

第七届淮安技能状元大赛数据标注 (职工组) 项目技术文件

第七届淮安技能状元大赛组委会技术工作组
2026年10月

目 录

一、技术描述	1
(一) 项目概要	1
(二) 基本知识与能力要求	1
二、试题与评判标准	3
(一) 试题内容	3
(二) 评判标准	3
三、竞赛细则	7
(一) 裁判员分组和职责	7
(二) 赛场管理制度	8
(三) 技术违规处理	10
四、竞赛场地、设施设备等安排	11
(一) 赛场规格要求	11
(二) 场地布局	12
(三) 基础设施清单	12
五、安全要求	13
(一) 赛场人员安全要求	14
(二) 场地设备安全要求	14
六、附件：样题	16

一、技术描述

（一）项目概要

本赛项以人力资源社会保障部发布的《人工智能训练师国家职业技能标准》和“人工智能训练师”新职业的能力要求为依据，面向人工智能数据标注、人工智能应用开发等相关岗位从业人员。重点考察选手对图像、文本和语音数据的理解、标注、检查及成果交付能力，重点评价标注准确性、完整性、一致性和规范性。

（二）基本知识与能力要求

本赛项是对数据标注项目相关技能的展示与评判，选手需掌握操作必备的理论知识，具有相应的知识水平。该项目不涉及理论考试，只进行实际操作竞赛。参加数据标注竞赛的选手，应具备的基本知识和工作能力如下表：

相关要求		权重比例 (%)
1	职业素养	5%
基本知识	- 掌握用电安全、设备与人身安全基础规范，识别安全隐患 - 了解人工智能伦理、算法公平性，知晓标注偏见的危害 - 熟悉数据隐私保护要求，严禁竞赛数据私自拷贝、外传 - 掌握赛场工位管理、垃圾清理等现场文明作业规定	
工作能力	- 作业前后排查隐患，规避人身与设备风险 - 规范操作系统，防止标注文件丢失、泄露 - 标注保持客观中立，减少主观偏见 - 保持工位整洁，及时清理垃圾余料	
2	图像标注	30%
基本知识	- 掌握图像尺寸、像素坐标及类别编号等基础知识 - 掌握目标检测边界框的表示方法及标注规范 - 掌握语义分割的像素级类别标注方法 - 掌握 IoU、mIoU 及 F1 等指标的含义 - 熟悉图像标注成果的导出及格式检查方法	

工作能力	- 能依据任务书和标注类别配置标注任务，正确导入数据并识别无效样本 - 能准确绘制边界框并赋予类别，规范处理遮挡和截断目标，避免漏标、多标及重复标注 - 能准确勾画语义区域并生成类别掩膜，规范处理复杂边界、背景和忽略区域，保持原图尺寸、坐标与类别值一致 - 能检查并修正类别、位置及区域标注错误，按指定格式导出图像标注成果	
3	文本标注	
基本知识	- 掌握命名实体识别的基本概念、实体类别及标签定义 - 掌握实体起止边界和字符偏移的表示方法 - 掌握空格、标点、数字和特殊字符的处理要求 - 掌握精确率、召回率及 F1 的含义和计算方法 - 理解实体类型与完整边界一致匹配的评价要求	
工作能力	- 能准确理解原文，依据标签表识别应标实体，不推断或补写文本中无法确认的信息 - 能准确确定实体类型和起止位置，按统一规范处理歧义及特殊情况，保持标签和边界一致 - 能保持样本 ID、原文、实体编号、类型和字符区间对应，检查并修正漏标、错标、重复标注和偏移错位 - 能按指定格式导出文本实体识别成果	30%
4	语音标注	
基本知识	- 掌握采样率、声道、时长、波形和时间戳等基础知识 - 掌握逐字转写规范 - 掌握按停顿和说话轮次切分语音的方法 - 掌握说话人区分及编号规则 - 掌握 CER、cpCER 和 DER 的基本含义	
工作能力	- 能借助回放和波形听辨语音，准确转写实际内容，不根据上下文补写或擅自改写口语表达 - 能合理切分语音并准确设置起止时间，按要求保留时间精度，避免漏段、重复片段及时间越界 - 能正确区分说话人并保持编号一致，规范处理交替发言、重叠语音、静音和不可辨认片段 - 能保证语音 ID、文字、时间和说话人对应，结合复听及 CER、DER、片段 F1 等反馈修正错误，按模板导出完整有效的成果	35%
合计		100%

