

2025中国机器人大赛暨 RoboCup 机器人世界杯中国赛

2025 年度赛事规则

（总决赛）

赛项：智源具身大模型挑战赛

智源具身大模型挑战赛技术委员会

2025 年 9 月

目录

一、项目简介.....	3
二、技术委员会与组织委员会.....	3
2.1 技术委员会	3
2.2 组织委员会	4
三、资格认证要求.....	4
3.1 报名信息提交.....	4
3.2 报名须知	4
四、参赛人员要求.....	5
五、技术与竞赛组织讨论群	5
六、赛事规则要求.....	5
6.1 赛事规则概述.....	5
6.2 赛事规则详情	6
6.3 赛事资源支持	7
七、赛事考察技术难点.....	7
八、赛题详情要求.....	7
8.1 模型训练赛题：	7
8.2 答辩赛题：	8
九、评分标准.....	9
9.1 战队总成绩	9
9.2 模型提交及评测排行榜评分方法	9
9.3 代码提交	10
9.4 技术方案提交要求及答辩评分方法	10
十、赛程赛制.....	11

一、项目简介

具身智能（Embodied Artificial Intelligence）即通过机器人等物理实体与环境交互，能进行环境感知、信息认知、自主决策和采取行动，并能够从经验反馈中实现智能增长和行动自适应的智能系统。具身大脑大模型是支撑具身智能体实现物理世界交互的核心智能引擎。它以大模型为技术基座，融合多模态感知、空间推理、动态规划等能力，能处理来自视觉、触觉等传感器的环境数据，实现场景理解、指令解析与自主决策，同时通过交互反馈持续优化认知与行动策略，是具身智能体“感知 - 思考 - 行动 - 进化”闭环的核心中枢。

当前我国正处于具身智能发展的热潮之中，关于具身智能机器人的讨论热度持续高涨。在此背景下，由北京智源人工智能研究院在中国机器人大赛暨 RoboCup 机器人世界杯中国赛中设立“具身大模型能力挑战赛”。赛事依托北京智源人工智能研究院开源的 RoboBrain2.0-7B 大模型，聚焦具身智能核心能力考核，全面覆盖指令规划、感知理解及可供性分析三大关键维度和若干子维度展开，深度考察模型在真实物理交互场景中的多任务处理能力——从精准解析复杂自然语言指令，到分步规划可行执行方案，再到结合环境感知进行逻辑推理，以及对任务物体与环境动态与静态可供性的分析能力。赛事采用国际权威 Benchmark 评测体系，通过标准化数据集验证模型性能，最终生成多维度能力排名，为具身智能技术发展提供权威参考，助力推动大模型在机器人交互、自主决策等领域的实用化突破。

二、技术委员会与组织委员会

2.1 技术委员会

后续更新

2.2 组织委员会

后续更新

三、资格认证要求

3.1 报名信息提交

参赛队伍需提交资料供技术委员会进行审核，包含以下主要内容：

- (1) 队伍名称
- (2) 机构或学校名称
- (3) 队伍其他成员（老师和领队）
- (4) 联系电话
- (5) 联系邮箱
- (6) 是否自带算力

3.2 报名须知

- 报名成功的选手需根据组委会要求，签署相关保密责任书、知识产权说明文件、数据安全承诺书等文件
- 10.17-10.19 期间，每个队伍每天至少 1 人在线下现场参加。答辩至少1人线下参加。如不符合本要求，将承担所在队伍被回收已发放资源、取消成绩的后果
- 报名成功后将进行组委会审核。审核通过的参赛队伍获得本次比赛的参赛资格
- 本次竞赛名额有限，先到先得
- 按照报名顺序进行相关资源分配和账号开通

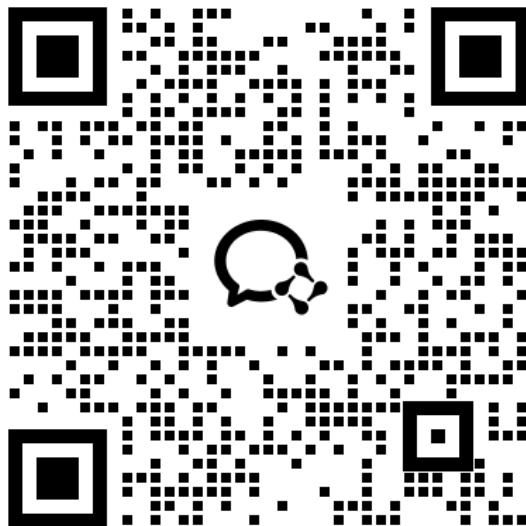
四、参赛人员要求

- 本次大赛欢迎国内外企业、高校、科研院所、创新团队、科研开发者等人员参加
- 每一个参赛队伍可以报名参加多个赛项
- 每一个参赛队的人数应不多于 5 人，支持 1 人组队
- 10.17-10.19 期间，每个队伍每天至少 1 人在线下现场参加。答辩至少1人线下参加。如不符合本要求，将承担所在队伍被回收已发放资源、取消成绩的后果

