



第二届北京市大学生“人工智能+”创新大赛
THE 2ND BEIJING COLLEGE STUDENTS' AI+INNOVATION COMPETITION

智领新质 共创未来

第二届北京市大学生“人工智能+”创新大赛

赛题规则

赛题

“临床智录”——住院首次病程
记录智能生成

首都医科大学宣武医院、润达医疗科技股份有限公司、华为技术有限公司

一、赛题背景

病历书写是临床工作中负担最重、重复性最高的环节之一。首次病程记录（首程）作为住院病历的核心文书，要求医生在入院初期基于有限的入院资料，凝练病例特点、形成拟诊讨论并制定诊疗计划，既考验临床思维，也耗费大量时间。随着大语言模型技术在医疗领域的快速发展，利用生成式 AI 辅助首程书写、减轻医生文书负担、规范病历质量，已成为“AI+医疗”落地的典型场景。本赛题以住院首次病程记录生成任务为载体，要求参赛团队开发一套可运行的人工智能应用：系统读取统一提供的教学病例资料（主诉、现病史、既往史、体格检查、辅助检查等，均设有资料截止点），在规定时间内调用主办方指定的大模型，提炼病例特点，形成有依据的拟诊讨论和诊疗计划，并展示关键内容的资料来源及需要医生补充确认的信息。赛题产出仅用于教学比赛与评审，不用于真实诊疗。

二、赛题考察点及技术创新

本赛题重点考察选手在事实忠实性、临床分析能力与工程可复现性三方面的综合能力：要求严格基于给定资料生成，不虚构事实、不将“未提供”改写为“正常”、不提前写入后续诊断或已实施治疗；拟诊须有依据、鉴别有针对性、正确处理信息不足；算法需可批量运行、配置规范、日志完整、结果可复核。技术创新的主线是围绕大模型调用构建“提取→整理→生成→校验”的流水线。参赛者可通过事实提取（抽取结构化事实以防遗漏）、时间整理（按截止点过滤时序）、分段生成（通过管道式多次调用提升逻辑一致性）、检索增强（RAG 辅助鉴别诊断）、生成后校验（数值回溯与 Schema 验证）等技术路线自主设计并论证改进方案，用可对比的实验证明每一步改进的实际收益。

三、参赛要求

本次大赛面向的参赛对象为北京地区普通高等学校在校本科生，不限制专业和年级，本赛题以团队形式报名参赛，每支队伍由 3-5 名队员组成，指导教师不多于 2 人。

四、赛程安排与赛题规则

本赛题按高校预选推荐、企业分赛场选拔、北京市线下决赛三阶段实施，

遵循大赛总体时间安排。**高校预选推荐阶段**由各高校自行组织校内选拔，推荐队伍进入企业选拔阶段；企业分赛场选拔阶段比赛形式为线上材料提交+线下演示，由企业统一组织，并选拔出队伍晋级总决赛；**北京市线下总决赛**安排由大赛组委会另行通知。具体要求如下：

（一）高校预选推荐

校赛阶段由各高校自行组织，各高校可结合发布的赛题任务组织校内选拔，每所高校可推荐不超过 5 支队伍参加企业赛。

（二）企业选拔题：系统开发与统一评测验证

企业分赛场选拔采用“线上开发+统一上传+线上评审+线下复核”形式。开发期内，参赛队伍使用 Python 开发可批量处理病例的程序。选拔阶段对进入评奖范围的队伍增加线下算法复现核验环节。

1.参赛队伍需完成以下核心任务：

任务 1：算法开发与模型调用。读取赛方提供的入院资料，远程调用主办方指定的 Qwen 大模型，生成包含“病例特点、拟诊讨论、诊疗计划”三段正文的首程草稿，并附关键资料依据（source_id 可溯源）和待核事项。

任务 2：统一评测运行。比赛日统一运行算法包处理正式病例并提交结果。程序需在主办方提供的封闭网络环境中运行，使用统一启动入口（main.py）和指定的依赖环境（Python 3.11）。

2.提交资料：

- (1) **算法包与源代码：**命名为“队伍编号_algorithm.zip”，含 README、requires.txt 依赖清单、统一启动入口 main.py、Prompt 文本及版本，固定模型参数与随机种子以确保结果可复现。
- (2) **结果包与运行记录：**命名为“队伍编号_results.zip”，含 results.json（覆盖全部正式病例，不人工修改正文）与 run_log.json。
- (3) **项目说明：**不超过 10 页，说明提示词设计、生成流程、改进点及局限。
- (4) **声明文件：**所用模型及版本、检索资料来源与版本声明、大模型使用声明和数据使用承诺书。