二、试题与评判标准

（一）试题内容

1.基本内容

依据《人工智能训练师国家职业技能标准》及数据标注相关岗位能力要求，参赛选手需在规定时间内，根据赛场提供的任务书、原始数据和标注要求，以现场上机操作方式完成下列四个模块。

模块 A：职业素养（贯穿全程，独立评价）

模块 B：图像标注

模块 C：文本标注

模块 D：语音标注

2.样题

数据标注赛项遵循公平、公正原则，命题流程借鉴第八届江苏技能状元大赛人工智能技术应用项目的命题方式，样题随技术文件公布。

（二）评判标准

1.竞赛试题配分

（1）竞赛方式

本赛项采用个人赛的比赛模式，通过阅读赛场提供的任务书（赛题）明确竞赛内容，完成数据标注赛项任务书中所规定的操作内容。

（2）竞赛内容

依据以上要求，参赛选手需按要求完成以下 4 项任务。

模块 A：职业素养

考核选手健康安全法规、人工智能伦理原则、算法公平性、隐私保护等基本知识的理解和应用能力，主要考核是否正确用电、保护数据与系统，能预判并规避安全风险。安全包括设备安全和人身安全，发生事故将按评分细则扣分；卫生包括竞赛工位场地和墙面的清洁，存在垃圾、余料、破损、污染将按评分细则扣分。

模块 B：图像标注

选手依据赛题给定标注要求，依次完成目标检测、语义分割两项独立图像标注子任务。

子任务 1-目标检测：按照规范完成目标检测矩形框标注、类别属性录入，定位图像内指定目标对象；本任务完成后开展自检，修正漏标、错标、类别混淆、框边界漂移等标注缺陷，按赛题要求导出规定格式标注文件，并按要求进行存储。

子任务 2-语义分割：对另一组独立图像素材开展像素级掩码标注，区分前景目标与背景；本任务完成后开展自检，修正掩码边缘失真、区域漏绘等标注缺陷，按赛题要求导出规定格式标注文件，并按要求进行存储。

考核选手图像理解能力、标注规范执行能力、标注质量自查能力与数据集工程基础能力。

模块 C：文本标注

选手依据赛题给定实体标注要求，开展命名实体识别（NER）标注，识别并标记文本中指定类型实体，完成实体边界定位、实体类别归类，处理歧义、简称等复杂文本场景。标注完成后开展自检，

排查实体漏标、边界切分错误、实体类别错分、冗余标注等问题；按赛题要求导出规定格式标注文件，并按要求进行存储。考核选手自然语言理解能力、实体边界判别能力、规范落地能力与文本数据质量管控能力。

模块 D：语音标注

选手依据赛题语音标注要求，对给定语音片段开展语音转写、多说话人语音分段及说话人身份标注。选手将语音内容转写为规范文本，按照规范处理数字、专有名词，完成后自检文字错字、漏字、专有名词不规范等问题；基于语音波形标定每段语音的起止时间戳，完成后自检时间戳偏移、分段错位、重叠语音处理不当等缺陷；在已完成分段的语音片段上标记对应的说话人编号，区分重叠语音、插话场景，完成后自检说话人编号混淆、身份标记错误等问题。全部标注工作完成后，按赛题要求导出带时间轴、说话人标签、转写文本的标注文件，并按要求进行存储。考核选手听觉文本转写能力、时间轴标定能力、多说话人区分能力与语音数据集质量管控能力。

分数权重，详见表 2-1。

表 2-1 竞赛模块分数权重表

模块 编号	模块名称	配分			评价方式
		评价分	测量分	总计	
A	职业素养	5	0	5	评价
B	图像标注	0	30	30	测量
C	文本标注	0	30	30	测量
D	语音标注	0	35	35	测量
总计		5	95	100	