五、技术与竞赛组织讨论群

参赛队员与指导老师需加入智源具身大模型挑战赛项目微信群。群内实行实名管理（名片格式：院校-教师 or 学生-姓名，群内定期清理非竞赛相关人员）。请求加群时，需要注明参赛队伍，否则可能不能入群。

本群可以进行学术讨论。同时，竞赛相关情况将在群内同步。



六、赛事规则要求

6.1 赛事规则概述

智源具身大模型挑战赛基于智源具身大脑系列大模型围绕具身场景任务需要的能力展开，参赛者选用比赛要求的模型、建议的数据集（部署在竞赛算力平台）或自建数据集训练模型，在评测集上完成评测，最终依据

评测结果和方案答辩综合评估排名。

6.2 赛事规则详情

以下是赛事的整体安排与要求（个别时间可能会根据赛事情况进行微调）：

6.2.1 报名与赛前备赛（线上）：即日起至 10 月 8 日

- 参赛队伍按照竞赛要求进行报名，提供相应信息
- 参赛队伍获取组委会提供的数据、模型以及参考训练脚本后，可以开展数据收集、数据处理、模型训练、日志与模型备份以及答辩设计等任务准备
- 备赛期间开放线上技术答疑通道，用来解答模型使用相关问题

6.2.2 数据准备（线上）：10 月 10 日-10 月 11 日

参赛队伍将前期自行准备的数据上传至组委会指定的存储空间，以便竞赛中使用。

6.2.3 模型训练与结果提交（线上+线下）：10 月 12 日 - 10 月 17 日

模型训练阶段主要包括四个任务：

- 算力环境检查：核查基础软件环境完整性与适配性，测试 GPU 功能，再按需补充库、配置虚拟环境与优化参数
- 数据准备：上传自有数据，处理自有数据且检查标注，整合数据并构建加载器
- 训练实施：加载验证基础模型，制定超参数、训练轮次等策略，编写脚本执行训练并监控，存档日志与模型检查点
- 结果整理：校验管理模型文件，归档含数据、策略等信息的训练日志并备份，用以完善答辩文档，执行模型提交

选手针对以上 4 个任务进行模型训练。选手 10 月 12 日领取算力即可开始训练。

- 智源研究院针对选手训练的模型进行评测，得到选手初赛排行榜成绩

6.2.4 答辩准备（线下）：10 月 18 日 - 10 月 19 日 9:00

评测工作在智源评测平台统一运行，期间选手无需等待，可同步进行答

辩准备

6.2.5 现场答辩（线下）：10 月 19 日 9:00-10 月 19 日 11:00

选手进行答辩，分别进行方案陈述，技术专家进行打分

6.2.6 公布成绩与颁奖（线下）：10 月 19 日 14:00-16:00

公布选手成绩与名次，进行线下颁奖典礼

6.3 赛事资源支持

组委会提供算力和存储资源支持，选手按组委会要求申请使用相关资源

七、赛事考察技术难点

本赛题对具身大脑在指令规划、感知理解等三大能力维度七个子维度能力进行考察，并设计了“自建数据 - 模型训练 - 推理验证”的流程，对技术考察难点主要聚焦以下几方面：

- 数据适配性与标注精准度难题：需从开源数据集中筛选适配能力维度的样本，而现有数据多缺乏具身场景的空间关系标注、多模态时序关联标注等关键信息，自主标注需平衡标注成本与精度，尤其对预测、规划类任务的“因果逻辑标注”难度极高。
- 多维度能力协同训练矛盾：能力维度需求差异显著，例如感知理解依赖几何特征学习，指令规划侧重逻辑序列建模，单一训练策略易导致能力失衡，如何设计多任务联合训练框架，避免维度间的优化冲突成为关键。
- 模型泛化与评测适配挑战：训练数据与统一评测集存在场景、任务差异，模型易出现“训练 - 评测偏差”，尤其在评测集的场景预测、复杂步骤规划等子维度上，泛化性能难以保障。
- 训练效率与参数优化平衡：指定模型参数量大，需在有限资源下优化训练设置，既要通过参数调整适配多维度任务需求，又要避免过拟合

八、赛题详情要求

8.1 模型训练赛题：

智源具身大脑大模型赛项指定具身大脑大模型，围绕指令规划、感知理

解、可供性分析三大能力维度和七个子维度能力展开，需要选手使用赛事指定数据集或在准备阶段与模型训练阶段下载并处理的数据集在指定模型上进行能力提升，将处理好的模型交予主办方在评测集上完成评测。还需要参赛选手准备方案设计演示，最终根据评测结果和方案设计进行综合评估排名。

