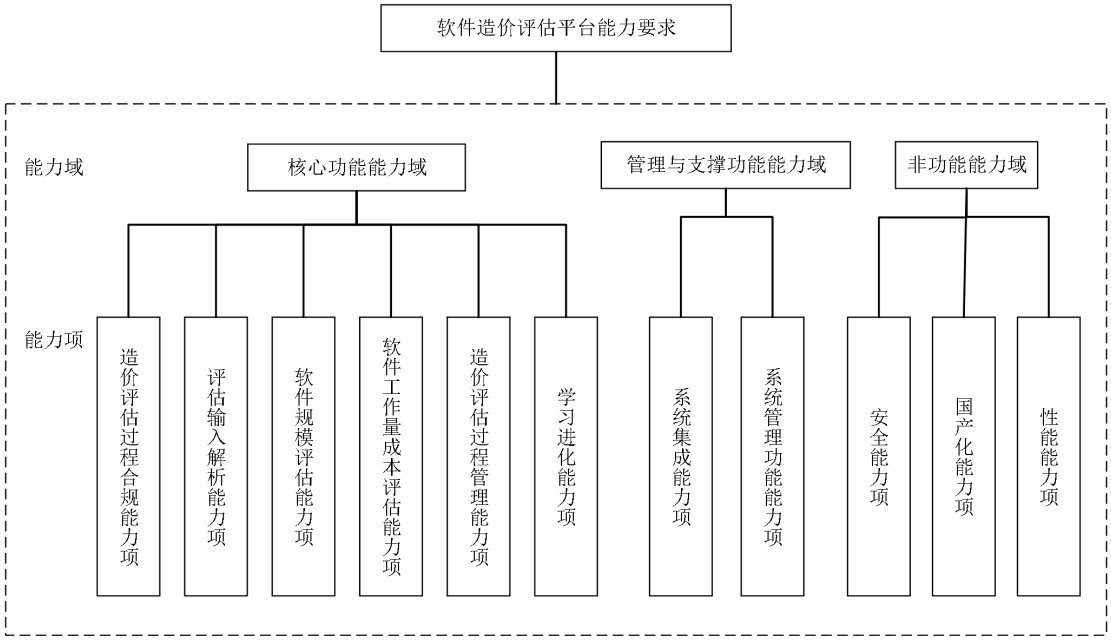

图 1　平台能力要求模型框架

6　平台能力等级

6.1　L1 级

适用于功能点人工计数结果审核及评估文档在线维护场景。平台能够指出特定格式评估结果的明显错误，并对评估过程相关文档在线维护。

6.2　L2 级

适用于项目审核及人工评估结果筛查场景。平台能够评估与审核大量送审项目的规模、工作量、成本，并较准确识别出评估结果偏差较大的项目；同时，平台具备管理基准数据、评估结果、评估报告、评估流程、在线维护评估过程相关文档的能力。

6.3　L3 级

适用于特定行业软件造价评估及审核场景。平台能够较为准确地评估或审核特定行业的软件开发造价，仅需要少量人工调整；同时，平台应具备定制评估流程、评估报告能力，与第三方平台集成与被集成能力，私有化部署能力，大模型安全可靠，运行环境国产化。

6.4　L4 级

适用于多个行业软件造价评估及审核场景。平台能够准确地评估多个行业软件开发造价，评估结果准确性不低于软件造价评估专家通常水平；同时，平台应具备自主学习能力，定制评估流程、评估报告

的能力，与第三方平台集成与被集成能力，私有化部署能力，大模型安全可靠、自主可控，运行环境国产化。

7 L1级能力要求

7.1 核心功能能力域

7.1.1 造价评估过程合规能力项

造价评估过程合规能力项要求如下：

a) 平台提供的软件开发成本估算基本流程应符合GB/T 36964—2018中7.1.1的规定；
b) 平台提供人工从CSBMK®基准数据集中选择调整因子、基准数据的功能，评估软件工作量和成本。

7.1.2 评估输入解析能力项

评估输入解析能力项要求如下：

a) 平台具备从特定格式评估结果解析基本功能组件的能力。

7.1.3 软件规模评估能力项

软件规模评估能力项要求如下：

a) 平台具备审核人工计数的特定格式评估结果，指出明显错误的能力；
b) 平台提供在线人工调整基本功能组件的功能。

7.1.4 软件工作量成本评估能力项

软件工作量成本评估能力项要求如下：

a) 平台应具备人工选择生产率、调整因子和人月单价的能力；
b) 平台应具备审核人工评估的基于功能点分析法评估的规模（FP），采用方程法评估软件开发工作量和成本的能力。

