

ICS 35.020
I653

T/ SIA

中国软件行业协会团体标准

T/SIA 038—2024

人工智能企业能力评估规范

Specifications for capability assessment of artificial intelligence enterprises

2024-6-15 发布

2024-6-15 实施

中国软件行业协会 发布

目录

前言.....	1
引言.....	2
人工智能企业能力评估规范.....	3
1 范围.....	3
2 规范性引用文件.....	3
3 术语和定义.....	3
3.1 人工智能.....	3
3.2 精确率.....	3
3.3 召回率.....	3
3.4 F1 值.....	3
3.5 混淆矩阵.....	3
3.6 ROC 曲线.....	3
3.7 AUC 值.....	3
3.8 交叉熵损失.....	4
3.9 鲁棒性.....	4
3.10 软件项目管理能力评估.....	4
4 人工智能企业能力评估要素.....	4
5 人工智能企业能力评估模型.....	5
5.1 企业基础.....	5
5.1.1 稳定性.....	5
5.1.2 人员规模.....	6
5.1.3 服务链.....	6
5.2 技术创新能力.....	6
5.2.1 研发投入.....	6
5.2.2 研发人员.....	6
5.2.3 研发成果.....	6
5.2.4 行业认可.....	6
5.3 项目管理与质量管理能力.....	6
5.3.1 项目级管理能力.....	6
5.3.2 组织级管理能力.....	7
5.3.3 质量管理能力.....	7
5.4 财务与信用状况.....	7
5.4.1 收入情况.....	7
5.4.2 成长性.....	7
5.4.3 信用状况.....	7
5.5 产品技术能力.....	7
5.5.1 人工智能企业通用产品能力.....	7
5.5.2 人工智能企业个性产品能力.....	8
6 评价标准等级划分.....	8
7 人工智能企业能力评估模型应用场景.....	8

前言

本标准由中国软件行业协会提出并归口。

本标准起草单位：北京软件和信息服务业协会、上海市软件行业协会、重庆市软件行业协会、山西软件行业协会、吉林省软件行业协会、辽宁省软件行业协会、山东省软件行业协会、福建省软件行业协会、浙江省软件行业协会、河南省软件服务业协会、湖南省软件行业协会、江苏省软件行业协会、安徽省软件行业协会、广东软件行业协会、广西软件行业协会、海南省软件行业协会、四川省软件行业协会、贵州省信息技术服务业协会、云南省软件行业协会、深圳市软件行业协会、大连软件行业协会、厦门市软件行业协会、中国软件行业协会项目管理专业委员会。

本标准起草人：陈宝国、龙飞、姚宝敬、刘谦、赵原、吕彦伟、吴开宇、李倩、刘灿军、范为民、王小号、陈云贺、于洋、喻晖、夏冰莹、董先权、吕晖、林明广、邓月琴、邓小华、杨兵、李巡生、郑飞、秦健、吴戈、黄旻、鹿麟、张劲明、张然。

引言

人工智能是当前全球最受关注的新兴科技。近年来，以生成式人工智能为代表的新一代人工智能技术加速创新发展，进入成果爆发期，不断催生新场景、新业态、新模式、新市场，展现出广泛应用前景。2024年全国两会期间，“人工智能+”首次被写入政府工作报告。报告提出，深化大数据、人工智能等研发应用，开展“人工智能+”行动，打造具有国际竞争力的数字产业集群，以加快发展新质生产力。人工智能（AI）以其强大的计算能力、自动化特征及精准的数据分析能力，全面渗透到各行各业中，驱动了行业融合向更高层次的智能化迈进。

为顺应人工智能领域的发展趋势及市场需求，维护市场秩序，加强行业自律，促进人工智能企业能力的不断进步，增强人工智能企业的技术创新及深化拓展能力，规范人工智能领域的产品质量与服务，特制订本标准。