3.评分标准参考维度：

评分维度	评价内容
事实准确与忠实性	重要事实、数值、时序和否定关系正确，无凭空补写
拟诊讨论与临床分析	判断有依据，鉴别有针对性，正确处理信息不足
诊疗计划针对性	检查、处理和观察安排与当前问题对应，区分计划与已实施事项
结构规范与个体化	三段完整，重点清楚，体现病例特征而非泛化套话
来源说明与待核提示	关键内容可核对来源，缺失和冲突提示明确
算法复现与改进验证	按说明能运行、配置依赖和日志完整；改进点明确，能用公开病例对比说明收益及局限

4.线下算法复现核验：

企业分赛场选拔设置线下算法复现核验：技术委员会在主办方提供的统一环境中，对进入评奖范围的参赛队伍算法包进行现场抽样复现——在干净环境中重新安装依赖，以 `python main.py` 统一入口运行，核对生成结果与提交结果包的一致性，并核查运行记录（调用次数、时间戳、病例编号对应关系等）的真实性；队伍代表须在场观看核验过程并接受技术评委问询。复现失败、生成结果与提交结果明显不一致或运行记录无法对应的队伍，算法复现部分记零分；涉嫌伪造运行记录或篡改日志的，按赛事违规程序处理。线下复现核验不改变现有评分权重，作为评奖与总决赛推荐资格的核验门槛。

五、设备要求及资源支持

（一）设备与软硬件要求

- 日常开发：**参赛队伍需自备具备联网条件的电脑开展日常开发、调试与测试。
- 评测环境：**正式评测不要求学生自备服务器或图形加速设备。统一运行环境由主办方提供，正式版本为 `Python 3.11`，依赖通过 `pip` 安装。
- 模型接口：**企业选拔赛统一评测在主办方提供的封闭网络环境中运行，模型由主办方以无密钥接口方式提供，运行环境、统一启动入口、模型配置及输入输出协议须严格执行赛题技术规范。

(二) 资源支持

比赛期间，命题企业统一提供以下资源支持：

支持内容	提供方式与范围
赛题资料	统一发布练习病例、模型清单、接口文档与技术规范。
评测资源	正式评测时提供封闭网络环境、统一启动入口、无密钥模型API。
技术答疑	由赛题技术委员会负责技术答疑与规则解释。

(四) 安全及风险提示

1. **数据安全：**比赛病例资料仅限参赛期间使用，参赛队伍不得对外传播或用于其他用途；生成结果仅供教学比赛评审，不得作为真实诊疗依据。
2. **合规使用：**参赛队伍应遵守模型使用限制，仅调用赛方公布的指定模型。
3. **网络安全：**参赛队伍不得对评测环境进行攻击、压测或越权访问，发现问题应及时报告组委会；不得伪造运行记录或篡改日志，违者按赛事违規程序处理。

六、附加说明

1. **评审与奖励：**高校校赛完成推荐；企业分赛场选拔开展正式评审，评出一、二、三等奖并推荐总决赛队伍；总决赛评奖按大赛统一规定执行。
2. **规则解释：**本规则未尽事宜，由赛题技术委员会与大赛组委会协商确定。本规则如与大赛组委会的其它规定不一致时，以大赛组委会规定为准。

“临床智录”——住院首次病程记录智能生成
赛题答疑群



附件 1 学生拿到的输入入院资料（示例）

以下为完全人工编写的格式示例，不来自真实患者，也不是正式比赛答案。
具体医学要点由赛方专家审定。

病例编号 DEMO001，资料截止点 T0，以下资料在 T0 已可获得。

source_id	资料内容
S01	成人，乏力 2 周，活动后心悸；现有记录未见明确黑便、肉眼血尿或其他出血表现
S02	结膜苍白、双下肢轻度水肿；未提供血压、尿量及其他体征
S03	血红蛋白 82 g/L，实验室标低；肌酐 185 μmol/L，实验室标高；未提供既往肌酐、红细胞指数、网织红细胞或尿检
S04	既往病史、药物史、肾功能基线和贫血既往检查未提供；入院初步判断为贫血及肾功能异常，原因待查

