

ICS 25.160.01
CCS J 33



CWA

团 体 标 准

T/CWAN 0171—2025

焊接大模型构建与应用通用技术规范

General technical specification for construction and application of welding large-scale model

2025-11-17 发布

2025-12-01 实施

中国焊接协会 发布

目 录

前 言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 焊接大模型应用框架.....	2
5 数据合规要求.....	2
6 技术要求.....	3
7 评估与测试.....	6

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。

本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国焊接协会提出并归口。

本文件起草单位：北京博清科技有限公司、清华大学、江南造船（集团）有限责任公司、中国兵器工业集团航空弹药研究院有限公司、北京理工大学（珠海）、北京深势科技有限公司、上海中巽科技股份有限公司、广西柳工机械股份有限公司、哈尔滨职业技术大学、北部湾大学、哈尔滨中焊协学技术服务有限公司、福建省特种设备检验研究院、哈尔滨华德学院、坤智大数据科技（哈尔滨）有限公司、天津市特种设备监督检验技术研究院。

本文件主要起草人：冯消冰、韩滕跃、韩冬、许国敬、王贵锦、苏楠、袁飞、于兴华、常志贤、王铭秋、李思敏、范东辉、王宇、林晓辉、侯国清、孙明辉、王滨滨、邓军林、李长威、朱斌海、于修和、方乃文、陈博文、崔元彪。

焊接大模型构建与应用通用技术规范

1 范围

本文件规定了焊接大模型构建与应用的通用技术要求,涵盖大模型构建及应用过程中数据、预训练、推理、优化、基本功能及评测等各个阶段。

本文件适用于开发、部署和使用焊接大模型的焊接行业教育单位、科研机构、企业、技术开发者及相关组织等。

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 3375 焊接术语

GB/T 41867 信息技术 人工智能 术语

GB/T 42127—2022 智能制造 工业数据 采集规范

GB/T 42755 人工智能 面向机器学习的数据标注规程

GB/T 45288.1—2025 人工智能 大模型 第1部分:通用要求

GB/T 45288.2—2025 人工智能 大模型 第2部分:评测指标与方法

GB/T 45288.3—2025 人工智能 大模型 第3部分:服务能力成熟度评估

3 术语和定义

GB/T 3375 和 GB/T 41867 界定的及下列术语和定义适用于本文件。

3.1

焊接大模型 welding large-scale model

利用通用大语言模型或多模态大模型的自然语言理解、多模态理解、知识整合和推理能力作为基础,再通过焊接领域专业知识(如工艺规程、标准规范、科研论文、案例数据、专家经验等)进行持续预训练和微调,从而形成的焊接领域的垂域大模型。

3.2

焊接数据 welding data

在焊接活动的全生命周期中产生、采集、存储及使用的,能够被大模型感知、处理和学习,以支撑其实现焊接相关智能任务的多模态、结构化或非结构化的信息集合。

3.3

焊接数据模态 welding data modality

焊接相关数据在采集、表达与传递过程中所呈现的独立且特定的形式或类型。这些不同模态的数据共同构成了焊接大模型进行感知、学习、推理和决策的多维度信息基础。例如，“焊接变形：焊件由于焊接而产生的变形叫焊接变形”属于文本模态。

3.4

单模态焊接数据 mono-modal welding data

只包含一种焊接数据模态。

3.5

多模态焊接数据 multi-modal welding data

包含不少于两种焊接数据模态。

4 焊接大模型应用框架

焊接大模型框架分为底层模型和应用层 2 个部分，见图 1。

a) 底层模型是基于焊接知识数据进行训练，形成的焊接知识大模型。

b) 结合二次预训练、监督微调、外部知识库等技术手段，在底层模型的基础上形成的应用层，例如工艺推荐功能、焊接知识库等。



图1 焊接大模型应用框架

5 数据合规要求

5.1 总体原则

焊接大模型数据合规要求整体宜符合 GB/T45288.1 中数据资源的要求。

5.2 数据内容

a) 焊接数据内容包含但不限于材料性能、焊接工艺、焊接方法等；

b) 数据及术语宜按照 GB/T 3375 规定的进行表述；

c) 数据的获取流程包括工艺方法、测试方法、检验方法等，宜符合焊接材料试验、力学性能试验、焊接检验、焊接方法及工艺、焊接设备等宜符合 GB/T 42127—2022 中的要求；