本标准结合人工智能领域的发展趋势和企业的发展实践，对人工智能企业的基本能力、技术创新能力、项目管理及质量管理能力、财务与信用状况、产品技术能力五个方面的要素提出了要求。

本标准是由相关软件行业协会、企业、评估机构基于市场和行业发展需要而共同制定，有利于发挥行业自律和示范作用，促进人工智能企业的健康、可持续发展。

人工智能企业能力评估规范

1 范围

正在开展人工智能业务的企业均适用于本标准，以对企业整体水平及开展人工智能业务的能力进行客观评价。

本标准可作为第三方机构的评估、认证依据。

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/T22900-2009 科学技术研究项目评价通则

GB/T40147-2021 《科技评估通则》

GB/T40148-2021 《科技评估基本术语》

GB/T11457-2006 信息技术 软件工程术语

3 术语和定义

3.1 人工智能 Artificial Intelligence

人工智能是一门研究如何使计算机以及机器模拟、延伸和扩展人的智能行为的学科。它涉及计算机科学、认知科学、哲学、神经科学、心理学、数学、语言学等多个领域。人工智能的核心目标是创造能够理解环境、学习新知识、解决问题、作出决策，并以类似于或超越人类智能的方式进行有效沟通的智能体。

3.2 精确率 Precision

精确率（Precision）是一种衡量分类模型或信息检索系统性能的指标，它专注于所检索或预测为正例的样本中真正为正例的比例。

3.3 召回率 Recall

在机器学习和统计学中，特别是在评估分类模型性能时，是一个重要指标。它衡量的是模型识别出的所有真正正类占所有实际正类的比例。

3.4 F1值 F1-score

F1值是评估分类模型性能的一个重要指标，尤其适用于那些关注精确率和召回率平衡的场景。它是精确率和召回率的调和平均数。

3.5 混淆矩阵 Confusion Matrix

混淆矩阵又称为误差矩阵，是一种用于评估和可视化分类模型性能的标准工具，特别是在监督学习任务中。它以表格的形式呈现，用来总结分类器预测结果与实际类别标记之间的对应关系。混淆矩阵的大小通常是n行n列，其中n代表类别的数量。

3.6 ROC曲线 Receiver Operating Characteristic Curve

ROC曲线是一种在二分类模型中评估其性能的图形表示方法。ROC曲线通过绘制真阳性率（True Positive Rate, TPR）对假阳性率（False Positive Rate, FPR）的图形，展现了分类器在不同阈值设置下的敏感性和特异性平衡。

3.7 AUC值 Area Under the Curve

AUC值是指ROC曲线下方的面积。ROC曲线通过图形化展示了一种二分类模型在所有可能的分类阈值下，其真阳性率(True Positive Rate, TPR)与假阳性率(False Positive Rate, FPR)之间的权衡关系。

3.8 交叉熵损失 Cross-Entropy Loss

交叉熵损失也常称为逻辑损失(Logistic Loss)或负对数似然损失(Negative Log-Likelihood Loss)，是一种在机器学习和深度学习中广泛使用的损失函数，尤其适用于分类任务。其核心目的是衡量模型预测的概率分布 $q(y|x)$ 与真实的概率分布 $p(y|x)$ 之间的差异。

3.9 鲁棒性 Robustness

在机器学习和统计学中，鲁棒性指模型在面对异常数据、噪声、输入变化或模型假设部分不满足时，仍然能够保持稳定性能和准确预测的能力。鲁棒性强的模型更能抵抗外界干扰，提供可靠的预测结果。

3.10 软件项目管理能力评估 Competency of Project Management Model

中国软件行业协会为促进软件及信息服务业企业项目管理能力的不断提高、增强企业创新能力和国际竞争力、支撑企业转型升级和高质量发展、规范企业软件项目管理能力而研制的一套软件项目管理能力评价体系。通过评估企业的组织级管理能力、项目级管理能力和组织前瞻能力等要素，评价企业的项目管理水平。