2.成绩计算方式

本项目采用 100 分制,各个评分项的分数精确到小数点后 2 位,各任务原始成绩直接相加即为参赛选手最终竞赛成绩。

3.评判方法

现场裁判依据评分表对参赛选手的操作规范、合理性以及数据标注质量、现场表现等进行评分。

（1）职业素养评分

该评分项由不少于 2 名裁判为一组。裁判各自单独评分,再除以 2 计算实际得分,分数精确到小数点后 2 位,四舍五入。该模块总分 5 分。

（2）图像标注评分

目标检测：在单张图像内,按类别将选手框与参考真值框进行一对一匹配,主要考查类别准确性、框的位置贴合度、漏标和多标等情况。基于类别正确性,统计每一类的精确率、召回率,计算各类别 F1 分数,并结合 IoU 综合计算本任务总分。该子任务总分 10 分。

语义分割：在单张图像内,读取选手导出的标注文件,将选手标注区域与参考真值区域进行像素级比对,主要考查区域类别归属准确性、分割轮廓贴合度、前景漏标和多余前景标注等情况,采用 IoU 综合计算本任务总分。该子任务总分 20 分。

（3）文本标注评分

读取选手导出的标注文件,将导出文件与标准答案进行比对,

主要考查实体类型及位置的准确性，以及漏标、多标等情况。采用逐题实体级 F1 计算成绩，该模块总分 30 分。

（4）语音标注评分

短语音文字转写：将选手转写文本与标准答案逐题比对，主要考查错字、漏字和多字等情况。采用字符错误率 CER 评价转写准确性。该子任务总分 30 分。

多人长语音标注：将选手标注结果与标准答案进行比对，主要考查文字转写及说话人归属的准确性、说话人区分的一致性，以及语音切分和时间边界的准确性。该子任务总分 5 分。

4.成绩并列

按比赛总成绩从高到低排列参赛人员的名次。如遇比赛总成绩相同则由模块 D 成绩高低进行排序，如果成绩还是相同，则依次由模块 C、模块 B、模块 A 成绩同理决定排名。如果仍然相同则按照全部竞赛模块完成的时间短者为优进行判定。

三、竞赛细则

本赛项竞赛总时长 180 分钟。参赛选手在规定时间内，根据赛场提供的赛项任务书（即赛题）和有关资料，以现场上机操作的方式完成全部赛项任务。图像、文本和语音三个模块均在该竞赛时段内完成，选手自主分配各模块操作时间。

（一）裁判员分组和职责

本次竞赛设立裁判组，裁判组由裁判长 1 名和若干裁判员组成。裁判长负责组织全体裁判员做好赛前技术准备及竞赛各环节的技术

工作，组织本项目开展技术总结和技术点评。裁判组接受竞赛组委
会的领导。

1.裁判长

裁判长按照本项目技术文件，对裁判员进行培训和工作分工，
带领裁判员对本项目比赛设备设施和现场布置情况进行检验；组织
选手进行安全培训并熟悉赛场及设备，保障所有选手在比赛前掌握
必备的安全知识和安全操作规范；比赛期间组织裁判员执裁，并按
照相关要求和程序，处理项目内出现的问题；组织统计、汇总并及
时录入大赛成绩等工作；赛后组织开展技术点评。裁判长应公平公
正组织执裁工作，不参与评分。

2.裁判员

全部裁判工作均采取回避制度，裁判员不对来自同一单位参赛
选手进行评判。裁判员应服从裁判长工作安排，认真做好本职工作；
熟练掌握竞赛技术规则，参加赛前培训和技术讨论；对有争议的问
题提出客观、公正、合理的意见和建议；公平公正执裁，不徇私舞
弊；坚守岗位，严格遵守执裁时间安排，保证执裁工作正常进行。