算法应输出的首程示例

病例特点。成人，近 2 周乏力、活动后心悸，现有记录显示结膜苍白和双下肢轻度水肿。血红蛋白 82 g/L、肌酐 185 μmol/L，实验室分别标示降低和升高。现有记录未见明确出血表现；既往血红蛋白、肾功能基线、血压和尿量不详。

拟诊讨论。现有资料支持贫血及肾功能异常，需要分别明确原因及两者关系。贫血需结合红细胞指数、网织红细胞和针对性检查进一步分析，现有记录不能排除隐匿失血、营养相关因素或其他原因。单次肌酐升高不足以确定急性或慢性过程，也不足以将贫血直接归因为肾性贫血。水肿需结合血压、尿量、尿检及相关检查综合判断。目前病因待查。

诊疗计划。由临床医生补充既往检查、症状变化、用药及出血相关病史，核实血压、尿量和容量状态；结合病情复核血常规及肾功能，按鉴别需要安排进一步检查。观察症状和指标变化，依据补充资料调整分析与处理，治疗需结合完整临床评估确定。

依据说明。血红蛋白 82 g/L 来自 S03，属于事实；急慢性过程未明由 S03 和 S04 支持，属于推断边界。其余关键症状和体征分别对应 S01、S02。

待核事项。既往肾功能、贫血病程、血压尿量及用药史。没有提供不等于正常；不能把初步判断写成已确诊病因，也不能补写已实施治疗。

比赛日，学生运行自己的算法，以同样方式处理正式病例，把全部结果与日志上传指定入口，并附算法包。命题方收到材料后组织评分。

附件 2 评分标准

评价维度	评价内容
事实准确与忠实性	重要事实、数值、时序和否定关系正确，无凭空补写
拟诊讨论与临床分析	判断有依据，鉴别有针对性，正确处理信息不足
诊疗计划针对性	检查、处理和观察安排与当前问题对应，区分计划与已实施事项
结构规范与个体化	三段完整，重点清楚，体现病例特征而非泛化套话
来源说明与待核提示	关键内容可核对来源，缺失和冲突提示明确
算法实现与可复现性	按说明能运行，代码提示词与提交结果对应，配置依赖和日志完整
改进方法与验证	改进点明确，能用公开病例对比说明收益及局限
其他	不对界面美观或产品功能另行加分

评分点细化

（一）每例五个临床维度的得分点

维度 1 事实准确与忠实性

得分点	做法说明
关键事实正确	症状、体征、检验数值与资料一字不差，数值带单位（如 Hb 82 g/L）
时序正确	保留“乏力 2 周”“入院时”等原始时间表述，不颠倒先后
否定与缺失如实保留	“未见黑便”“血压未提供”原样保留“未”字；缺失的写“不详/未提供”
无凭空补写	资料里没有的检查、病史一律不写；不将“未提供”改写为“正常/否认”

维度 2 拟诊讨论与临床分析

得分点	做法说明
初步判断有依据	每个判断对应到具体资料，如“贫血待查（Hb 82 g/L，见 S03）”
鉴别有针对性	围绕本例特点列 2—3 个鉴别方向并说明理由，不照搬教科书清单

正确处理信息不足	资料不够时写“需进一步检查明确”；写“病因待查”不为缺点，不硬下结论
----------	------------------------------------

维度 3 诊疗计划针对性

得分点	做法说明
计划对应当前问题	检查、处理与观察安排跟本例问题对应，不列无关的大而全清单
区分计划与已实施	用“拟/建议/复查”表述，绝不把计划写成“已予/已完成”
有观察安排	写明观察什么症状、复查什么指标

维度 4 结构规范与个体化

得分点	做法说明
三段完整	病例特点、拟诊讨论、诊疗计划每段非空
重点清楚	病例特点按“重要阳性→有价值的阴性→病程”排列，不写流水账
个体化表达	输出体现本例特征，换一个病例无法直接套用

维度 5 来源说明与待核提示

得分点	做法说明
依据可溯源	evidence 列 3—5 条（不足 3 条按实际写），每条标注 source_ids 且编号真实存在于本病例
事实推断区分	资料中原有内容标 fact，由算法推理得出的标 inference
待核事项明确	缺失、冲突、需医生确认的问题列入 review_items；确实没有时写[]

