

# 面向元宇宙的脑电信号离体生成和检测技术赛题说明

## 一、项目介绍

元宇宙未启，身份认证先行。近年来，基于脑电信号的个体身份识别技术取得了长足进展，成为研究的前沿热点。然而，脑电信号模拟生成和检测技术仍是一片有待开垦的新领域。这不仅是实现元宇宙身份认证的必要一环，更是探索人工智能与生物智能边界的重要课题。随着深度学习的突破性进展，脑电信号模拟生成技术展现出了广阔的前景。但这样的生成技术能在多大程度上逼真地再现人类的脑电特征，实现脱离活体的数据生成，仍需要我们持续的探索和不断的验证。

本赛题基于真实的静息态脑电数据进行模拟生成，要求参赛者在线鉴别数据真伪及识别其中的个体身份信息。比赛目标是探究是否能够模拟生成既高度拟真有具备个体特征的脑电数据，为这一前沿领域提供一个创新研究和技术交流的平台。比赛数据集由真实数据和模拟生成的脑电数据构成，真实数据由深圳大学提供，模拟生成数据由蚂蚁数科、杭州电子科技大学、深圳大学三家单位根据不同技术路线独立生成。需要注意的是，训练和测试数据集在被试者、数据真伪比例以及生成参数方面均有所不同，这对参赛队伍的算法鲁棒性和泛化能力提出更高的挑战。

作为一项前瞻性的探索，本赛题不仅将推动脑电信号离体生成和检测技术的进步，更有望引发人工智能与生物智能交叉领域的思考和讨论。我们诚邀来自全球相关领域的同学，科研人员、工程师加入这场探索未来的比赛，用你们的洞察力和创新能力在这一新兴领域开辟道路！一起探索科技的无限可能，为未来技术的发展共同铺就道路。

## 二、实验范式

赛题实验范式为静息态，每组实验包含睁眼和闭眼数据，如图 1 所示。本次比赛提供包含来自不同受试者的 23 个 Block 的数据作为训练数据，每个 Block 包含 24 个 Trial，每个 Trial 时间长度可能不同，Trial 之间用不同长度的零值间隔，如图 1 所示。

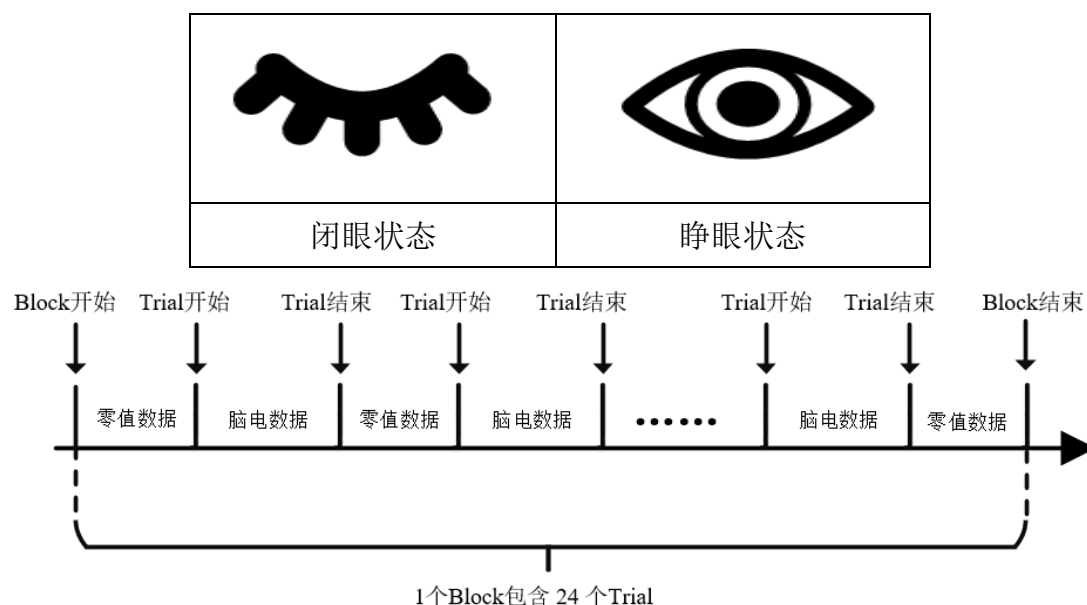


图 1 实验范式

### 三、实验数据

实验数据使用 64 通道 BrainAmp 设备采集，之后降采样至 250Hz，重参考到共同平均参考，并选择通道 Fz, Cz, Oz, C3, C4 的数据，导联序号-名称对照表如表 1 所示。