- 指令规划包括语义理解与任务规划两个子维度，作为系统的“大脑中枢”，模型需要先通过语义理解解析输入需求，再进行任务规划，明确行动目标与步骤逻辑。
- 感知理解包括视觉识别、环境分析与空间理解三个子维度，具身大脑需要借助视觉识别捕捉物体等视觉信息，通过空间理解和环境分析形成感知的结构认知。
- 可供性分析包括物体功能性理解和场景可供性分析两个子维度，为导航和操纵任务提供“环境交互判断”职能，包括理解物体功能属性、分析场景中可导航点位、物体可放置区域等交互可能性。

模型链接：

- <https://huggingface.co/BAAI/RoboBrain2.0-7B>

数据链接：

- <https://huggingface.co/datasets/BAAI/ShareRobot>
- <https://huggingface.co/datasets/JingkunAn/RefSpatial>

8.2 答辩赛题：

参赛队伍需提交技术方案。技术方案需包含以下内容：

（1）如何收集训练数据

- 如数据选择哪些，旨在提升哪方面的能力
- 数据标注基于哪些数据集进行，prompt 设计，旨在提升哪方面的能力

（2）如何设计训练方案

- 整体实验设计
- 训练设置，是否需要分阶段训练，每个阶段旨在学习什么能力？
- 参数设置

（3）如何验证模型效果？

- 精准度层面，测试模型效果是否符合预期目标
- 泛化性层面，测试模型效果能否泛化到更广泛的层面

答辩时各个战队依据文档展开陈述和答辩，评审团重点考察参赛者对数据与训练方法的理解与设计。

九、评分标准

9.1 战队总成绩

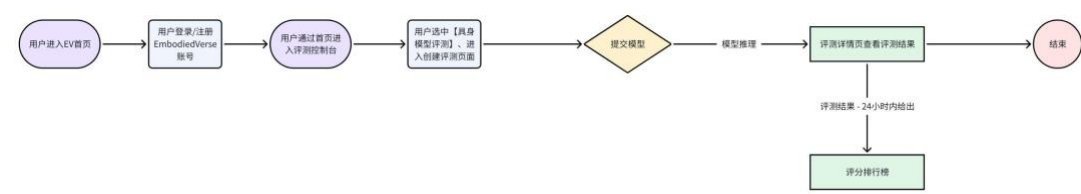
由模型评测排行榜成绩（占比 60%）和答辩成绩（占比 40%）加权得出，组委会将根据总成绩颁发奖项

9.2 模型提交及评测排行榜评分方法

9.2.1 模型提交方法

通过智源 FlagEval-EmbodiedVerse 具身智能评测平台以 API 的形式进行模型的能力评测。参赛选手需要注册 FlagEval-EmbodiedVerse，按照平台要求封装 API 接口，支持评测调用。

模型评估方式流程如下：



9.2.2 模型评测维度

模型评测维度如下（有可能根据参赛情况进行微调，以开赛前确认为准）：

表 1：评测维度

能力维度	子能力维度
指令规划	语义理解
	任务规划
感知理解	视觉识别

	环境分析
	空间理解
	物体功能性理解
	场景可供性分析
可供性分析	

9.3 代码提交

选手需按照组委会的要求统一将代码提交至魔乐社区相应代码仓库

9.4 技术方案提交要求及答辩评分方法

9.4.1 技术方案提交要求

技术方案统一要求为 PDF 格式（不少于 5 页容），以战队名命名。技术方案提交至魔乐社区相应空间。鼓励提交其他补充内容，辅助展示方案可行性

9.4.2 答辩评审方式

每个参赛队伍都要根据组委会要求，依据技术方案展开陈述。组委会评审团将根据方案的完整性、创新性、可行性、实验设计与复现性进行打分。评审团将由智源研究院技术专家组成

9.4.3 答辩评审标准

- 方案完整性（30%）：是否完整覆盖赛题要求的所有核心要点，逻辑是否清晰
- 方案创新性（30%）：数据收集、训练设计等环节是否具备新颖思路或技术亮点
- 方案可行性（20%）：方案在技术实现、资源适配等方面是否具备可落地性
- 实验设计与复现性（20%）：实验设计是否科学合理，步骤是否清晰可复现

十、赛程赛制

赛程安排如下：

- 报名与赛前备赛（线上）：即日起至 10 月 8 日
- 数据准备（线上）：10 月 10 日-10 月 11 日
- 模型训练与结果提交（线上+线下）：10 月 12 日 - 10 月 17 日
- 榜单评测（线下）：10 月 18 日 - 10 月 19 日 11:00
- 答辩准备（线下）：10 月 18 日 - 10 月 19 日 9:00
- 现场答辩（线下）：10 月 19 日 9:00-10 月 19 日 11:00
- 公布成绩与颁奖（线下）：10 月 19 日 14:00-16:00

赛事最终解释权归中国机器人大赛暨 RoboCup 机器人世界杯中国赛智源具身大模型挑战赛所有