（二）整体评分部分的得分点

评分项	得分点	做法说明
算法复现	可重新运行	干净环境 pip install -r requires.txt 后 python main.py 能自动处理全部病例并正常退出
算法复现	材料相互对应	代码、提示词、参数与提交结果一致；固定随机种子
算法复现	配置日志完整	main.py、requires.txt、.env.example、README 齐备；run_log 字段齐全且不含密钥
改进验证	改进点明确	说清改进内容及解决的问题，如“数值回溯核对拦截抄错”
改进验证	同条件对比	基线版与改进版在同一批公开练习病例上的结果对比
改进验证	局限分析	说明代价（耗时/调用量）或尚未解决的问题

评审与复核建议

每例由两名医学评委独立评分，依据赛前准备的逐例参考要点，允许合理的不同表达，不以文字相似度代替医学评价。总分分小计差异超过 10 分，或重

大错误等级不一致时，由第三位医学专家复核争议项。算法部分由技术评委结合线下复现核验评定：在统一环境现场抽样重跑队伍算法包，核对代码、提示词、配置与提交结果的对应关系；复现未通过或运行记录不实的队伍，算法复现记零分，并视情节按赛事违规程序处理。抽样病例与条件对各队一致；差异先核查模型随机性和版本，不直接认定作弊。

所有正式病例进入分母。未提交、病例错配、无法读取或缺任一正文必需部分的病例记 0 分；合格输出中一般错误在相应维度扣分。编造关键事实、关键时序或数值错误并明显影响判断的，该例总分小计最高 50% 的分；基于题内明确证据会导致严重伤害的建议或严重遗漏，经第三位医学专家复核后该例记 0 分。先合并双评分及争议裁决，再应用一次封顶，不对同一错误另加重复扣分。

按未舍入总分排序，展示两位小数；同分依次比较临床前四项得分、事实忠实性、严重错误例数较少。用于同分比较的分项按封顶后总分与原始小计之比同比缩放，零分病例分项也记 0；仍相同则并列，由组委会决定名额。算法复现抽样病例与条件对各队一致；差异先核查模型随机性和版本，不直接认定作弊。

故意人工代写正式结果、获取答案、篡改日志或使用未获准外部模型，按赛事违规程序处理。公共 API 故障由赛方统一处理，不让个别队伍自行选择重跑结果。正式发布前由技术委员会负责解释。

附件 3 技术规范

1. 源码包与运行文件

提交文件命名为 `队伍编号_algorithm.zip`。ZIP 解压后的根目录须直接包含以下文件，不再套一层目录。

<code>main.py</code>	# 统一启动入口
<code>requires.txt</code>	# Python 依赖清单
<code>.env.example</code>	# 配置模板，仅包含 <code>MODEL_NAME</code> 和 <code>ENDPOINT</code>
<code>README.md</code>	# 方法、运行步骤、模型参数、改进点及局限
<code>src/</code>	# 其他代码，可选
<code>prompts/</code>	# Prompt 文件，可选

主办方在正式运行时，将以下文件放入源码根目录：

<code>.env</code>	# 主办方提供的模型配置
<code>input.json</code>	# 正式测试输入

程序运行完成后，须在源码根目录生成 `output.json`。根目录指 `main.py` 所在目录；所有相对路径均以该目录为基准。

提交包须包含运行所需的全部代码、Prompt 和静态资源。不得依赖参赛者本机绝对路径、未提交的文件或人工操作。模型由主办方提供接口，提交包无需包含模型权重。

2. 运行环境与统一启动入口

本技术规范采用 Python 3.11 作为统一运行环境，正式版本随赛前技术附件发布。主办方按以下命令安装依赖并启动程序：

```
cd <源码解压目录>
python -m pip install -r requires.txt
python main.py
```

`main.py` 为唯一启动入口，无需额外命令行参数。程序启动后自动读取 `.env` 和 `input.json`，完成全部病例处理、写入 `output.json` 后退出，不依赖交互输入、IDE 或额外启动的本地服务。

依赖文件统一命名为 `requires.txt`，采用可被 `pip install -r requires.txt` 解析的标准格式。只列实际使用的依赖，并建议固定已测试版本。以下仅为依赖名称示例，参赛者可按实现调整：

```
python-dotenv
requests
```

3. 模型配置与无密钥调用

`.env` 只包含以下两项模型配置：

```
MODEL_NAME=replace_with_model_name
ENDPOINT=http://127.0.0.1:8000/v1/chat/completions
```