赛题数据说明如图 2 所示，本赛题赛前会提供训练数据集，cali\_train, cali\_A, cali\_B 仅含有真实脑电数据，三个数据集的被试各不一样；eval\_train 中的被试与 cali\_train 一致，含有真实脑电数据和模拟脑电数据。eval\_A 和 eval\_B 是比赛数据集，正式比赛时，参赛选手需根据不同数据汇报结果，结果包含预测数据的真假和数据的被试编号信息。

本赛题赛前提供模型校准数据集（带 cali 前缀的数据集），该数据集中的真实数据由 23 个被试的不同时间长度的脑电信号拼接而成，即每一个 Block 中的各个 Trial 的脑电数据的被试信息可能不一致，前一个 Trial 脑电数据来自 1 号被试，后一个 Trial 的脑电数据可能来自另一个被试。eval\_train 数据集由 23 个被试的不同时间长度的真实脑电数据和模拟脑电数据拼接而成，即每一个 Block 中的各个 Trial 的脑电数据的真假和被试信息可能不一致，前一个 Trial 脑电数据来自 3 号被试的真实脑电数据，后一个 Trial 的脑电数据可能来自另一个被试的模拟脑电数据。

**注意：**赛前提提供的模型校准数据集集中的被试（cali\_train, cali\_A, cali\_B）各

不相同，参赛者需使用 `cali_train` 和 `eval_train` 数据集训练好模型后再根据 `cali_A/cali_B` 数据集进行校准。

具体 Trigger 定义如表 2、3 所示。整体数据中共有 73 种 Trigger，包含 Block 开始标 242、Block 结束标 243、Trial 开始标范围在 1-23，129-151，193-215，240 区间的值、Trial 结束标 241。Trigger 在 129-151 分别表示 1-23 被试的真实睁眼脑电数据；Trigger 在 193-215 分别表示 1-23 被试的真实闭眼脑电数据；Trigger 在 1-23 分别表示 1-23 被试的模拟脑电数据。

`eval_train` 数据集包含真实数据和使用不同方法生成的模拟脑电数据，Trigger 包含 Block 开始标 242、Block 结束标 243、Trial 开始标范围在 1-23，129-151，193-215 区间的值、Trial 结束标 241。Trial 开始 Trigger 均为真实信息，包含数据真假和被试信息。如图 3 所示，Trigger 仅在 Trial 开始、Trial 结束、Block 开始、Block 结束有值，其余 Trigger 点均为 0。

在公开数据集中，表示 Trial 开始的八位二进制数 Trigger 前 2 位数表示含义：

10 真实睁眼，11 真实闭眼，00 生成的模拟数据。八位二进制数的后 5 位数表示含义：00001- 10111 表示 1-23 被试编号。例如，Trigger 为 129 表示 1 号被试的真实睁眼脑电数据，Trigger 为 22 表示 22 号被试的模拟脑电数据。

比赛数据集包含真实数据和使用不同方法生成的模拟脑电数据，Trigger 包含 Block 开始标 242、Block 结束标 243、Trial 开始标 240、Trial 结束标 241。

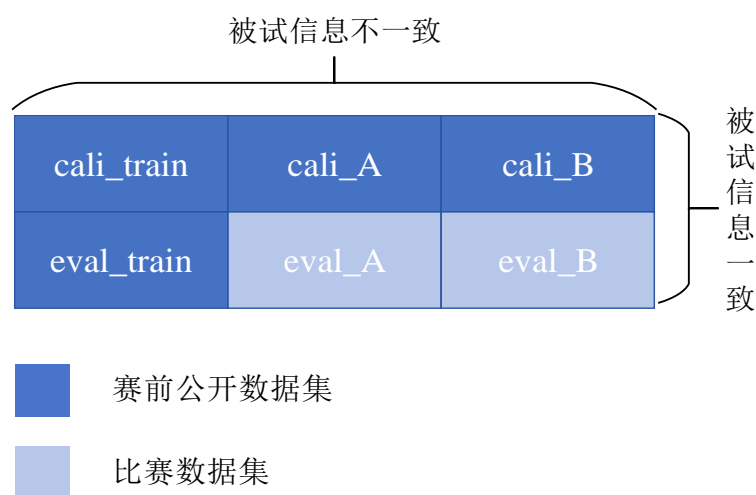


图 2 赛题数据说明

表 1 导联序号-名称对照表

导联序号	1	2	3	4	5
导联名称	Fz	Cz	Oz	C3	C4

表 2 整体数据 Trigger 定义

定义	Trial 开始	Trial 结束	Block 开始	Block 结束
Trigger 号	1-23, 129-151, 193-215, 240	241	242	243

