

# 消防大模型评测指南

(征求意见稿)

## 编制说明

2025年09月

# 《消防大模型评测指南》

## 团体标准征求意见稿编制说明

### 1. 工作简况，包括任务来源、协作单位、主要工作过程、团体标准主要起草人及其所做的工作等

#### 1.1. 任务来源

《消防大模型评测指南》由上海市人工智能行业协会提出。

#### 1.2. 协作单位

本标准的起草单位有上海库帕思科技有限公司、上海人工智能实验室、应急管理部上海消防研究所、上海脉策数据科技有限公司等。

#### 1.3. 主要工作过程

标准编制期间，其主要工作过程如下：

1) 2025年7月，标准起草单位调研了行业标准现状，根据目前消防大模型评测标准缺失的现状，提出了标准起草需求；

2) 2025年8月，标准起草单位结合消防行业的专业知识，业务特点的特点，起草了标准草案；

3) 2025年8月20日，由上海市人工智能行业协会进行了立项评审，会后根据专家意见对草案进行了修改并提交评审资料，2025年8月，下达标准计划；

4) 2025年8月,标准牵头单位对外公开征集,组建了标准起草工作组;

5) 2025年8月29日,标准牵头单位收集标准草案意见并召开了第一次标准研制会,讨论了标准草案,要集中在评测框架、评测体系等方面进行讨论,共收集到3条书面意见,会后标准牵头单位根据收集的意见及现场意见对标准草案进行了标准框架优化与内容修改;

6) 2025年9月8日,标准牵头单位组织召开第二次标准研制会,讨论标准草案,并根据讨论情况对标准进行了修改,共收集到5条意见;

7) 2025年9月,标准牵头单位根据研制会的建议对标准进行了修改,形成征求意见稿。

#### 1.4. 标准主要起草人及其所做的工作

上海库帕思科技有限公司负责了本标准的架构制定工作,并对标准内容进行初审;上海库帕思科技有限公司主要负责本标准的全文编制工作并主导了标准的会议讨论和标准的整体审核工作;上海人工智能实验室主要负责大模型评测方法的编制内容;应急管理部上海消防研究所、上海脉策数据科技有限公司参加了标准工作组会议,提出标准内容的具体修改意见。

## 2. 标准编制原则和确定标准主要内容(如技术指标、参数、公式、性能要求、试验方法、检验规则等)的论据(包括试验、统计数据),修订标准时,应增列新旧标准水平的对比;

### 2.1. 标准制定的原则

标准编制遵循“统一性、适用性、一致性、规范性”的原则，注重标准的可操作性。本标准编写是执行 GB/T1.1-2020《标准化工作导则第1部分：标准化文件的结构和起草规则》。

## 2.2. 标准的主要内容（增加体现自创新部分的说明）

本文件规范了消防大模型的评测框架、评测方法和评测内容。

除了标准的规范性要素（范围、规范性引用文件、术语和定义等）外，本标准的正文部分主要内容包括：

a) 第5章语评测框架：从模型通用基础能力、消防安全与价值对齐、应急响应与快速决策、消防专业认知能力、消防业务场景应用能力五个维度进行要求；

b) 第6章评测内容：对模型通用能力、专业能力和业务能力等方面的评测内容进行定义；

c) 第7章 评测方法：规范了评测数据集、评测环境、评测工具和评测实施的要求。

## 3. 主要试验（或验证）的分析、综述报告，技术经济论证，预期的经济效果；

### 3.1. 试验验证分析

无；

### 3.2. 综述报告

当前智慧消防系统内部数据缺乏统一标准，数据质量参差不齐、表达形式不一，难以满足构建消防语料，实现人工智能训练与智能化应用的需要。具体包括以下几个必要性：

一是填补行业语料标准体系空白。目前消防行业虽已有部分数据元及信息代码标准，但尚缺专门面向 AI 应用和语料资源建设的系统性规范，尤其缺乏对多模态数据（文本、语音、图像、空间数据等）协同表达和结构化标注的要求。

二是应对“AI 原生智慧消防”发展的迫切需求。随着 AI 原生智慧消防平台的推进，行业对高质量、多源异构语料的统一表达和集中管理提出了更高要求，缺乏标准将严重制约模型训练效果与智能体场景的适配能力。

三是构建行业语料共享与安全机制的基础支撑。语料作为关键的战略资源，其规范化建设不仅是行业内部协同的基础，也为跨部门相关数据共享、平台化语料服务和产业化智能应用提供了安全与合规保障。

四是对接国家战略与产业发展趋势。在“人工智能+”上升为国家战略的背景下，消防行业需通过标准化手段，提升语料生产、组织与交付的专业化水平，增强我国在智慧应急和城市公共安全领域的技术话语权。

### 3.3. 技术论证

国标《GB/T45288.2-2025 人工智能大模型第 2 部分：评测指标与方法》规范了通用大模型的评测指标和评测内容；团体标准《T/SAIAS 019—2024 金融大模型应用测评指南》、《T/SAIAS 022—2025 医疗大模型应用测评指南》分别规范了金融与医疗领域的垂类大模型评测要求。

而消防行业特殊性，一方面需要大模型掌握覆盖防火、灭火、火灾调查、消防装备、消防产品等行业专业知识，另外一方面需要针对灭火的应用场景，需要模型具备快速的响应能力。目前国内缺乏对消防行业大模型在评测指南要求。

### 3.4. 预期的经济效果

对社会公共安全方面，指南确保应用于应急救援领域的人工智能技术，其行为和输出始终与生命至上的核心价值观和公共安全总体利益保持一致。

对消防总队应用方面，指南提供了一套科学的选型和验收“标尺”，能够有效评估和检验大模型在消防实战场景中的真实能力，确保采购和部署的模型能够真正满足时效性高的严苛要求，避免技术与实战的脱节。

对模型开发单位，指南指明了清晰的技术研发方向和能力建设目标，引导企业将研发资源投入到真正解决消防核心痛点、具备实战价值的功能上，而不是停留在通用问答层面，从而推动技术的精准赋能。

## 4. 采用国际标准和国外先进标准的程度，以及与国际、国外同类标准水平的对比情况，或与测试的国外样品、样机的有关数据对比情况；

国标《GB/T45288.2-2025 人工智能大模型第2部分：评测指标与方法》规范了通用大模型的评测指标和评测内容，本指南中对于模型通用基础能力的评测要求，引用该标准；团体标准《T/SAIAS 019—2024 金融大模型应用测评指南》、《T/SAIAS 022—2025 医疗大模型应用测评指南》分别规范了金融与医疗领域的垂类大模型评测要求。

本规范首次从模型通用基础能力、消防安全法规与价值对齐、应急响应与快速决策、消防专业认知能力、消防业务场景应用能力五个维度提出了消防大模型的评测内容与要求。

**5. 与有关的现行法律、法规和强制性国家标准的关系；**

与现行相关法律、法规、规章及相关标准无矛盾。

**6. 重大分歧意见的处理经过和依据；**

无。

**7. 标准的建议；**

建议作为团体标准实施，建议标准发布后立即实施。

**8. 贯彻标准的要求和措施建议（包括组织措施、技术措施、过渡办法等内容）；**

本标准发布后，标准起草组将作为标准应用推广的主体，组织科研院所、相关企业、第三方检测认证机构等进行标准的宣贯。同时，起草组对标准的核心内容进行解读，方便后续的应用。

**9. 废止现行有关标准的建议；**

无，本标准为首次制定标准。

**10. 其他应予说明的事项。**

无。

起草工作组

2025年09月