7.1.5 造价评估过程管理能力项

造价评估过程管理能力项要求如下：

a) 平台应具备在线管理每一个项目的需求、人工评估结果、审核结果表和审核报告的能力。

7.1.6 学习进化能力项

学习进化能力项要求如下：

a) 平台应具备维护规模审核规则的能力。

7.2 管理与支撑功能能力域

7.2.1 系统集成能力项

系统集成能力项要求如下：

a) 平台应具备通过页面、接口、服务等方式被外系统集成的能力。

7.2.2 系统管理功能能力项

系统管理功能能力项能力要求如下：

a) 平台应具备用户、权限、菜单管理的能力；
b) 平台应具备报表管理的能力；
c) 平台应具备监控和消息管理的功能；
d) 平台应具备评估调整因子、评估规则、人月单价管理的能力。

7.3 非功能能力域

7.3.1 安全能力项

安全能力项要求如下：
a) 平台应达到网络安全等级保护规定的系统相关要求。

7.3.2 国产化能力项

国产化能力项要求如下：
a) 平台应具备在中国信息安全测评中心安全可靠测评公告公布的软硬件环境上运行的能力。

7.3.3 性能能力项

性能能力项要求如下：
a) 平台应具备管理与支撑功能快速响应的能力；
b) 平台应具备在较短时间内提供评估结果审核结果的能力。

8 L2级能力要求

8.1 核心功能能力域

8.1.1 造价评估过程合规能力项

造价评估过程合规能力项要求如下：
a) 平台的软件开发成本估算基本流程应符合GB/T 36964—2018中7.1.1的规定；
b) 平台软件造价评估总体流程应符合T/CCUA 005—2024中5.1的规定；
c) 平台具备人工选择CSBMK®基准数据的能力；
d) 平台应具备根据T/CCUA 005—2024中5.7的规定审核报告的能力。

8.1.2 评估输入解析能力项

评估输入解析能力项要求如下：
a) 平台具备经手工调整后的需求中解析基本功能组件的能力。

8.1.3 软件规模评估能力项

软件规模评估能力项要求如下：
a) 平台应符合功能点分析法的要求，基于需求，通过推理识别基本功能组件类型筛查人工评估结果；
b) 估算方法应选用已发布的GB/T 42588—2023或GB/T 42449—2023等功能规模测量标准中的一种；
c) 平台多次评估同一份需求结果应一致；
d) 平台应具备在线人工调整基本功能组件的能力；

e) 平台应具备基于需求，审核人工计数的特定格式评估结果，指出明显错误，并给出纠正建议的能力。

8.1.4 软件工作量成本评估能力项

软件工作量成本评估能力项要求如下：

a) 平台应具备基于功能点分析法评估的规模（FP），采用方程法评估软件开发工作量和成本的能力；
b) 平台应具备人工选择生产率、调整因子和人月单价的能力；
c) 平台应具备审核人工评估的基于功能点分析法评估的规模（FP），采用方程法评估软件开发工作量和成本的能力。

8.1.5 造价评估过程管理能力项

造价评估过程管理能力项要求如下：

a) 平台应具备管理每个项目的评估方法、评估规则、评估调整因子、人月单价与项目需求、评估结果表、评估报告和审核报告的能力；
b) 平台应具备查询每个项目的评估流程的每个节点的能力；
c) 平台应具备查询每个项目调用大模型信息的能力。

8.1.6 学习进化能力项

学习进化能力项要求如下：

a) 平台使用的大语言模型应具备微调和优化的能力；
b) 平台应具备规则库持续迭代更新、接入、查询和使用的能力。

8.2 管理与支撑功能能力域

8.2.1 系统集成能力项

系统集成能力项要求如下：

a) 平台应具备通过页面、接口、服务等方式被外系统集成的能力。

8.2.2 系统管理功能能力项

系统管理功能能力项能力要求如下：

a) 平台应具备用户、权限、菜单管理的能力；
b) 平台应具备数据分析统计的能力；
c) 平台应具备监控和消息管理的功能；
d) 平台应具备调整因子、评估规则、人月单价管理、规则库管理的能力；
e) 平台应具备评估模板管理、评估报告模板管理的能力。