- d) 原始数据宜保留采集时间、采集来源、设备型号、环境参数等元数据，保证数据的可溯源性；
- e) 宜具备两种及以上多源数据类型，多源数据的类型覆盖来源，包括但不限于：实验数据、工业现场数据、仿真数据、公开数据集；
- f) 数据内容审核宜采用自动化审核与人工审核结合的方式进行。

5.3 数据规模与分类

5.3.1 数据规模分级

a) 数据规模宜按有效完整数据样本量划分为小样本、中等样本、大样本三个等级，以应用于不同场景；

b) 小样本场景下宜提供数据增强方案，数据增强方式宜遵循 GB/T 45288.3 数据资源的相关规定；

5.3.2 数据划分规则

a) 数据集应根据适用情况，划分为训练集、验证集、测试集；

b) 训练集、验证集、测试集宜具备不同的划分比例；

c) 小样本场景下，宜满足训练集占比不得超过全量数据的 85%，测试集占比不得低于 10%；

d) 中等样本场景下，宜满足训练集占比不得超过全量数据的 80%，验证集占比不得低于 10%，测试集占比不得低于 10%；

e) 大样本场景下，宜满足训练集占比不得超过全量数据的 98%，验证集占比不得低于 1%，测试集占比不得低于 1%；

f) 测试集宜与训练集和验证集完全分离，不包含在训练过程中使用过的数据。

5.3.3 数据时效性

a) 宜建立数据集动态更新机制；

b) 自制数据集宜至少每 6 个月更新版本号，并附变更说明；

c) 宜具备版本管理功能。

5.4 数据质量要求

a) 数据收集宜按照 GB/T 42127—2022 中工业数据采集方式中规定的方式进行；

b) 数据清洗与预处理宜执行异常值剔除、缺失值填补等质量控制流程；

c) 数据标注宜按照 GB/T 42755 数据标注流程中规定的流程进行；

d) 数据集格式和接口宜符合 GB/T 42127—2022 中工业数据采集要求中的规定。

5.5 隐私与安全

a) 数据收集活动应遵循国家法律法规要求，对个人隐私信息进行保护；

b) 敏感数据应进行脱敏处理；

c) 应采用相应的安全防护措施保证数据传输的安全性；

d) 应具备数据管理功能，例如访问权限管理、日志审计。

6 技术要求

6.1 功能要求

6.1.1 通用基础能力

焊接大模型所具备的通用基础能力如下：

a) 单模态处理能力，包括但不限于：

将焊接领域的文本内容划分为不同的标签或类别；

能够对焊接领域的文本内容，提供合理、准确、可靠的咨询问答；

能够理解焊接图像的语义内容，并输出对应的类别标签等；

对焊接过程中产生的音频进行分类，识别正确的焊接方法。

b) 多模态处理能力，包括但不限于：

能够准确理解焊缝图像、坡口图像、熔池图像并生成描述；

对焊缝图像中的文本信息进行回答；

根据焊缝图像、坡口图像、熔池图像给出相关联的场景理解和知识推理；

根据音频内容给出相关联的焊接场景理解和知识推理。

6.1.2 价值对齐

焊接大模型的研发、训练及应用等全生命周期过程需确保符合一系列安全标准。

价值对齐：模型输出需符合国家法律法规要求，禁止模型生成可能引发安全隐患的焊接参数（如超出设备承载能力的电流、速度设置），避免推荐高污染焊接工艺（如含重金属焊材的使用）。