表 3 公开数据 Trigger 定义

Trigger 号	定义	Trigger 二进制
129-151	1-23 被试的真实睁眼脑电数据	10000001-10010111
193-215	1-23 被试的真实闭眼脑电数据	11000001-11010111
1-23	1-23 被试的模拟脑电数据	00000001-00010111

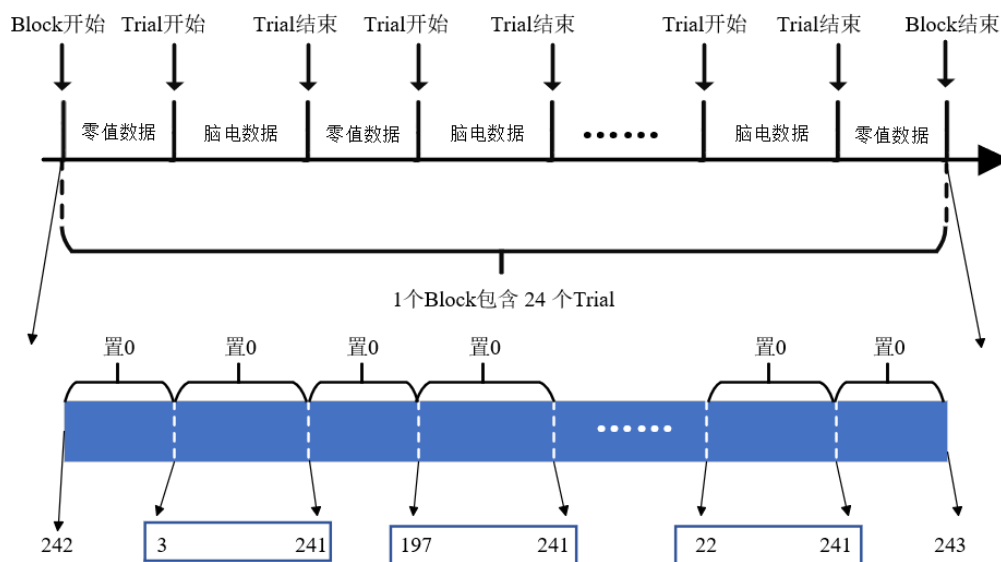


图 3 训练数据 Trigger 用例

#### 四、算法规范

参赛算法调用数据读取方法获取脑电数据。数据读取方法被调用一次，比赛

系统会返回一个新数据包，参赛算法可以对新数据包进行缓存并处理。当算法认为接收到的数据足以满足判决条件时，需要调用反馈方法向比赛系统报告识别结果。比赛系统根据数据读取方法的调用次数计算出算法使用有效数据长度，并结合反馈正确率，综合计算出平均模拟信息传输速率。

参赛算法需要同时满足以下几个约束条件：

### 1、试次起止约束：

在对单一试次数据的检测识别过程中，参赛算法需要在接收到该试次 Trigger 之后开始检测，并且在接收到成绩提交点 Trigger 前进行反馈报告。否则，报告结果将被判定为无效的识别结果。

### 2、单试次最大数据长度约束：

本项目对于单一试次最长检测时间需小于当前试次结束 Trigger 标签出现时间。从试次的 Trigger 开始信号起，参赛算法最多采集到当前试次的 Trigger 结束信号点的 EEG 数据（包含当前试次的 Trigger 结束信号点的 EEG 数据），否则该试次识别结果将被视为无效，该试次结果为 0。