8.3 非功能能力域

8.3.1 安全能力项

安全能力项要求如下：

a) 平台应达到网络安全等级保护规定的系统相关要求；
b) 平台应具备安全存储大模型模型数据的能力。

8.3.2 国产化能力项

国产化能力项要求如下：

a) 平台应具备在中国信息安全测评中心安全可靠测评公告公布的软硬件环境上运行的能力。

8.3.3 性能能力项

性能能力项要求如下：

a) 平台应具备管理与支撑功能快速响应的能力；
b) 平台应具备可接受时间范围内提供规模、工作量和成本评估结果的能力；
c) 平台应具备较短时间内提供规模、工作量和成本审核结果的能力。

9 L3级能力要求 显著降低评估人员工作量

9.1 核心功能能力域

9.1.1 造价评估过程合规能力项

造价评估过程合规能力项要求如下：

a) 平台的软件开发成本估算基本流程应符合GB/T 36964—2018中7.1.1的规定；
b) 平台软件造价评估流程应符合T/CCUA 005—2024中5.1的规定；
c) 平台应具备人工选择包含CSBMK®在内的基准数据的能力；
d) 平台自动生成的评估报告应符合T/CCUA 005—2024中5.7的规定；
e) 平台应披露规模、工作量和成本的解析算法；
f) 平台使用的大模型应合规。

9.1.2 评估输入解析能力项

评估输入解析能力项要求如下：

a) 平台具备从输入文本文件中自动抽取功能性需求，并解析长文本自然语言为基本功能组件的能力。

9.1.3 软件规模评估能力项

软件规模评估能力项要求如下：

a) 平台应符合功能点分析法的要求，基于需求，推理识别的基本功能组件类型和数量准确率较高，与需求高度相关，且离散度小；
b) 估算方法应选用已发布的GB/T 42588—2023或GB/T 42449—2023等功能规模测量标准中的一种；
c) 平台多次评估同一份需求结果应一致；
d) 平台应具备在线人工调整基本功能组件的能力；
e) 平台应具备基于需求，审核人工计数的特定格式评估结果，指出明显错误，并给出纠正建议的能力；
f) 平台应具备规范基本功能组件命名的能力；
g) 平台应具备高度适配特定行业特点，推理识别的基本功能组件的能力。

9.1.4 软件工作量成本评估能力项

软件工作量成本评估能力项要求如下：

a) 平台应具备基于功能点分析法评估的规模（FP），采用方程法评估软件开发工作量和成本的能力；
b) 平台应具备人工选择生产率、调整因子和人月单价的能力；
c) 平台应具备审核人工评估的基于功能点分析法评估的规模（FP），采用方程法评估软件开发工作量和成本的能力；
d) 平台应适用于特定行业的预算、招投标、需求变更及决算结算等场景。

9.1.5 造价评估过程管理能力项

造价评估过程管理能力项要求如下：

a) 平台应具备管理每个项目的评估方法、评估规则、评估调整因子、人月单价与项目需求、评估结果表、评估报告和审核报告的能力；
b) 平台应具备查询每个项目的评估流程的每个节点的能力；
c) 平台应具备查询每个项目调用大模型信息的能力；
d) 平台应具备查询历史每个评估结果的基本功能组件与需求对应关系的能力；
e) 平台应具备定制软件造价评估流程的能力。

9.1.6 学习进化能力项

学习进化能力项要求如下：

a) 平台使用的大语言模型应具备训练、微调和优化的能力；
b) 平台应根据特定行业需求构建专属的规则库，具备规则库持续迭代更新、接入、查询和使用的能力；
c) 平台应具备模型历史版本和微调迭代过程的日志及查询功能,包含准确率、损失、参数更新等信息；
d) 平台使用的大语言模型数据资源应深度覆盖单个行业特征。

9.2 管理与支撑功能能力域

9.2.1 系统集成能力项

系统集成能力项要求如下：

a) 平台应具备通过页面、接口、服务等方式被外系统集成的能力；
b) 平台应具备通过页面、接口、服务等方式集成第三方系统的能力。

9.2.2 系统管理功能能力项

系统管理功能能力项能力要求如下：

a) 平台应具备用户、权限、菜单管理的能力；
b) 平台应具备数据分析统计的能力；
c) 平台应具备监控和消息管理的功能；
d) 平台应具备调整因子、评估规则、人月单价管理、规则库、知识库管理的能力；
e) 平台应具备评估模板管理、评估报告模板管理的能力；
f) 平台应具备系统备份管理的能力；
g) 平台应具备模型管理的能力。