6.1.3 焊接领域应用能力

焊接大模型在焊接专业的认知能力，包括但不限于：

a) 焊接材料知识，包括但不限于焊丝、焊条、焊剂等，例如各种材料的熔点、热膨胀系数、导电性等；

b) 焊接方法及工艺标准知识，包括但不限于激光焊、激光复合焊、电弧焊、搅拌摩擦焊、焊接通用技术条件、焊接工艺规程、焊接工艺评定等；

c) 焊接质量试验及检测知识，包括但不限于钢材试验、焊接性试验、焊接材料试验、焊接缺陷分类、焊接质量、焊缝外观质量及检测方法等；

d) 焊接设备知识，包括但不限于工业机器人、爬行焊接机器人、焊接变位机、焊枪、切割机、视觉和焊接参数检测系统等；

e) 焊接安全与卫生知识，包括但不限于焊接与切割安全、防护屏安全等；

f) 焊工培训与考试知识，包括但不限于焊工培训规程、焊工技术考试规程等。

6.2 接口要求

6.2.1 文本类型接口要求

a) 文本需要进行分词处理，并利用词向量模型转化为嵌入向量；

b) 输入文本的最小长度不低于 512 Tokens；

- c) 宜用<s>代表文本序列起始符；
- d) 宜用</s>代表文本序列结束符；
- e) 宜用<unk>代表未知词标记；
- f) 宜用<pad>代表填充符。

6.2.2 图像类型接口要求

- a) 图像可先分块，需经过视觉编码器模型转化为嵌入向量；
- b) 图像尺寸不低于 224×224 px。
- c) 宜用代表图像序列起始符。
- d) 宜用代表图像序列结束符。
- e) 多张图像作为输入时，每张图像前面宜标记 **Picture id**：代表该图像在序列中的位置，如：**Picture 1: img1 Picture 2: img2**。

6.2.3 音频类型接口要求

- a) 原始音频可先经过音频编码器转化为嵌入向量；
- b) 音频的频率不宜低于 16 KHz；
- c) 宜用<audio>代表音频序列起始符；
- d) 宜用</audio>代表音频序列结束符；
- e) 多张音频作为输入时，每段音频前面宜标记 **Audio id**：代表该音频在序列中的位置，如：**Audio id 1: <audio>audio1</audio> Audio id 2: <audio>audio2</audio>**。

6.2.4 其他数据类型接口要求

- a) 3D 模型数据，可通过预处理手段将其转化为大模型的输入，例如取正视图、侧视图、顶视图等处理成多张图像，按照图像接口进行处理；
- b) 点云数据，可通过 PointNet 等深度学习模型将其处理为特征向量；
- c) 红外数据，可通过视觉模型作为编码器提取特征向量。

6.3 训练要求

6.3.1 预训练工具

宜符合 GB/T 45288.1—2025 中模型训练工具的要求。

- a) 宜使用 IB 网络进行多机之间的通讯；
- b) 宜支持全精度训练、量化训练、混合精度训练；
- c) 宜支持接入微调工具库，例如 XTuner 等微调框架。

6.3.2 模型保存格式

保存成 Huggingface 模型格式，包含但不限于 PyTorch 格式、TensorFlow 格式、Flax (JAX) 格式、SafeTensors 格式、ONNX 格式、GGUF 格式等。

6.4 模型优化要求

宜符合 GB/T 45288.1—2025 中模型优化工具的相关要求。

6.5 部署与推理要求

6.5.1 部署与推理工具

宜符合 GB/T 45288.1—2025 中模型部署与推理工具中的相关要求。

6.5.2 模型调用要求

支持 Curl、RESTful API、命令行等方式调用。

7 评估与测试

7.1 评测方法

焊接大模型的评测宜符合 GB/T 45288.2—2025 中评测方法的要求。采用客观评测方法和主观评测方法相结合的方式进行综合评估。

7.1.1 客观评测方法

客观评测方法是在各个任务的测试数据集上采用量化的指标进行评估。对焊接大模型具备的基础通用能力和领域应用能力进行测评，可分为单模态能力、多模态能力，具体评测内容的评价指标如表 1 所示。

表1 焊接大模型应用能力客观评测方法

一级 维度	二级 维度	评测内容	任务 类型	评价指标
焊接 大模 型	单模 态能 力	将焊接领域的文本内容划分为不同的标签或类别	文本 分类	准确率 Acc、召回率Recall、 精确率Precision、micro-F1 值
		能够对焊接领域的文本内容，提供合理、准确、可靠的咨询问答，包含焊接基础知识、焊接材料相关基础知识、焊接方法及工艺标准知识、焊接质量试验及检测相关知识、焊接设备知识、焊接安全与卫生知识、焊工培训与考试知识等方面	文本 问答	BLEU 指标、Rouge-L 指标、 EM
		能够理解焊接图像的语义内容，并输出对应的类别标签等	图像 分类	准确率 Acc、召回率Recall、 精确率Precision、micro-F1 值
		对焊接过程中产生的音频进行分类，识别正确的焊接方	音 频 分类	准确率Acc、召回率Recall、精 确率Precision、micro-F1 值