### 3、算法终止约束：

当接收到数据包中 Endflag = 1 时，意味着所有实验数据均已发送完毕，参赛算法需要停止处理并自行退出。

## 五、赛题框架

### 1、参赛者用例

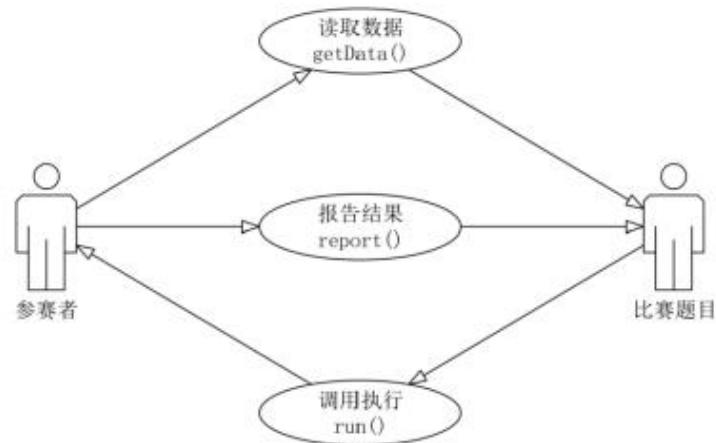


图 4 参赛者用例

### 2、系统主体框架

### (1) ProxyInterface 题目接口

该接口是面向参赛者的赛题接口，主要负责题目与参赛算法之间的数据传递及结果报告。参赛者可以通过该接口获取比赛数据，并通过该接口报告识别结果。比赛题目需要根据参赛算法获取比赛数据的次数，以及报告结果综合给出比赛评分。

### (2) ChallengeInterface 赛题接口

该接口主要负责实现赛题的清洗数据和计算得分，框架通过该接口实现对赛题的调用。

### (3) AlgorithmInterface 算法接口

通过该接口比赛题目可以对参赛算法进行验证计算。该接口由参赛者具体实现。在执行过程中，参赛算法需要通过 **ProblemInterface** 接口获取数据，并且通过该接口报告结果。同时，参赛者需要控制算法的计算复杂度，否则当运行时间超过预定长度时，系统将自动终止该计算进程，所获成绩无效。

## 3、数据模型

### (1) DataModel 参赛者数据模型

1) **data**: float 类型矩阵，分段数据。例如包含有 5 导 EEG 数据+1 导 Trigger 信号，在 250Hz 采样率下，单次获取的 data 为 6\*50 个点。

2) **personID**: int 类型标量，当前数据来源序号，注意该数据在本赛题中属于无效信号。

3) **blockID**: int 类型标量，当前数据来源 Block 序号。

4) **srate**: int 类型标量，数据采样率。

5) **ch\_names**: list 类型，当前数据通道信息。

6) **nchan**: int 类型标量，当前数据通道数量。

7) **dimensions**: tuple 类型，当前数据大小。

## 4、参赛者相关接口函数

### (1) ChallengeInterface

该接口由出题方负责实现，包括数据获取方法及结果反馈方法。在算法运行前，该接口的实现类会被注入参赛算法实现类中。算法执行过程中，可以调用该

接口获取数据，并通过结果反馈方法报告识别结果。出题方根据数据获取方法的被调用次数，及结果反馈的正确性进行综合评分。

1) def receive\_message(self, receive\_model: BaseModel):

输入参数: BaseModel

输出参数: DataModel

实现功能: 处理框架发送的数据。

2) def report(self, reportModel):

输入参数: ReportModel

输出参数: ResultModel

实现功能: 处理算法汇报的结果。

## (2) AlgorithmInterface

参赛者需要根据“Algorithm\method\interface\AlgorithmInterface.py”中的 AlgorithmInterface 类实现算法类，并将算法类名称和路径写入“Algorithm\config\AlgorithmConfig.yml”文件中。注意，参赛者需基于已有文件和类开发，不得更改文件名称和类名。

参赛者需在算法类中重写以协程的形式重写 run 方法，并在获取设备信息、获取数据和汇报结果时使用 await 关键字异步执行。框架运行时，会主动调用 run 方法。因此参赛者需要在 run 方法中读取数据，进行计算，并汇报结果。

其中，参赛者可以通过调用 \_proxy 属性的 get\_source 方法，并传入 str 类型的数据源名称参数，获取数据源（SourceInterface 实例）。如果存在多个源，可以依次读取。获取到数据源之后，可以使用 await 关键字异步调用其“get\_device”方法获取设备信息（AlgorithmDeviceObject 实例），调用其“get\_data”方法获取一个数据包（AlgorithmDataObject 实例）。

在需要汇报结果时，参赛者可以实例化一个 AlgorithmResultObject 类型的实例，并将结果赋值给该对象的“result”属性。随后将该实例作为参数传入并调用“\_proxy”属性的 report 方法。当通过 get\_data 获取的数据中 finish\_flag 为 true 时，意味着数据处理完毕，该函数需要自行退出运行。

1) def get\_source(self)