9.3 非功能能力域

9.3.1 安全能力项

安全能力项要求如下：

a) 平台应达到网络安全等级保护规定的系统要求；
b) 平台应具备安全存储大模型模型数据的能力；
c) 平台应具备大模型训练数据隐私保护的能力；
d) 平台应具备公开透明服务，公开使用的模型信息和使用用途的能力；
e) 平台应具备大模型反作弊系统,黑白名单管理,管理作弊用户和风险用户的能力;
f) 平台应具备内容安全审核机制，拒绝接受与生成不良信息的能力；
g) 平台及使用的大模型应具备私有化部署能力；
h) 平台应使用已备案的基础模型。

9.3.2 国产化能力项

国产化能力项要求如下：

a) 平台应具备在中国信息安全测评中心安全可靠测评公告公布的软硬件环境上运行；
b) 平台使用的大多数组件应国内自主可控。

9.3.3 性能能力项

性能能力项要求如下：

a) 平台应具备高并发情况下，管理与支撑功能应快速响应的能力；
b) 平台应具备可接受时间范围内提供规模、工作量和成本评估结果的能力；
c) 平台应具备高并发情况下，较短时间内应提供规模、工作量和成本审核结果的能力。

10 L4 级能力要求 人调整能达到

10.1 核心功能能力域

10.1.1 造价评估过程合规能力项

造价评估过程合规能力项要求如下：

a) 平台的软件开发成本估算基本流程应符合GB/T 36964—2018中7.1.1的规定；
b) 平台软件造价评估流程应符合T/CCUA 005—2024中5.1的规定；
c) 平台应具备人工选择包含CSBMK®在内的基准数据的能力；
d) 平台自动生成的评估报告应符合T/CCUA 005—2024中5.7的规定；
e) 平台应披露规模、工作量和成本的算法；
f) 平台使用的大模型应合规。

10.1.2 评估输入解析能力项

评估输入解析能力项要求如下：

a) 平台具备从文本、音频等不同模态的输入文件中自动抽取功能性需求，并解析长文本自然语言为基本功能组件的能力。

10.1.3 软件规模评估能力项

软件规模评估能力项要求如下：

a) 平台应符合功能点分析法的要求，基于需求，推理识别的基本功能组件类型和数量准确率高，达到软件造价评估专家通常水平；
b) 估算方法应选用已发布的GB/T 42588—2023或GB/T 42449—2023等功能规模测量标准中的一种；
c) 平台多次评估同一份需求结果应一致；
d) 平台应具备在线人工调整基本功能组件的能力；
e) 平台应具备基于需求，审核人工计数的特定格式评估结果，指出明显错误，并给出纠正建议的能力；
f) 平台应具备规范基本功能组件命名的能力；
g) 平台应具备高度适配多个行业特点，推理识别的基本功能组件的能力；
h) 平台应具备分析分散在多个文件中一份完整需求的能力。

10.1.4 软件工作量成本评估能力项

软件工作量成本评估能力项要求如下：

a) 平台应具备基于功能点分析法评估的规模（FP），采用方程法评估软件开发工作量和成本的能力；
b) 平台应具备选择生产率、调整因子和人月单价的能力；
c) 平台应具备审核人工评估的基于功能点分析法评估的规模（FP），采用方程法评估软件开发工作量和成本的能力；
d) 应适用于多个行业的预算、招投标、需求变更及决算结算等场景。

10.1.5 造价评估过程管理能力项

造价评估过程管理能力项要求如下：

a) 平台应具备管理每个项目的评估方法、评估规则、评估调整因子、人月单价与项目需求、评估结果、评估报告，审核评估结果和评估报告的能力；
b) 平台应具备查询每个项目的评估流程的每个节点的能力；
c) 平台应具备查询每个项目调用大模型信息的能力；
d) 平台应具备查询历史每个评估结果的基本功能组件与需求对应关系的能力；
e) 平台应具备定制软件造价评估流程的能力。