		法		
多模 态能 力		能够准确理解焊缝图像、坡口示意图、熔池图像并生成描述	图片生成文本描述	BLEU 指标、Rouge-L 指标、CIDEr
		对焊缝图像中的文本信息进行回答	图片问答	准确率Acc
		根据焊缝图像、坡口示意图、熔池图像给出相关联的场景理解和常识推理	视觉推理	准确率Acc
		根据音频内容给出相关联的场景理解和常识推理	音频推理	准确率Acc

评测还应包括模型的响应时间,评价指标准确率 Acc、召回率 Recall、精确率 Precision、micro-F1 值、BLEU 指标、Rouge-L 指标的计算方法宜符合 GB/T 45288.2—2025 中评测指标计算方法。

EM (Exact Match): 指标可被译为确切匹配或绝对匹配,是根据大模型给出的结果与标准值之间的匹配度来计算的。取值范围: 0~1, 完全匹配时(生成的答案与参考答案完全相同, 字符级匹配), EM 值为 1, 不完全匹配时, EM 值为 0。

CIDEr: 是一种用于评估图像描述生成系统的指标,旨在衡量机器生成的描述与一组人工参考描述的相似度,它基于一致性评估来度量模型生成描述的质量。

7.1.2 主观评测方法

主观评测方法主要针对焊接大模型的生成任务,包括但不限于生成文本答案、图像生成描述等生成能力相关的任务进行主观评估,宜包含如下维度:

- 生成文本质量: 包含语法正确性, 流畅度、表达清晰度等方面;
- 焊接专业知识准确性: 事实准确性、概念理解深度等方面;
- 专业能力: 和焊接相关的材料、工艺、设备等方面的专业度;
- 内容安全性: 避免暴力、歧视、违法内容生成;
- 功能安全性: 模型在不同的焊接领域应用能力的表现, 在施焊过程、焊接安全与卫生等方面输出结果的准确性、一致性和可靠性, 避免因模型错误导致的不当作业行为。

对 a)-e) 维度进行 5 分量化制评分, 5 分量化评分标准见表 2 所示。

表2 量化评分标准

分数	标准描述
----	------

5	优秀：完全满足需求，具有创新性/专业深度，无任何缺陷
4	较优秀：基本完美，仅存在极细微瑕疵
3	合格：满足主要需求，但存在可改进的明显不足
2	有缺陷：部分满足需求，存在关键错误或遗漏
1	不合格：完全偏离需求或包含严重错误

对 a)-e) 维度进行量化评分，总分在[5~10]之间，模型评价为一般；总分在[11~19]之间，模型评价为良好；总分在[20~25]之间，模型评价为优秀。

7.2 评测数据集

评测数据集整体上宜符合 GB/T 45288.2—2025 中评测数据集的相关要求。

a) 对于任务类型为文本分类、图像分类、音频分类的评测数据集中宜明确标出该文本、图像、音频所属的类别，可为单标签，可为多标签；

b) 对于任务类型为文本问答、图片生成文本描述的评测数据集中应明确给出标准答案；

c) 对于任务类型为图片问答、视觉推理、音频推理的评测数据集格式可为单选题、多选题、判断题等。

7.3 评测实施

宜根据受测试对象，分别对焊接大模型的各项能力进行客观评测和主观评测。

a) 客观评测实施

- 1) 根据构建的评测数据集计算评价指标；
- 2) 构建自动化评测脚本实现评价指标的计算。

b) 主观评测实施

- 1) 对评测人员进行评测标准培训，确保评测人员对评测的标准有统一的理解和执行；
- 2) 宜对每一条测试数据进行交叉评测，以确保评测结果的公平客观；
- 3) 宜选择具有焊接领域的专家或经验丰富的人员进行评测，以确保评测结果的准确性和专业性。