

基于大语言模型的慢性病患者随访管理系统

一、项目背景及现有技术综述

1. 项目背景

随着我国人口步入老龄化，各种慢性病的发病率逐渐升高。根据国家统计局的数据，2022年末，我国60岁及以上人口为28004万人，占全国人口的19.8%，相比2021年增加1268万人，比重上升0.9个百分点，而65岁及以上人口为20978万人，占全国人口14.9%，增加922万人，比重上升0.7个百分点，老龄化程度加深^[1]。伴随老龄化，我国的流行病学已经从传染病转向慢性非传染性疾病。《2017年中国卫生和计划生育统计年鉴》数据显示，从2003年到2013年十年间，我国慢性病患病率增长近2倍，2012年全国居民慢性病死亡率为533/10万，占总死亡人数86.6%，其中心脑血管病、癌症和慢性呼吸系统疾病是主要死因，占总死亡的79.4%，慢性病已经成为我国居民健康的头号杀手^[2]。2018年，我国慢性病患病率已达到34.29%。据统计，截至2019年，慢性病导致的疾病负担占总疾病负担的近70%，导致的死亡占我国总死亡人数的88%^[3]。同时，糖尿病等多种慢性病呈现出年轻化的发病趋势，对人民的身体健康和生活质量构成越来越大的挑战。在国家政策层面也越来越重视慢性疾病的管理，在2016和2017年相继发布的《国家慢性病综合防控示范区建设管理办法》和《中国防治慢性病中长期规划（2017-2025年）》都对慢病提出了要求。

慢性病健康管理（health management for chronic diseases）是指在收集个人健康信息的基础上，对个体未来一定时间内某种慢性病的发生风险进行预测。在风险预测的基础上，针对生活方式和危险因素制定个体化干预和行为校正计划并实施，定期进行跟踪和效果评估。在效果评估的基础上进一步收集信息，然后进入下一个循环。在个体健康管理的基础上，也可对服务人群信息进行汇总和分析，并对人群的慢性病预防、治疗和管理工作提出建议、指导和咨询^[4]。

然而我国当前在慢性病管理方面面临诸多困难。大量医疗资源投入集中在急性疾病和重症，导致没有足够的医疗资源用于持续跟踪慢性病患者的健康情况，使得原本严峻的医疗资源紧张问题在慢性病领域更加突出。同时由于医生精力的限制和慢性病本身病程长的特点，患者在院外期间需要有一定质量的健康监测与管理服务，因此近年市场上也涌现出各类健康管理APP和智能穿戴设备。但由于缺少医生的直接参与，单纯通过线上APP的跟踪、教育和激励对患者的依从性提升效果不明显，另一方面用户的支付意愿很低，无法持续运营从而形成良性循环。Latitude Health在2016年进行的调研显示只有28%的用户愿意为慢病管理付费，而慢病管理服务又因其效果难以定量分析、与医保控费的直接经济联系不明确等因素导致其很难纳入报销范围。因此，充分利用现有技术手段，提高慢病管理的服务质量，并从人文关怀角度让医生更多地参与进来，有利于提升现有的慢病管理水平。

2022年年底以来，随着越来越多以ChatGPT为代表的人工智能大语言模型（Large Language Model, LLM）面向公众开放，除了为人们展现出人工智能酷似真人的强大聊天功能外，也为包括医学在内的许多领域带来潜在的颠覆性变革。已有研究指出GPT-4这样的大语言模型可以用于作为医学聊天机器人完成为患者提供“非正式医疗咨询”、撰写医疗信息记录、作为医疗百科全书以及协助分析医学研究资料等工作^[5]。而随着Google开发的Med-PaLM通过美国医学执照考试（USMLE），并在回答医学问题的准确性与潜在危害程度等多个方面接近甚至超过人类医生，人工智能大语言模型的专

业性与可靠性得到了进一步证实。尽管出于伦理等多种因素人工智能还不能完全取代医生的工作，但在慢病管理方面以 ChatGPT 为首的大语言模型有广阔的应用空间。人工智能大语言模型由于其自身的专业性和“类似”人类的表达方式使其可以在慢性病患者随访工作中部分代替医生的工作，起到辅助作用，结合提示工程技术（Prompt Engineering），合理利用大语言模型代替医生对随访患者定期“问诊”，同时将获得的信息以规范化记录的形式反馈给医生，并为患者提供复诊建议，为医生提供风险评估参考，一方面能够大大减轻医生的工作量，另一方面也能为慢病患者提供更优质的院外健康服务，提升慢病管理水平。

2. 现有技术综述

为将人工智能更好地应用于慢性病患者的随访管理，本项目可能涉及到的技术包括：以 ChatGPT 为代表的大语言模型（LLM）和给予深度学习的疾病风险预测模型，下面分别概述其在医学领域的应用现状。

（1）以 ChatGPT 为代表的大语言模型

大语言模型（Large Language Model, LLM）是基于人工智能神经网络开发的自然语言处理模型，旨在理解和生成人类语言，可以执行包括文本总结、翻译、情感分析等多种任务。大语言模型由预训练语言模型（Pre-trained language models, PLM）发展而来，主要区别在于增大了模型的规模，参数量可达到十亿（billion）级别，科学家在研究中发现将语言模型不断放大达到一定规模后，其性能得到了显著的飞跃，并出现了小规模预训练语言模型所不具备的诸如上下文学习的能力（in-context learning），因此学界提出了大语言模型（LLM）的概念以区分前者^[6]。大语言模型的训练过程主要包括了预训练（pre-training）和针对任务的微调（fine-tuning），主要是通过无监督学习（unsupervised learning）完成的。目前已有很多面向大众开放的大语言模型，包括 OpenAI 的 ChatGPT、Google 的 PaLM、清华大学 ChatGLM 的开源版本 ChatGLM-6B 等。