10.1.6 学习进化能力项

学习进化能力项要求如下：

a) 平台使用的大语言模型应具备训练、微调和优化的能力；
b) 平台应根据多个行业需求构建专属的规则库，具备规则库持续迭代更新、接入、查询和使用的能力；
c) 平台应具备模型历史版本和微调迭代过程的日志及查询功能,包含准确率、损失、参数更新等信息；
d) 平台使用的大语言模型数据资源应深度覆盖多个行业特征；
e) 平台应具备行业知识库接入和使用,根据多个行业需求构建专属的行业知识库或知识图谱,提升模型和问答效果的能力。

10.2 管理与支撑功能能力域

10.2.1 系统集成能力项

系统集成能力项要求如下：

a) 平台应具备通过页面、接口、服务等方式被外系统集成的能力；
b) 平台应具备通过页面、接口、服务等方式集成第三方系统的能力。

10.2.2 系统管理功能能力项

系统管理功能能力项能力要求如下：

a) 平台应具备用户、权限、菜单管理的能力；
b) 平台应具备数据分析统计的能力；
c) 平台应具备监控和消息管理的功能；
d) 平台应具备调整因子、评估规则、人月单价管理、规则库、知识库管理的能力；
e) 平台应具备系统备份管理的能力；
f) 平台应具备评估模板管理、评估报告模板管理的能力；
g) 平台应具备模型管理的能力。

10.3 非功能能力域

10.3.1 安全能力项

安全能力项要求如下：

a) 平台应达到网络安全等级保护规定的系统要求;
b) 平台应具备安全存储大模型模型数据的能力；
c) 平台应具备大模型训练数据隐私保护的能力；
d) 平台应具备查询大模型训练语料来源、内容、标注的能力；
e) 平台应具备公开透明服务，公开使用的模型信息和使用用途的能力；
f) 平台应具备大模型反作弊系统,黑白名单管理,管理作弊用户和风险用户的能力;
g) 平台应具备内容安全审核机制，拒绝接受与生成不良信息的能力；
h) 平台应符合行业特殊安全要求；
i) 平台及使用的大模型应具备私有化部署能力。
j) 平台应使用已备案的基础模型；
k) 平台使用的大模型应抗对抗攻击。

10.3.2 国产化能力项

国产化能力项要求如下：

a) 平台应具备在中国信息安全测评中心安全可靠测评公告公布的软硬件环境上运行；
b) 平台使用的所有组件、算法、大模型应国内自主可控。

10.3.3 性能能力项

性能能力项要求如下：

a) 平台应具备高并发情况下，管理与支撑功能应快速响应的能力；
b) 平台应具备高并发情况下，较短时间内提供规模、工作量、成本评估结果的能力；
c) 平台应具备高并发情况下，较短时间内提供规模、工作量、成本审核结果的能力。

附 录 A
（资料性）
平台能力等级主要差异

表A.1 平台能力等级主要差异

能力域	能力项	L1级	L2级	L3级	L4级
核心功能能力域	造价评估过程合规能力项	1）评估流程符合GB/T 36964—2018要求	1）根据T/CCUA 005—2024要求审核评估报告	1）评估流程符合GB/T 36964—2018和T/CCUA 005—2024要求 2）自动生成的评估报告符合T/CCUA 005—2024要求 3）人工选择包含CSBMK®在内的基准数据 4）披露规模、工作量和成本的算法 5）使用的大模型应合规	1）与L3级一样
	评估输入解析能力项	1）从特定格式评估结果解析基本功能组件	1）经手工调整后的需求中解析基本功能组件	1）从输入文本文件中自动抽取功能性需求，并解析长文本自然语言为基本功能组件	1）从文本、音频等不同模态的输入文件中自动抽取功能性需求，并解析长文本自然语言为基本功能组件
	软件规模评估能力项	1）审核人工计数的特定格式评估结果，指出明显的错误 2）在线人工调整基本功能组件	1）符合功能点分析法的要求，基于需求，通过推理识别基本功能组件类型筛查人工评估结果 2）估算方法应选用已发布的GB/T 42588—2023或GB/T 42449—2023等功能规模测量标准中的一种 3）多次评估同一份需求结果应一致 5）基于需求，审核人工计数的特定格式评估结果，指出明显的简单错误，并给出纠正建议	1）符合功能点分析法的要求，基于需求，推理识别的基本功能组件类型和数量准确率较高，与需求高度相关，且离散度小 2）规范基本功能组件命名 3）高度适配特定行业特点，推理识别的基本功能组件	1）符合功能点分析法的要求，基于需求，推理识别的基本功能组件类型和数量准确率高，达到软件造价评估专家通常水平 2）高度适配多个行业特点，推理识别的基本功能组件 3）分析分散在多个文件中一份完整需求