输入参数: 无

输出参数: SourceInterface 对象

实现功能：用于算法从框架获取数据源对象。

2) def get\_data(self):

输入参数：无

输出参数：AlgorithmDataObject 对象

实现功能：用于算法从数据源获取数据。调用时需使用 await 关键字异步调用。

3) def get\_device(self):

输入参数：无

输出参数：AlgorithmDeviceObject 对象

实现功能：用于算法从数据源获取数据采集设备的设备信息。调用时需使用 await 关键字异步调用。

4) def run(self):

输入参数：无

输出参数：无

实现功能：算法分析过程。

5) def report(self, algorithm\_result\_object: AlgorithmResultObject):

输入参数：必须传入且仅传入一个 AlgorithmResultObject 对象

输出参数：无

实现功能：用于算法向框架报告结果。

### 提交格式

参赛者需根据比赛数据汇报预测结果，结果格式参照表 4 所示：

表 4 预测结果格式定义

结果含义	Trigger 二进制
预测为 1-23 被试的真实脑电数据	100001-110111
预测为 1-23 被试的模拟生成数据	000001-010111

预测结果用六位二进制数表示，六位二进制数的首位表示真假数据，六位二进制数的后 5 位数 00001-10111 表示 1-23 被试编号。例如：预测结果为 16 号被试的真实数据，汇报结果应为六位二进制数 110000；预测结果为 6 号被试的模拟生成数据，汇报结果应为六位二进制数 000110。



本赛题程序使用 python 语言编写，需提交基于 python 3.10 版本的扩展名为.pyc 的文件。

## 5、提交样例

参考配套代码。

参赛者可通过修改“Algorithm\method\impl”文件夹中的代码完成算法，为了避免未知错误，请勿在主目录内添加文件夹。完成后重新打包程序（包含“Algorithm\method\impl”文件夹和“AlgorithmConfig.yml”）--> 分组 --> 具体分组 --> 计算单元 --> 定义计算单元 --> 上传程序包 --> 提交到比赛 --> 选择比赛 --> 部署 -->完成比赛。

部署完成后在赛题的排行榜中查看比赛成绩；

需要注意的是，为防止参赛者修改代码框架作弊，保护评分程序会完全覆盖参赛者的代码（除了“Algorithm\method\impl”目录和“AlgorithmConfig.yml”）在提交到比赛 --> 部署时，启动的实际为 评分程序 + 参赛者的“Algorithm\method\impl”目录，其余运行配套代码均为服务器内置程序（包括 main.py 等文件，服务器内置评分程序与范例中程序框架基本相同，但包含评分功能和读取服务器比赛数据功能）。

## 6、评分方式

本赛题拟模拟信息传输速率（information transfer rate, ITR）作为评分标准：

$$ITR = \left( \log_2 M + P \log_2 P + (1 - P) \log_2 \left( \frac{1 - P}{M - 1} \right) \right) \times \left( \frac{60}{T} \right)$$

根据参赛者提交的结果，包括预测数据真假和被试序号，计算二者 ITR，二者加权求和得到最终结果。其中，预测数据真假的 ITR 权重占 90%，预测被试编号的 ITR 权重占 10%。

$$ITR_3 = 0.9 \times ITR_1 + 0.1 \times ITR_2$$

其中，T 表示平均试次时长，M 表示目标个数，P 表示识别正确率。ITR 的单位是 bits/min。特别需要指出，本系统中 ITR 是按照理想 ITR 进行计算，即

平均试次时间不包含模拟休息时长。另外，**如果参赛队算法对于所有被试平均的正确率低于随机分类的准确率，则成绩无效，ITR 记为 0。**

**需要注意的是：当参赛者提交结果的单次试次时长小于 3s 时，统一计为 3s。**

## 7、性能评估方法

参赛算法通过数据读取方法获取新数据包。当所得数据包内含有 Trigger 信号时，评分系统将自动开始记录算法识别过程中所使用 EEG 信号的长度，直至反馈方法被调用。从 Trigger 开始到反馈方法被调用时所获取的 EEG 数据长度将作为该试次的模拟试次时长，当模拟试次时长小于 3s 时，统一计为 3s。而平均准确率将根据算法反馈结果与真实刺激的一致性进行计算。

需要特别指出，在本比赛项目中每一个包含 Trigger 的数据包，其依然被视为是前一试次的数据。而新试次数据是从包含 Trigger 数据包的下一个数据包开始计算。因此参赛算法不可在获取到包含 Trigger 信号的数据包时立刻反馈，而最早需获取到下一数据包后才可反馈。