4 人工智能企业能力评估要素

一级指标	二级指标	三级指标
企业基础	稳定性	成立年限
		长期目标与核心竞争力
		风险管理能力
	人员规模	人员数量规模
		企业员工结构
	服务链	客户群体
供应商群体		
技术创新能力	研发投入	研发投入占比
		研发资金增长率
	研发人员	研发人员数量与占比
		研发人员素质
		研发人员流动性
	研发成果	专利申请数量
		专利质量
		技术创新频率
	行业认可	国家级、省部级奖项
全国性协会组织创新类奖项		
国家科技重大专项		
项目管理与质量管理能力	项目级管理能力	项目规划与立项能力
		项目执行与监控能力
		项目结项能力
	组织级管理能力	获得 CPMM 相关评估证书情况
		组织级项目管理能力
		人力资源管理能力

	质量管理能力		知识积累情况
			获得 CPMM 相关评估证书情况
			质量管理体系的建设
			客户反馈机制的建设
			持续改进机制的建设
			应急响应机制的建设
财务与信用状况	收入情况		履约情况
			收入规模
			净利润
	成长性		净利润率
			主营业务收入增长率
			净利润增长率
	信用状况		总资产增长率
获得信用评估证书情况			
产品技术能力	人工智能技术考核指标		未在国家相应信用黑名单中
			准确率
			精确率
			召回率
			F1 值
			混淆矩阵
			ROC 曲线
			AUC 值
			交叉熵损失
			计算速度
			鲁棒性
			可扩展性
	可解释性		
	人工智能技术个性指标	计算机视觉	视频精度
			感知能力
		机器人技术	运动能力
			操作能力
			学习能力
			沟通能力
			决策能力
			任务完成能力
		生物识别	易用性
稳定性			

5 人工智能企业能力评估模型

5.1 企业基础

5.1.1 稳定性

对企业稳定性的考察包括以下内容：

- a) 考察企业成立年限；
- b) 考察企业的长期目标与核心竞争力；
- c) 考察企业的风险管理能力。

5.1.2 人员规模

对企业人员规模的考察包括以下内容：

- a) 考察企业的人员数量规模；
- b) 考察企业员工结构，包括年龄结构、学历结构、岗位结构等。

5.1.3 服务链

对企业在服务链中的稳定性考察内容如下：

- a) 考察企业客户群体的稳定性及其所属行业前景；
- b) 考察企业供应商的信誉与服务能力, 包括供应商的经营历史、业绩表现、业界声誉及客户评价，以判断其可靠性和专业性；
- c) 考察企业与供应商的合作稳定性，包括企业与供应商的合作年限、单一采购比例，以判断企业与供应商对整体供应链稳定性的影响；
- d) 考察企业供应商的可替代情况，以判断企业供应链的稳定性。

5.2 技术创新能力

5.2.1 研发投入

对企业的研发投入考察包括以下内容：

- a) 考察企业的研发投入总量以及研发支出占其总营业收入的比例，以判断企业对技术创新重视程度；
- b) 考察企业研发投入资金的增长率。

5.2.2 研发人员

对企业研发人员的考察包括以下内容：

- a) 考察企业中研发人员的数量与占比，以判断企业在研发团队建设上的投入力度；
- b) 考察企业研发人员的整体素质，包括专业技能、学历背景、工作年限等；
- c) 考察企业研发人员的流动率，以判断企业研发团队的稳定性。

5.2.3 研发成果

对企业的研究成果考察内容如下：

- a) 考察企业拥有的专利数量；
- b) 考察企业拥有专利的质量，包括专利的技术难度、创新性、经济效益等；
- c) 考察企业的专利产出频率，以判断企业技术创新频率。

5.2.4 行业认可

对企业的行业认可考察内容包括如下：

- a) 企业研究成果获得的国家、省级科技进步奖的数量；
- b) 企业研究成果获得全国性协会组织颁发的创新类奖的数量；
- c) 企业承担国家科技重大专项的数量。