表 A.1 平台能力等级主要差异（续）

能力域	能力项	L1级	L2级	L3级	L4级
核心功能能力域	软件工作量成本评估能力项	1）人工选择生产率、调整因子和人月单价 2）审核人工评估的基于功能点分析法评估的规模（FP），采用方程法评估软件开发工作量和成本	1）基于功能点分析法评估的规模（FP），采用方程法评估软件开发工作量和成本	1）适用于特定行业的预算、招投标、需求变更及决算结算等场景	1）适用于多个行业的预算、招投标、需求变更及决算结算等场景
	造价评估过程管理能力项	1）在线管理每一个项目的需求、人工评估结果、审核结果表和审核报告	1）管理每个项目的评估方法、评估规则、评估调整因子、人月单价与项目需求、评估结果表、评估报告和审核报告 2）查询每个项目的评估流程的每个节点 3）查询每个项目调用大模型信息	1）查询历史每个评估结果的基本功能组件与需求对应关系 2）定制软件造价评估流程	1）同L3级
	学习进化能力项	1）维护规模审核规则	1）使用的大语言模型应微调和优化 2）规则库持续迭代更新、接入、查询和使用	1）大语言模型训练、微调和优化 2）根据特定行业需求构建专属的规则库，规则库持续迭代更新、接入、查询和使用 3）模型历史版本和微调迭代过程的日志及查询功能，包含准确率、损失、参数更新 4）使用的大语言模型数据资源应深度覆盖特定行业特征	1）根据多个行业需求构建专属的规则库，具备规则库持续迭代更新、接入、查询和使用 2）使用的大语言模型数据资源应深度覆盖多个行业特征 3）行业知识库接入和使用，根据多个行业需求构建专属的行业知识库或知识图谱，提升模型和问答效果

表 A.1 平台能力等级主要差异（续）

能力域	能力项	L1级	L2级	L3级	L4级
管理与支撑功能域	系统集成能力项	1）被外系统集成	1）同L1级	1）集成外系统	1）同L3级
	系统管理功能能力项	1）用户、权限、菜单管理 2）数据分析统计 3）监控和消息管理 4）调整因子、评估规则、人月单价管理	1）调整因子、评估规则、人月单价管理、规则库管理 2）评估模板管理、评估报告模板管理	1）调整因子、评估规则、人月单价管理、规则库、知识库管理 2）系统备份管理 3） 模型管理	1）同L3级
非功能功能域	安全能力项	1）达到网络安全等级保护规定的系统相关要求	1）安全存储大模型模型数据	1）大模型训练数据隐私保护 2）公开透明服务，公开使用的模型信息和使用用途 3）大模型反作弊系统，黑白名单管理，管理作弊用户和风险用户 4）内容安全审核机制，拒绝接受与生成不良信息 5）私有化部署	1）查询大模型训练语料来源、内容、标注 2）符合行业特殊安全要求
	国产化能力项	1）在中国信息安全测评中心安全可靠测评公告公布的软硬件环境上运行	1）同L1级	1）使用的大多数组件应国内自主可控	1）使用的所有组件、算法、大模型应国内自主可控
	性能能力项	1）管理与支撑功能应快速响应 2）较短时间内提供评估审核结果	1）管理与支撑功能快速响应 2）在可接受时间范围内提供规模、工作量和成本评估结果 3）较短时间内提供规模、工作量和成本审核结果	1）高并发情况下，管理与支撑功能应快速响应 2）可接受时间范围内提供规模、工作量和成本评估结果 3）高并发情况下，较短时间内应提供规模、工作量和成本审核结果	1）高并发情况下，管理与支撑功能应快速响应 2）备高并发情况下，较短时间内提供规模、工作量、成本评估结果 3）高并发情况下，较短时间内提供规模、工作量和成本审核结果