3.工作人员

包括技术支持人员、录分员及赛务保障人员等。按照大赛统一
要求，在裁判长领导下做好相应的竞赛保障工作。

（二）赛场管理制度

1.所有参观人员的活动必须在参观通道内，不得进入竞赛区域；

2.现场保持安静，不得大声交谈及喧哗；

3.现场参观允许拍照，严禁使用闪光灯，赛场内部禁止拍照（若需拍照由裁判长指定人员进行）；

4.在比赛前选手可以在工位内准备自己物品和工具，在裁判宣布开始前禁止触碰竞赛设备或开启电源，否则做扣分处理；

5.竞赛期间选手禁止携带拍照、存储及通信设备，如带到赛场，需要交给本单位场外人员保管或由赛场工作人员集中保管；

6.选手必须在任务区内对题目进行仔细审核，如有问题及时向现场裁判反映，由裁判长决定是否修改或调整题目，如有修改必须对所有参赛队公示说明；

7.选手上交的电子文档由工作人员用赛场指定 U 盘进行拷贝传递或指定网络上传，比赛成果由工作人员打印并由选手确认签字；

8.各参赛单位场外人员在竞赛过程中严禁与任何选手交谈或作出任何提示、影响、干扰行为，如被发现将相应扣除选手个人成绩；

9.竞赛期间，选手需要通过提示牌与现场裁判进行应答或举手交流，本单位裁判需要回避，由其他单位裁判员前去处理；

10.比赛期间，本单位的裁判与选手禁止一切的交流形式；

11.场内现场裁判执裁过程中，除选手示意禁止主动进入选手工位内，如需要裁判进入工位必须 2 名以上非选手单位裁判同时前往处理；

12.选手如怀疑设备问题，可向裁判示意，并选择两种处理方式：1 是技术工作人员检查设备时同时工作，不予补时，2 是离开工位让技术工作人员检查设备，如是设备问题给予相应补时，如设备无恙

则不予补时；

13.严禁在竞赛过程中向赛场内传递任何物品，如有需要必须经过现场裁判确认后由裁判转交；

14.竞赛现场发布的试卷禁止带出场外，竞赛结束后由现场裁判统一收回存档；

15.竞赛现场任何位置严禁吸烟；

16.其他未尽事宜，参照省技能大赛相关标准要求。

（三）技术违规处理

1.不得携带其他未经组委会认可的设备、工具、机具、材料等参赛，不听劝告的取消比赛资格。

2.竞赛过程中，选手不得接受场外送进的材料、加工过的半成品等。

3.选手不得损坏、拆卸、改装赛场提供的设备、工具和工作台等设施。

4.选手不得在任何竞赛区域、位置、赛件上作任何涉嫌作弊的标记。如比赛开始前发现有明显痕迹，可上报裁判员进行处理，严重者可按作弊处理。

5.在完成竞赛任务的过程中，有不符合职业规范的行为或因操作不当导致事故，扣 10-20 分，情况严重者取消比赛资格。

6.因违规操作损坏赛场提供的设备、污染赛场环境等不符合职业规范的行为，视情节扣 5-10 分。

7.扰乱赛场秩序，干扰裁判员工作，视情节扣 5-10 分，情况严

重者取消比赛资格。

四、竞赛场地、设施设备等安排

（一）赛场规格要求

1. 赛场整体规划

赛场设置相互独立的选手工位，标明工位号，必要时设置隔板，保证选手正常操作并便于裁判巡视。根据赛项流程设置选手集合报到区、选手休息区、技术支持休息区、赛事办公工作（储物）区、录分室、裁判休息交流等区域（如有需要，可再设其他空间）。

竞赛区与观摩通道隔离，服务器、参考答案及评分环境置于受控区域。场地应具有稳定供电、良好通风和照明条件，线缆固定布置，不得影响通行和疏散。

2. 竞赛工位规划

每个工位提供竞赛计算机 1 套、桌椅 1 套、有线耳机 1 副，桌面满足标注操作需要。计算机的软件版本、显示条件、网络权限和数据目录统一配置；不得因工位不同造成标注条件差异。