上述值仅为占位示例，实际值由主办方提供。

配置项	说明
<code>MODEL_NAME</code>	模型服务实际使用的模型标识，作为请求体中的 <code>model</code> 值
<code>ENDPOINT</code>	完整推理请求地址，包含实际请求路径；程序直接使用，不再拼接 <code>/v1</code> 或 <code>/chat/completions</code>

模型接口无需 **API Key**、**Token** 或其他密钥。程序不得要求配置密钥，不得因缺少密钥而拒绝启动，模型请求不发送 ``Authorization`` 鉴权头。

程序须从 `main.py` 所在目录的 `.env` 文件读取上述两项配置，不得将实际模型标识或地址硬编码在源码中。输入输出文件名固定为 `input.json` 和 `output.json`，不通过 `.env` 配置。

本技术规范按 **Chat Completions** 兼容协议说明调用方式，最终请求和响应协议由主办方随技术附件确认。请求方式为 `POST`，请求体示例：

```
{
  "model": "replace_with_model_name",
  "messages": [
    {
      "role": "system",
```

"content": "生成教学比赛中的住院首次病程记录草稿。仅依据截止点及此前可获得的资料，输出病例特点、拟诊讨论、诊疗计划、依据说明和待核事项。区分事实与推断，不虚构事实，不把计划写成已实施治疗，仅返回约定的 JSON 对象。"

```
    },  
    {  
        "role": "user",  
        "content": "此处填入当前病例的原始资料。"  
    }  
],  
"stream": false  
}
```

以下为配置读取和模型调用示例，不是完整参赛程序；参赛者仍须实现输入读取、病历生成、输出解析、异常处理和结果写入。

```
from pathlib import Path  
  
import requests  
  
from dotenv import dotenv_values  
  
BASE_DIR = Path(__file__).resolve().parent  
INPUT_FILE = BASE_DIR / "input.json"  
OUTPUT_FILE = BASE_DIR / "output.json"  
  
config = dotenv_values(BASE_DIR / ".env")  
  
model_name = (config.get("MODEL_NAME") or "").strip()  
  
endpoint = (config.get("ENDPOINT") or "").strip()  
  
if not model_name or not endpoint:  
    raise ValueError(".env 必须提供非空的 MODEL_NAME 和 ENDPOINT")  
  
# 不从系统代理或凭据配置中引入额外设置。  
  
session = requests.Session()  
  
session.trust_env = False  
  
def call_model(messages):  
    response = session.post(  
        endpoint,  
        headers={"Content-Type": "application/json"},
```

```

    json={
        "model": model_name,
        "messages": messages,
        "stream": False,
    },
    timeout=120, # 示例值，正式限制由主办方发布
)

response.raise_for_status()

return response.json()["choices"][0]["message"]["content"]

```

生成参数、调用次数、并发、超时和重试限制由主办方统一发布。参赛代码应遵守这些限制，不自行调用其他模型或未经允许的外部服务。

4. 输入文件规范

输入文件固定为 `input.json`，采用 UTF-8 编码，内容为标准 JSON 数组。数组中的每个对象代表一例待生成首次病程记录的入院资料，不采用 JSONL 格式。

字段	类型	说明
<code>case_id</code>	string	非空的病例唯一标识，文件内不得重复
<code>record_cutoff</code>	string	资料截止点，可以是时间或赛方约定的标识，例如 <code>T0</code>
<code>sources</code>	array	资料片段数组，每个对象包含 <code>source_id</code> 和 <code>content</code>
<code>sources[source_id].source_id</code>	string	该病例内唯一、非空的资料片段标识
<code>sources[source_id].content</code>	string	资料片段正文，保留原始事实、数值、时序及不确定性

资料可包括主诉、现病史、既往史、体格检查、辅助检查和入院初步判断。未提供的资料保持缺失，仅使用截至资料截止点（含该时点）已可获得的信息。赛方在发题前筛选资料，排除目标首次病程、后续确诊、出院记录及截止点之后的检查结果。

以下沿用赛题中的人工编写教学样例，不来自真实患者，也不是正式比赛答案；具体医学要点由赛方专家审定。sources 和 content 是本执行手册约定的资料承载字段。

```
[
  {
    "case_id": "DEMO001",
    "record_cutoff": "T0",
    "sources": [
      {
        "source_id": "S01",
        "content": "成人，乏力 2 周，活动后心悸；现有记录未见明确黑便、肉眼血尿或其他出血表现"
      },
      {
        "source_id": "S02",
        "content": "结膜苍白、双下肢轻度水肿；未提供血压、尿量及其他体征"
      },
      {
        "source_id": "S03",
        "content": "血红蛋白 82 g/L，实验室标低；肌酐 185 μmol/L，实验室标高；未提供既往肌酐、红细胞指数、网织红细胞或尿检"
      },
      {
        "source_id": "S04",
        "content": "既往病史、药物史、肾功能基线和贫血既往检查未提供；入院初步判断为贫血及肾功能异常，原因待查"
      }
    ]
  }
]
```

