

2026 年（第 19 届）中国大学生计算机设计大赛大数据主题赛

“健康数据洞察” 赛题说明

一、比赛题目

题目：健康数据洞察

题目描述：

促进全民健康、确保医疗卫生服务的可及性与公平性，是联合国可持续发展目标（SDG 3）的核心内容，也是我国“健康中国”战略的基石。在全球范围内，疾病谱正从传染性疾病为主转向非传染性疾病为主，但不同国家与区域间仍存在巨大的健康不平等。研究全球疾病负担的演变规律、健康风险因素的归因贡献以及卫生资源配置的效率与公平，对于优化全球卫生治理、完善本国公共卫生体系至关重要。例如，空气污染、不健康饮食等风险因素不仅直接推高特定疾病的发病率和死亡率，其影响程度又因各国的经济发展阶段、医疗资源可及性和公共卫生政策有效性而产生巨大差异，最终体现为人均预期健康寿命的国际差距。同时，卫生资源的投入规模与配置结构，直接关系到健康产出的效率与公平性。因此，健康结果、风险暴露与卫生资源共同构成了一个全球性的健康生态系统。如何通过多源数据，量化评估各国在这一系统中的位置、识别关键短板并提出优化路径，对于提升全球健康水平具有重大现实意义。

随着人工智能技术的快速发展，智能体已成为数据密集型科学研究与决策支持的重要赋能工具。本次比赛鼓励参赛者探索智能体在健康数据分析中的创新应用，例如利用智能体进行自动化数据探索、关联模式发现、预测建模以及资源优化策略的模拟推演，以人机协作的方式深化对健康生态系统复杂性的理解，并生成更具洞察力与可操作性的解决方案。

可参照的思考维度包括并不限于（鼓励学生提出自己的思考维度）：

1. 描述全球疾病谱系（从传染性疾病为主转向非传染性疾病为主）的时空变迁，并量化国家间、区域间在关键健康指标上的差距及其变化趋势。
 - (1) 形成全球疾病谱系的时空分布，挖掘疾病谱系的时空变化特征
 - (2) 形成不同国家在关键健康指标上的分布，分析影响健康指标的因素

- (3) 建立地区疾病谱系和健康指标的关联关系，对疾病谱系和健康指标的变化进行预测
 - (4) 给出导致变化的解释，并提出优化健康指标的方法
2. 分析不同风险因素（如 PM2.5、吸烟、不健康饮食）对不同地区疾病负担的具体贡献度，并预测未来主要疾病负担的变化。
- (1) 利用归因分析模型，划分不同风险因素对各国主要疾病（如心血管疾病、肺癌）死亡人数的具体贡献比例
 - (2) 建立风险因素暴露水平变化与未来一段时间后疾病负担变化之间的关联模型
 - (3) 模拟在不同公共卫生政策力度（如控烟立法、空气治理目标）下，未来 20 年重点疾病负担的变化区间
 - (4) 根据不同地区风险贡献的“主次矛盾”，提出最具成本效益的、循序渐进的优先干预措施清单
3. 评估各国卫生资源（如医生密度、卫生支出）配置与疾病负担的匹配程度，识别“资源配置不足”或“健康产出低下”的区域，并提出数据驱动的优化见解。
- (1) 测算各国“实际资源配置水平”与“基于其疾病负担的理论需求水平”之间的缺口，并在地图上进行分级标注
 - (2) 建立评估矩阵，将各国分为“高投入-高产出”、“高投入-低产出”、“低投入-高产出”、“低投入-低产出”四种类型，并分析其健康结果在不同人群间的公平性
 - (3) 设定“最大化健康产出”或“最小化健康不平等”等不同目标，通过数学规划模型，模拟卫生资源（如医疗人力、资金）在各国或一国内部的最优再分配方案
 - (4) 根据评估与模拟结果，为“高投入-低产出”地区提供以提升管理效率为核心的建议，为“低投入-高产出”地区总结可推广经验，为“低投入-低产出”地区设计增加基础投入与改善效率并行的综合性方案

数据说明：

本次比赛提供以下数据集，参赛选手必须选取其中的一个或多个数据集作为**基础数据**。同时，鼓励选手补充其他数据集以充实数据分析，但需要提供完整的数据集和描述。
(以下数据集需要先在开发平台完成报名，才能查看和使用)

1) [全球各国核心疾病与死亡数据](#)

- 2) [全球各国健康风险因素数据](#)
- 3) [全球各国健康营养和人口统计数据](#)
- 4) [全球社会经济发展背景数据](#)
- 5) [全国近 20 年卫生数据](#)

数据分析模型协作支持: DeepAnalyze

DeepAnalyze 是中国人民大学自研的面向数据科学的大语言模型，具备自动完成数据理解、分析建模与结果解释等多类数据科学任务的能力。该模型支持以“用户-智能体”协作的方式开展数据分析工作，可作为数据分析与智能决策场景中的低门槛、高效率开发工具。