供电采用安全接地和必要的过载、漏电保护措施。服务器和核心网络设备配备必要的备用电源，赛场预留同配置备用计算机及有线耳机，用于故障替换。

3. 场地消防和逃生要求

（1）赛场必须留有安全通道。竞赛前必须明确告知选手和裁判员安全通道和安全门位置。

（2）赛场必须配备灭火设备，并置于显著位置。

(3) 赛场应具备良好的通风、照明和操作空间的条件。

(4) 承办单位应做好竞赛安全、健康和公共卫生及突发事件预防与应急处理等工作。

(二) 场地布局

实际工位数量、门窗及安全出口位置以承办场地为准，布置时不得占用疏散空间。

公众观摩在外围通道进行，不得围观工位或拍摄屏幕。裁判评分区与选手操作区分开管理，评分资料不得向无关人员展示。

(三) 基础设施清单

1. 赛场提供的设备工具清单

赛场需提供的设施、设备清单见表 4-1。

表 4-1 赛场提供的设备工具清单表

序号	名称	规格或要求	单位	数量
1	竞赛计算机	CPU: i5 及以上; 内存: 8GB 及以上; 显示器: ≥22 英寸; 硬盘: ≥500GB;	套	20
2	有线耳机	能够清晰播放语音	套	20
3	操作系统	Win 10/11 x64	套	20
4	标注工具 (图像、文本、语音)	Label Studio v1.23.2	套	20

2. 赛场辅助设施

根据竞赛需要，赛场还需准备的辅助设施见表 4-2。

表 4-2 赛场提供辅助设备表

序号	名称	规格	数量	备注
1	音响及扩音器	能涵盖整个赛场	1 套	

序号	名称	规格	数量	备注
2	无线麦克风		2 个	与音响配套
3	赛场时钟	具有时/分/秒计时	1 套	赛场都可见
4	打印机		1 台	
5	打印纸	A4	1 包	
6	签字笔	红、黑	若干	
7	订书机及钉		1 套	
8	评分夹		若干	
9	灭火器		2 个	
10	U 盘		1 个	
11	计算器		1 个	
12	信封		20 个	用于存放手机

3.竞赛用耗材

竞赛用耗材根据竞赛需要，赛场统一提供。

4.参赛选手禁止使用的物品和材料

选手禁止携带的设备和材料，见表 4-3 所示，违规者不得参赛。

表 4-3 参赛选手禁止使用物品和材料清单表

序号	名称
1	存储设备，如 U 盘、移动硬盘、录音笔等；电子设备，如平板、手机、多媒体播放器、录音器，照相机，摄影机等
2	带有身份标示的物品
3	防锈清洗剂、酒精、汽油、有毒有害物、易燃易爆物
4	气动工具、特制工具

五、安全要求

根据国家相关法规要求，结合本项目实际，提出安全、健康要求及职业操作规范要求，并明确违反后的处理规定。特别是根据本项目具体情况的诸如人身防护，有毒、有害物品携带、存放，防火、防爆等措施。

（一）赛场人员安全要求

以参赛选手为重点，说明进入竞赛区和非竞赛区等竞赛场地的各类人员需进行哪些检测、所需的注意事项（如废弃物不能随意丢、不能使用明火等）、赛场文明要求（竞赛场地禁止吸烟、不能携带手机、照相机等）、所带物品安全检测以及参观人员和宣传人员的安全要求（不能进入竞赛区等）。

1. 现场裁判、选手、工作人员在竞赛期间应该遵守组委会和执委会的安全规定和要求。

2. 参赛选手进入竞赛场地后，须听从并尊重裁判人员的管理，文明参赛。

3. 参赛选手必须在确保人身安全和设备安全的前提下开始竞赛，发现或发生有关安全问题，应立即向裁判报告。

4. 参赛选手必须按照主办地的安全标准要求，配备个人防护用品，包括工作服、绝缘防砸鞋。

5. 参赛选手在本竞赛工位内操作，不得影响其他选手操作。

6. 未经许可，不得进入标有警告标示的危险区。

（二）场地设备安全要求

场地设备安全要求包括设施设备安全操作要求、赛场消防安全要求、安全标识张贴要求、设备安全操作规程。

1. 设施设备安全操作要求

（1）禁止选手及所有参加赛事的人员携带任何有毒有害物品进入竞赛现场。

（2）承办单位应设置专门的安全防卫组，负责竞赛期间健康和
安全事务。主要包括检查竞赛场地、与会人员居住地、车辆交通及
其周围环境的安全防卫；制定紧急应对方案；监督与会人员食品安
全与卫生；分析和处理安全突发事件等工作。