示例数组展示一例病例；正式文件可包含多例，每例按相同结构组成一个对象。公开样例和正式测试数据遵循同一字段协议，正式数据的内容及数量可不同。输入为空数组时，输出也应为空数组。

5. 输出文件规范

输出文件固定为 `output.json`，采用 UTF-8 编码，内容为标准 JSON 数组。

- 输出条数和顺序与输入一致，并原样保留每条记录的 `case_id`。
- 每条记录包含 `case_id`、`status`、`medical_record`、`error`，不得遗漏失败记录。
- 成功时 `status` 为 "success"，`medical_record` 为病历对象，`error` 为 `null`。
- 失败时 `status` 为 "failed"，`medical_record` 为 `null`，`error` 为包含 `code` 和 `message` 字符串的对象。
- 成功结果的三段正文均为非空字符串；资料不足时在相应正文和待核事项中说明，不补造事实。
- `evidence` 为依据对象数组，`review_items` 为待核事项字符串数组；没有待核事项时使用 `[]`，不省略字段。
- 最终文件须可直接解析为 JSON，不包含 Markdown 代码围栏、说明文字或日志。

`medical_record` 固定包含以下五个医学字段：

字段	类型	内容要求
<code>case_features</code>	string	病例特点：凝练重要阳性、有鉴别价值的已知阴性及病程信息
<code>diagnostic_discussion</code>	string	拟诊讨论：有依据的初步判断、鉴别分析及需要进一步明确的问题
<code>plan</code>	string	诊疗计划：对当前问题的进一步检查、处理及观察安排
<code>evidence</code>	array	依据说明：列出 3—5 条关键事实或判断及对应资料标

		识；关键事实不足 3 条时按实际列出
review_items	array	待核事项：列出资料缺失、冲突及需医生确认的问题

每条 `evidence` 对象包含以下字段：

字段	类型	内容要求
statement	string	可核对的事实或判断，说明简短清楚
source_ids	array	该条依据对应的一个或多个 <code>source_id</code> ，须属于当前病例
kind	string	<code>fact</code> 表示资料中的事实， <code>inference</code> 表示基于资料的推断或推断边界

三段正文建议合计 800—1600 字，以清楚完整为准，不按长短加分。不求输出模型内部思维链，只需要可核对的结论与简短依据。以下教学样例用于展示结构。

与上述输入对应的输出示例：

```
[
  {
    "case_id": "DEMO001",
    "status": "success",
    "medical_record": {
      "case_features": "成人，近 2 周乏力、活动后心悸，现有记录显示结膜苍白和双下肢轻度水肿。血红蛋白 82 g/L、肌酐 185 μmol/L，实验室分别标示降低和升高。现有记录未见明确出血表现；既往血红蛋白、肾功能基线、血压和尿量不详。",
      "diagnostic_discussion": "现有资料支持贫血及肾功能异常，需要分别明确原因及两者关系。贫血需结合红细胞指数、网织红细胞和针对性检查进一步分析，现有记录不能排除隐匿失血、营养相关因素或其他原因。单次肌酐升高不足以确定急性或慢性过程，也不足以将贫血直接归因为肾性贫血。水肿需结合血压、尿量、尿检及相关检查综合判断。目前病因待查。",
      "plan": "由临床医生补充既往检查、症状变化、用药及出血相关病史，核实血压、尿量和容量状态；结合病情复核血常规及肾功能，按鉴别需要安排进一步检查。观察症状和指标变化，依据补充资料调整分析与处理，治疗需结合完整临床评估确定。",
      "evidence": [
        {
```

```

    "statement": "近 2 周乏力、活动后心悸，现有记录未见明确出血表现。",
    "source_ids": ["S01"],
    "kind": "fact"
  },
  {
    "statement": "结膜苍白、双下肢轻度水肿，血压和尿量未提供。",
    "source_ids": ["S02"],
    "kind": "fact"
  },
  {
    "statement": "血红蛋白 82 g/L、肌酐 185 μmol/L，实验室分别标低和标高。",
    "source_ids": ["S03"],
    "kind": "fact"
  },
  {
    "statement": "仅有单次肌酐及缺失的肾功能基线资料，肾功能异常的急慢性过程尚不能确定。",
    "source_ids": ["S03", "S04"],
    "kind": "inference"
  }
],
"review_items": [
  "既往肾功能及贫血病程",
  "血压、尿量及容量状态",
  "既往病史、用药史及出血相关病史",
  "红细胞指数、网织红细胞及尿检等缺失资料"
]
},
"error": null
}
]