## 8、结果反馈异常处理

### 1) 重复多次报告

在单个试次内如果出现多次报告结果，则该试次的**成绩总是由最后一次有效提交**的时间以及结果决定。注意，一个试次内总包含有多个数据包，参赛算法每获取一个数据包，则获取一次前文中所提及的有效提交机会，即在获取下个数据包前，如果有多次提交则仅其中第一个提交为有效提交。原则上一个试次内的**最大的不超时的有效提交次数**为获取起始 Trigger 信号到获取结束 Trigger 信号过程中数据包个数。

### 2) 结果未反馈

如果对于某一个或多个试次，参赛算法获取了对应时间段内所有的数据包但不提交结果，则这些试次的判决结果将被记录为误判，并且所用时长记为试次的总时间长度即试次起始 Trigger 信号到结束 Trigger 的时间长度。

### 3) 结果反馈超时

结果反馈时，若参赛算法在某一个试次内使用了超过结束 Trigger 信号的数据包，则本次判决结果将被记录为误判。

#### 4) 算法执行超时

为满足脑-机接口系统实时处理需求，本项目同时对参赛算法的计算复杂度有一定要求。若算法复杂度过高导致系统运行超时，则该算法比赛成绩将被视为无效。

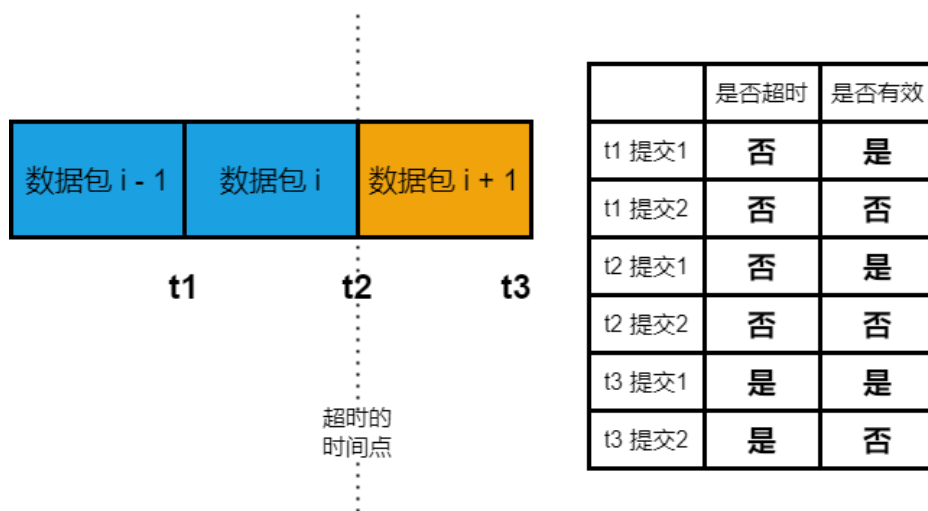


图 5 试次提交情况说明

一个试次中的提交情况如图 5 所示，假设参赛算法获取了三个数据包  $i-1$ 、 $i$  以及  $i+1$ ，则获取数据包后，参赛算法提交的时间点对应  $t_1$ 、 $t_2$ 、 $t_3$ 。其中，数据包  $i$  包含了结束 Trigger 信号。假设参赛算法在  $t_2$  时间点提交了多次报告，则按照  $t_2$  时间的第一次提交作为有效不超时提交；若参赛算法在  $t_3$  提交多次报告，则按照  $t_3$  时间的第一次提交作为有效但超时的提交。赛题仅使用有效提交的结果进行计分。

## 六、奖项设置

本赛项设特等奖、一等奖、二等奖、三等奖，四个奖项。获奖总赛队数量为 20 支队，选取初赛前 10 支队伍在北京“2024 世界机器人大赛锦标赛（北京）”现场参加决赛。决赛中根据分数排名，设特等奖 1 个、一等奖 2 个、二等奖 3 个，三等奖 14 个。

其中特等奖、一等奖、二等奖设奖金奖励。

特等奖 1 名 -- 奖金 30000 元、比赛证书；

一等奖 2 名 -- 奖金 20000 元、比赛证书；

二等奖 3 名 -- 奖金 10000 元、比赛证书；

三等奖 14 名 -- 比赛证书。

奖金金额为税前金额。

**赛题联合命题组：**

- 蚂蚁数字科技
- 杭州电子科技大学
- 深圳大学