5.3 项目管理与质量管理能力

5.3.1 项目级管理能力

对企业项目级管理能力的考察包括以下内容：

- a) 考察企业的项目规划与立项能力；

- b) 考察企业的项目执行与监控能力;
- c) 考察企业的项目结项能力;
- e) 考察企业获得 CPMM 相关证书的情况。

5.3.2 组织级管理能力

对企业组织级管理能力的考察包括以下内容:

- a) 考察企业组织级项目管理能力;
- b) 考察企业人力资源管理能力;
- c) 考察企业的知识积累情况;
- e) 考察企业获得 CPMM 相关证书的情况。

5.3.3 质量管理能力

对企业质量管理能力的考察包括以下内容:

- a) 考察企业质量管理体系建设情况;
- b) 考察企业客户反馈机制建设情况;
- c) 考察企业客户满意度及质量管理持续改进机制建设情况;
- d) 考察企业应急处理机制建设情况;
- e) 考察企业的履约情况。

5.4 财务与信用状况

5.4.1 收入情况

对企业收入情况的考察包括以下内容:

- a) 考察企业的收入规模;
- b) 考察企业的净利润;
- c) 考察企业的净利润率。

5.4.2 成长性

对企业的成长性考察包括以下内容:

- a) 考察企业的主营业务收入增长率;
- b) 考察企业的净利润增长率;
- c) 考察企业的总资产增长率。

5.4.3 信用状况

对企业信用状况的考察包括以下内容:

- a) 考察企业获得国家级协会 AA 以上级别的信用评价证书情况;
- b) 被考察企业须不在以下名单中:
 - i) 国家信用信息公示系统严重违法失信组织名单;
 - ii) 中国政府采购网政府采购严重违法失信行为记录名单;
 - iii) 国家公共资源交易平台交易诚信黑名单。

5.5 产品技术能力

5.5.1 人工智能企业通用产品能力

考察人工智能企业产品技术能力如: 准确率、精确率、召回率、F1 值、混淆矩阵、ROC

曲线、AUC 值、交叉熵损失、计算速度、鲁棒性、可扩展性、可解释性等指标。

5.5.2 人工智能企业个性产品能力

考察部分人工智能企业产品技术能力如：

- a) 针对计算机视觉类人工智能产品，需要考察视频精度等指标；
- b) 针对机器人技术类人工智能产品，需要考察感知能力、运动能力、操作能力、学习能力、沟通能力、决策能力、任务完成能力；
 - i) 感知能力包括视觉感知、听觉感知、触觉感知；
 - ii) 运动能力包括自主性、灵活性、稳定性；
 - iii) 操作能力包括精确性、效率、多样性；
 - iv) 学习能力包括快速学习和调整策略能力；
 - v) 沟通能力包括语言理解、语言表达、人机协作；
 - vi) 决策能力包括决策准确性、决策效率、鲁棒性；
 - vii) 任务完成能力包括任务完成率、任务执行效率。
- c) 针对生物识别类人工智能产品，需要考察其易用性、稳定性以及在不同环境和条件下的性能一致性。

6 评价标准等级划分

等级	定义	分值
一级	企业拥有完善的管理与高效的运营，拥有领先技术，人工智能技术能力高，市场竞争力强，能满足客户多样需求。	80分以上
二级	企业管理体系良好，可跟进行业创新趋势，产品技术先进，人工智能技术能力良好，市场竞争力良好。	70-80分
三级	企业管理体系健全，创新能力一般，产品技术基本达标，人工智能技术能力一般，市场竞争力平。	60-70分

7 人工智能企业能力评估模型应用场景

使用者可根据人工智能企业能力评估模型考察企业的整体水平及其人工智能产品技术能力，实现寻找可信、可靠、稳定的人工智能领域合作商的目标。应用场景包括以下主要内容：

- a) 采购方及招标机构可根据人工智能企业能力评估模型来对有能力承接对应项目需求的企业进行选择；
- b) 人工智能企业可通过此标准进行自评估。