本赛题鼓励参赛队伍通过 Modelwhale 平台内嵌功能或指定 API 接口调用 DeepAnalyze，结合人工决策与模型推理能力完成数据分析任务。参赛作品中对 DeepAnalyze 的调用方式、协作流程及实际效果将作为评审的重要参考因素。

DeepAnalyze 和 API 的具体使用请参考：

- 技术报告: <https://arxiv.org/abs/2510.16872>
- GitHub 仓库: <https://github.com/ruc-datalab/DeepAnalyze>

二、作品提交要求

1. 提交材料

参赛选手需至少提交的材料：主题分析报告、数据处理源程序、作品展示视频、自增的数据集（若有）、及智能体交互过程记录及五部分（其他的请参见大赛的“大数据应用”大类的作品提交要求）。

- **主题分析报告：**围绕题目撰写分析报告，运用数据思维，发现与解释经济社会现象，探讨与解决社会问题。以 PDF 形式提交。请参考附录的模板。请注意篇幅要求：正文部分不超过 20 页（不含参考文献）。如有需要，可在正文之后另行附加附录，附录篇幅不计入 20 页限制。请务必将核心观点与主要结论集中呈现在正文中，并在正文中对附录内容作出明确指引。附录仅作为补充参考材料，不作为主要评价依据。
- **数据处理源程序：**从技术应用的角度描述数据来源、数据处理、数据分析方法。另外，分析报告中的分析结果（包括并不限于图表、统计数字等）均应在程序源代码注释

或说明中列出详细的分析步骤，否则相应的结果不计入作品评审中。需要采用 Python/R 语言的源程序，请提交已在和鲸平台完成复现的源代码。

- **作品展示视频：**初赛选手采用视频（mp4）演示程序运行和结果；决赛选手需在参赛现场做汇报和演示。
- **数据集（可选）：**如引入其他支持数据集，则需提交使用的数据集。
- **智能体交互过程记录：**为体现智能体在作品创作中的协作角色，参赛队伍需提交与 DeepAnalyze（或其他自选 AI 智能体）的交互记录。记录应清晰展示智能体在需求理解、方法建议、代码生成、结果解释等一个或多个环节中的具体贡献。提交形式可以是对话日志（JSON/TXT）、关键交互截图或包含交互过程说明的 Markdown 文档。该记录将作为评估“智能体协作”维度的重要依据。

json 文件形式的交互记录样例：

```
{  
  "id": "xxx",  
  "messages": [  
    {  
      "role": "user",  
      "content": "...",  
    },  
    {  
      "role": "assistant",  
      "content": "...",  
    },  
    ...  
  ]  
}
```

2. 开发和提交方式

- 1) 作品开发：登陆和鲸平台上进行作品开发。
- 2) 生成作品链接：在和鲸平台的提交页面进行提交，提交完成后在提交记录处查看提交内容，复制作品链接。



作品提交

剩余2天

作品提交

时间
2023/01/04 00:00 – 2023/01/06 00:00

*主题分析报告
上传
文件格式要求为: pdf; 大小限制: 100MB;

*数据处理源程序
选择已有项目, 或者[创建新项目](#)

提交

提交记录

提交内容	提交时间	提交者	状态	分数
查看 复制链接	2023/01/04 10:51	...	待评审	-

请每支团队队长在【提交】入口的提交窗口查看提交须知后进行材料提交。

- 除提供数据外，鼓励选手通过其他方式补充数据进行支撑分析。是否对于数据进行充分利用也将是赛题考察的重要环节。
- 对于外部数据源，可在代码内部采用爬虫方式直接获得，也可将现有数据源随项目一同提供。
- 对于获奖作品，可能需要补充数据获取方式与清洗、保存的代码，因此，请保留相关代码及数据直至比赛结束。
- 可重复提交，每次提交均可以生成作品链接。

三、比赛日程

初赛阶段

- 选手登陆和鲸平台上进行作品开发，并生成作品链接；
- 按校赛或省赛规定的时间和提交方式提交作品链接。

决赛阶段

- 决赛于 2026 年 7 月 17 日-7 月 21 日进行，以国赛公告为准。
- 决赛形式：答辩

四、竞赛培训

为使参赛选手尽快熟悉和鲸平台，并了解 DeepAnalyze 智能体的使用，赛题方将为参赛选手提供竞赛培训。培训内容包括：Python/R 技能提升、赛题解读、数据分析报告的设计要点以及智能体协作数据分析的最佳实践等。

五、大赛组织

主办单位：中国大学生计算机设计大赛组委会

承办实施单位：东华大学

承办单位：中国人民大学、华东师范大学、东北大学、华东理工大学、上海大学、上海电力大学

协办单位：北京正阳恒卓科技有限公司、上海和今信息科技有限公司

联系方式：

- 赛题咨询：黄老师，邮箱：keman@ruc.edu.cn；吴老师，邮箱：service@heywhale.com
- 开发平台交流群：扫描下方二维码添加赛事助手→点击客服链接→回复或直接点击关键词【大数据主题赛】→即可扫码进群，群内会进行重要赛事通知及答疑。