（3）赛场须配备相应医疗人员和急救人员，并备有相应急救设
施。

2.消防与应急保障要求

消防设施、器材和消防安全标志全都在位且功能完整。消防安
全重点部位人员正常在岗工作。

3.安全表示张贴要求

安全出口、疏散通道保证畅通，安全疏散指示标志、应急照明
完好无损，竞赛场地安全疏散通道禁止被占用。

4. 设备安全操作规程

（1）现场电力规格为单相 220V 交流电，安全用电，禁止使用
不符合安全要求的机具，禁止使用连接 220V 电线供电的手电钻，
禁止擅自使用电气设备。

（2）在进行任何安装或维护工作前，必须确认操作对象处于停
止或断电状态。

六、附件：样题

重要说明

- 1.竞赛时间 180 分钟，选手不可以弃赛，不可提前离开赛场。
- 2.比赛共包括 4 个模块，总分 100 分，任务及配分见下表所示。

序号	模块	配分	备注
1	图像标注	30	
2	文本标注	30	
3	语音标注	35	
4	职业素养	5	
合计		100 分	

3.如发现任务书缺页、字迹不清等问题，请及时向裁判申请更换任务书；对照设备清单认真检查设备，如发现问题，请及时向裁判申请处理。

4.竞赛资料存放在计算机桌面“竞赛资料”文件夹下；选手完成数据标注后所导出的文件格式及存储位置，需严格按照赛题要求。

5.选手提交的资料不得出现学校、企业、姓名等与身份有关信息，擅自离开竞赛工位、与其他选手交流、不服从裁判指令，将依据扣分表进行处理。

6.由于操作不当等原因引起设备的损坏，将依据扣分表进行处理。严重损坏比赛设备将取消竞赛资格。

7.在完成任务过程中，请及时保存程序及数据，未能及时保存程序及数据，由于断电等意外情况造成的程序及数据丢失的责任将由选手自负。

8.比赛过程中需裁判确认部分，选手应当先举手示意，等待裁

判前来处理。

9.选手必须认真填写各类文档，竞赛完成后所有文档按页码顺序一并上交；赛场提供的任何物品，不得带离赛场。

模块 A 职业素养

【任务描述】

严格遵循相关职业素养要求及安全、伦理与数据规范，安全文明参赛；着装规范，标注操作规范；软硬件工具、外设摆放整齐；标注工程文件与资料归档完整。恪守数据隐私保护要求，杜绝数据泄露。

模块 B 图像标注

【任务描述】

本模块包括目标检测和语义分割两项子任务，满分 30 分。

（1）目标检测：数据源自公开数据集 SoccerDB，为足球比赛场景图像（包含人、球及球门）。

（2）语义分割：数据源自公开数据集 PASCAL VOC 2012，为包含动物目标的自然场景图像。

【标注要求】

（1）目标检测：

① 标注类别为“person”、“ball”和“goal”，分别对应人、球和球门。其中，“person”包括球场上出现的球员、守门员和裁判员。

② 对规定类别的目标逐一进行矩形框标注，并选择对应标签。标注应准确、完整，避免漏标、多标、重复标注和类别错误。

③ 标准目标数量、类别及位置以数据集配套标签为依据。

④ 标注工具：Label Studio v1.23.2。参赛资料中包含 Labeling Interface 标注界面 XML 编码，选手自行复制编码至标注工具代码视图，即可完成标注界面配置。

（2）语义分割

① 标注类别统一为“animal”，规定范围内的动物目标均使用该标签，不再细分动物种类。

② 沿动物目标的实际轮廓勾画前景区域。标注应准确、完整，避免前景漏标、多余标注及将背景区域误标为前景。

③ 标准前景区域及其边界以数据集配套标签为依据。

④ 标注工具：Label Studio v1.23.2。参赛资料中包含 Labeling Interface 标注界面 XML 编码，选手自行复制编码至标注工具代码视图，即可完成标注界面配置。

【评分】

（1）目标检测：将选手标注框与标准框进行比对。类别相同且 IoU 不低于 0.50 时，允许进行一对一匹配，每个选手框和每个标准框均只能匹配一次。将选手标注框与同一张图像中的标准框按类别进行一对一匹配，主要考查类别准确性、框的位置贴合度，以及漏标、多标和重复标注等情况。按全部考核图像累计统计各类别的 F1 值和平均 IoU，并进行加权融合。融合得分达到 0.90 及以上得满分，低于 0.90 时按比例折算得分。