```

单条处理失败时，在结果数组中的对应位置保留以下结构：

```
{
  "case_id": "DEMO001",
  "status": "failed",
  "medical_record": null,
  "error": {
    "code": "MODEL_TIMEOUT",
    "message": "模型请求超时，已达到允许的重试次数。"
  }
}
```

医学内容须遵守以下要求：

- 忠实于给定资料，正确保留重要事实、数值、时序及否定关系，不虚构正常结果或已完成治疗。
- 初步判断、鉴别分析和诊疗计划须有资料依据，并说明信息不足及不确定性。
- 不将“未提供”改写为“正常”或“否认”，不将入院初步判断写成已确诊病因。
- 不把计划写成已实施的检查或治疗，不提前引入后续诊断或住院结局。
- 病例特点和诊疗计划应体现该病例的问题，避免泛化套话；资料缺失及冲突须在待核事项中提示。

6. 执行步骤与异常处理

程序按以下流程执行：

1. 读取 `.env`，校验 `MODEL_NAME` 和 `ENDPOINT`。
2. 读取 `input.json`，校验 JSON 格式和病例标识。
3. 在允许的调用次数及并发范围内处理各条病例。
4. 解析并校验模型结果，整理为约定的病历结构。

5. 写入完整的 `output.json`。

6. 返回退出码并结束运行。

退出码	含义
0	全部输入均已处理，并生成符合协议的输出文件；允许存在单条失败记录
1	发生任务级运行错误，未能完成输出
2	必要配置缺失，或输入文件不符合基本协议

输入无法解析、顶层不是数组、病例标识缺失或重复，按任务级输入错误处理。单条病例的截止点或资料片段无效、模型请求失败或结果无法解析时，应记录该条失败并继续处理后续病例。

结果校验须检查三段正文、依据和待核事项的类型及完整性，并确认 `evidence[].source_ids` 均来自当前病例。模型接口响应与最终结果文件属于不同结构，不能直接把完整 HTTP 响应写入 `output.json`。

程序应先写入同一目录下的临时文件，全部处理完成后再替换为 `output.json`，避免将部分结果误认为完整结果。正常完成空输入时也须写入 `[]`。日志可输出到终端，正式评测以本次成功完成运行后生成的 `output.json` 为准。

7. 赛前技术约束与评测要求

主办方将在赛前统一发布以下详细约束：

类别	需明确的内容
运行环境	操作系统、Python 版本、CPU、内存及网络访问范围
模型服务	实际 <code>MODEL_NAME</code> 、完整 <code>ENDPOINT</code> 、请求及响应协议；接口无需密钥
调用限制	生成参数、每条病例调用预算、并发上限
时间限制	依赖安装时限、单次请求超时、任务总运行时限
重试规则	可重试的错误类型、最大次数，以及重试是否计入预算
数据协议	截止点表示方式、资料片段格式、首次病程输出字段及缺失值规则
评测规则	格式检查、内容质量、失败记录及超时任务的计分方式

参赛者可设计 Prompt、事实提取、时间整理、分段生成和生成后校验等处理流程。使用检索资料时，须限于赛方提供或赛前登记的允许资料，固定版本

并附来源清单。正式评测仅允许使用指定模型和主办方允许的服务，不临时联网搜索病例答案，不使用预先缓存的测试答案，也不得绕过调用限制。

正式评测使用同一提交版本批量处理全部病例。可对全部病例一致执行字段解析、格式校验和规则检查，不得逐例人工改写医学正文、调整提示词或挑选某次最好结果。服务错误按赛方统一重试规则处理。

提交前应确认：源码包能够在干净环境中安装依赖，通过 `python main.py` 自动运行，从 `.env` 读取模型名称和完整请求地址，在无需密钥的情况下调用模型，并生成可解析、条数一致、病例标识正确的 `output.json`。