注：表A.1中的L2级到L4级的每个能力项要求均包含了其前一个等级的能力项要求，此处仅列出其新增或差异部分。

附 录 B
（资料性）
典型应用场景

B.1 L1 级应用场景

L1级适用于功能点人工计数结果审核及评估文档在线维护场景。

某企业的信息化项目建设主要是基于成品软件的定制化开发，项目有限。该企业有一个软件造价评估人员，评估软件定制开发的规模、工作量和成本，但需要对评估结果进行审核。

达到能力要求L1级的平台适用于功能点人工计数结果审核及评估文档在线维护场景。该企业选择L1级平台审核人工计数的基本功能组件正确性，修改平台发行明显有问题的基本功能组件，最后生成报告。

B.2 L2 级应用场景

L2级适用于项目审核及人工评估结果筛查场景。

某大型集团企业数字化部负责全集团各分子公司信息化大型项目立项审核，审核内容包括软件开发造价。立项审核流程有一个软件开发造价合理性定性环节，初步筛查上报金额异常的项目。大数据局具备软件开发造价评估能力的人员数量有限，无法在短时间内审核大量的送审项目，需要借助软件开发造价评估平台能力筛查送审金额与需求偏离度非常大的项目，以快速完成审核任务。

达到能力要求L2级的平台能够对大量送审项目的规模、工作量、成本进行评估与审核，并能够较准确识别出评估结果偏差较大的项目。该集团选择L2级平台，先设置一个偏差阈值，然后人工选择行业调整因子基准值，企业人月单价，再使用平台评估每个项目的软件开发成本，比较评估结果与送审金额，最后筛查出超过阈值的项目，进一步重点审核。

B.3 L3 级应用场景

L3级适用于特定行业软件造价评估及审核场景。

某大型金融企业，每年以项目外包方式在信息化项目软件开发上投入上亿的资金，建设上百个项目，其中涉及关基系统。为了合理使用资金，该大型金融企业在招标阶段为每个项目设置最高限价，项目开发造价评估准确度要求较高。金融行业因其软件特殊性，不同系统有不同的规模评估要求。为了符合管理及审计要求，将评估流程嵌入到采购流程和项目管理流程中，实现评估输入材料和输出结果线上流转，每次评估过程和结果可追溯，平台本地化部署。该金融企业的软件开发造价评估人员少，可投入评估的资金有限。

达到能力要求L3级的平台准确性高，仅需要少量人工调整即可出具评估报告。该金融企业私有化部署L3级平台，并将评估流程各节点通过接口方式集成到企业的采购管理系统和项目管理系统中，完成一键登录和协同办公。在评估软件开发造价时，将需求文件输入平台，人工选择各调整因子和人月单价，待平台输出评估结果后，简单人工在线调整基本功能组件，输出报告，供后续采购流程使用，并可追溯。

B.4 L4 级应用场景

适用于多个行业软件造价评估及审核场景。

某省级大数据局，每年有大量维护、升级改造的项目需要结算，结算金额和工单规模、工作量密切相关。项目涉及医疗、交通、公安等多个行业，因行业不同，规模评估规则不同。由于涉及最后结算、决算，造价评估结果准确性不低于软件造价评估专家通常水平。而该大数据局的软件开发造价评估人员少，可用于评估的资金有限，无法聘请第三方专家，故大数据局希望通过平台直接给出用于结算的软件开发造价，同时将结算评估流程嵌入到结算流程和项目管理流程中。为了符合审计要求，每次评估过程和结果可追溯，平台本地化部署。

达到能力要求L4级的平台准确性高，其评估结果准确性不低于软件造价评估专家水平。该大数据局私有化部署L4级平台，并将评估流程节点通过接口方式集成到企业的采购管理系统和项目管理系统中，完成一键登录和协同办公。当发起结算流程后，平台获取需求文件，根据不同行业，选择不同规则，短时间内即可输出评估结果和报告，供后续结算使用，并可追溯。

参 考 文 献

[1] GB/T 5271.1—2000 信息技术 词汇 第1部分：基本术语
[2] GB/T 8566—2007 信息技术 软件生存周期过程
[3] GB/T 11457—2006 信息技术 软件工程术语
[4] GB/T 36964—2018 软件工程 软件开发成本度量规范
[5] GB/T 42588—2023 系统与软件工程 功能规模测量 NESMA方法
[6] GB/T 45288.1—2025 人工智能 大模型 第1部份 通用要求
[7] GB/T 45288.2—2025 人工智能 大模型 第2部份 评测指标与方法
[8] GB/T 45288.3—2025 人工智能 大模型 第3部份 服务能力成熟度评估
[9] T/CCUA 005—2024 软件造价评估实施规程