（2）语义分割：将选手标注区域与标准前景区域进行像素级比

对，主要考查前景区域的完整性、分割轮廓贴合度，以及漏标、多标等情况。同一图像内的多个前景区域合并比较，重叠部分不重复计算。累计全部考核图像的前景交集与并集像素，计算整体 IoU。整体 IoU 达到 0.98 及以上得满分，低于 0.98 时按比例折算得分。

模块 C 文本标注

【任务描述】

数据源自公开数据集 MSRA NER，为包含人名（PER）、地名（LOC）和机构/组织名（ORG）的中文文本。

【标注要求】

（1）标注类别为“PER”、“LOC”和“ORG”。

（2）对规定类别的实体逐一选取完整文本范围，并选择对应标签。标注应准确、完整，避免漏标、多标、重复标注、类别错误和边界错误。

（3）标准实体数量、类型及起止位置以数据集配套标签为依据。不得增删或改写原文，实体位置按原文字符确定，空格和标点均计入位置。

（4）标注工具：Label Studio v1.23.2。参赛资料中包含 Labeling Interface 标注界面 XML 编码，选手自行复制编码至标注工具代码视图，即可完成标注界面配置。

【评分】

将选手标注实体与标准实体进行一对一对比，主要考查实体类型及位置的准确性，以及漏标、多标和重复标注等情况。实体类型

与起止位置均一致时，判定为正确匹配。采用实体级 F1 评价各题标注质量，按单题分值折算后汇总。

模块 D 语音标注

【任务描述】

本模块包括短语音文字转写和多人长语音标注两项任务。

（1）短语音文字转写：数据源自公开数据集 AISHELL-1，为中文朗读语音。选手需对提供的 20 段短语音完成文字转写，准确记录实际语音内容。

（2）多人长语音标注：数据源自公开数据集 AliMeeting，为包含多人交谈及重叠讲话的中文会议语音。选手需对提供的 1 段多人长语音完成标注，准确记录讲话内容，区分说话人，并标出各段讲话的起止时间。

【标注要求】

（1）短语音：逐题听取录音，完整转写实际语音内容。

（2）多人长语音：标出每段有效讲话的开始时间、结束时间、说话人编号及完整原话。同一说话人使用一致编号；不同说话人同时讲话时，应分别标注，允许时间区间重叠。

（3）按实际听到的内容转写，保留可辨认的语气词、重复和口误，不概括、不润色。数字统一使用中文书写。不考核标点准确性。

（4）其他说话人插话时，不强制截断原说话人的连续片段；同一说话人不得重复框选重叠区间。片段外边界应对准实际发声起止，不包含前后静音；纯静音、纯背景声不建立文字区间。

(5) 完成后检查文字、人物归属及时间边界，确保提交结果与对应录音一致。

(6) 标准说话内容、说话人、起止时间以数据集配套标签为依据。

(7) 标注工具：Label Studio v1.23.2。参赛资料中包含 Labeling Interface 标注界面 XML 编码，选手自行复制编码至标注工具代码视图，即可完成标注界面配置。

【评分】

(1) 短语音文字转写，采用字符错误率 CER 评价错字、漏字和多字。各题按文字正确程度分别折分后累加；字符错误率达到 100% 时，该题计 0 分。

(2) 多人长语音标注，分项评价如下：

① 文字转写与说话人归属，2.5 分。采用 cpCER 指标，综合评价转写文字的准确性及其所属说话人的正确性。错字、漏字、多字及说话人归属错误均影响得分。说话人编号不要求与参考答案相同，但同一人的编号应前后一致。

② 说话人分离，1.5 分。采用说话人分离错误率 DER，评价各时间段内讲话人员的标注是否正确，主要考查漏标讲话、将静音误标为讲话及说话人混淆等情况，多人同时讲话的情况也纳入评价。

③ 语音切分与时间边界，1 分。采用片段 F1，评价讲话片段的切分及起止时间是否准确。起点和终点与标准答案的误差均不超过 0.20 秒时，可判定为时间匹配成功，每个片段仅匹配一次。漏标、

多标、重复标注及时间边界偏差均影响得分。

（3）文字评分忽略空白和标点，数字应按标注要求使用中文书写。各分项根据评价结果折算得分，最低为 0 分。

（4）缺失题目按漏答处理。