以 ChatGPT 为例，现有的大语言模型虽然功能强大，然而由于其不是为循证医学领域设计和训练的，在将其运用到医学领域仍然面临一些列问题与挑战，主要来自于以下四个方面^[7]：

a. 临床证据的检索：

循证医学对纳入研究的临床证据有严格的标准，因此筛选符合“PICO（患者、干预、对照、结果）”框架的随机对照试验（RCT）十分重要。大语言模型由于其大批量文本处理与生成的优势，用于循证医学证据的筛选能大大提高筛选效率，然而其可靠性还有待进一步的研究评估。基于大语言模型的临床证据检索还需要与传统语言模型以及人工提取评估进行对比，同时也需要在不同的医学领域内进行效果评估以确定其是否可靠。

b. 临床证据的综合：

即将多个研究的临床证据整合分析从而得出有依据的临床结论，从而进一步指导医疗实践的过程。随着医学的发展，临床研究的数目呈现“爆炸式”增长，临床医生通过传统的文献搜索方式很难跟上最新的医学研究，而大语言模型在未来可能会成为综合大量临床研究对复杂临床问题给出解答的有力工具。然而大语言模型给出的答案可能会有过时、自我矛盾或不准确的缺陷，而且其不具有随新研究被发表而及时学习、更新自我的能力，大语言模型的每次更新都意味着需要进行重新训练。此外，大语言模型目前不能很好地区分新研究和过时的研究，未来的大模型需要能够

具备时间推理能力，能够意识到临床证据的不断演进，能够在例如新诊断指南代替旧指南时给出合理的结果。同时，医生在综合分析临床证据时也要合理评估其适用性和可靠性。

c. 临床病例的医学推理：

大语言模型能够用逻辑清晰且易懂的语言回答问题，这一特点使得其在进行临床推理、分析临床问题时具有得天独厚的优势。尽管有证据表明大语言模型可以进行推理，但未来的研究需要将基于知识的推理能力引入模型中，使得其具有在复杂问题中进行有力且可解释的推理能力，这将有利于其在循证医学中的应用。同时我们也亟需一套标准来评价大语言模型的推理能力，评估其是否能像专业医生一样思考。

d. 临床证据的传播科普：

由于医学本身的专业性，导致患者和公众没能更多地主动参与到医疗过程中，但大语言模型能将复杂、专业的问题用通俗语言解释的能力使得其可以缩小患者与一生的信息差，有利于将最新、最前沿的临床研究成果及时在公众中普及，这对于推动医疗水平进步和改善医患关系都有帮助。但由于现有的大语言模型提供的信息可能会忽略部分事实、模糊不清甚至具有误导性，对于 AI 生成的内容还需要有相关专家进行严格的审查与修正。

此外，为了让其能在我们的应用场景中充分发挥作用，需要用到提示工程技术（Prompt Engineering Techniques）。提示工程，就是创建一系列指令，提示（询问、指导）ChatGPT 这类语言模型输出语料文本。提示工程技术能帮助用户控制语言模型输出，让其生成符合用户需求的、高质量的文本^[8]。

（2）基于深度学习的疾病风险预测模型

基于深度学习的疾病风险预测模型是主要利用卷积神经网络（CNN）和循环神经网络（RNN）等深度学习算法，从大量的临床数据和医学影像数据中提取有意义的特征，从而分析个体的生物数据，以预测患病风险的一种方法。这些模型通常使用神经网络架构，可以从大规模的数据集中学习复杂的模式和关联性，以更准确地评估患病风险。

其主要流程包括：

- a. 数据预处理：对原始数据进行清洗和标准化处理，剔除错误或无效的数据。
- b. 特征提取：利用深度学习算法从处理后的数据中提取有意义的特征。
- c. 模型训练：使用监督学习算法对提取的特征进行训练和优化，建立疾病风险预测模型。
- d. 预测：利用已训练的模型对新的患者数据进行分析和预测，得出该患者患病风险的评估。

其技术特点主要包括：

- a. 自适应性强：基于深度学习的疾病风险预测模型可以自适应不同的疾病类型和数据特征，提高模型的预测精度和可靠性。
- b. 决策支持能力强：该模型可以为医生提供辅助决策支持和个性化治疗方案定制，帮助医生更加有效地管理患者的健康信息。
- c. 处理速度快：基于深度学习的算法具有并行计算和高效处理大规模数据的能力，因

此可以在较短时间内对大量患者数据进行分析 and 预测。

- d. 精度高：由于深度学习算法具有强大的特征提取和模式识别能力，所以基于深度学习的疾病风险预测模型具有较高的预测精度和准确度。

目前，基于深度学习的疾病风险预测模型已经在多个医疗健康领域得到应用，如肺癌、糖尿病等疾病的风险预测和预后评估。此外，该技术还可以应用于计算机辅助诊断系统中，帮助医生更加准确地进行疾病诊断和治疗。

二、本项目目标、预计关键技术难点及解决方案

本项目拟利用 ChatGPT 和其他慢性病风险预测人工智能模型，以糖尿病为例，主要面向糖尿病患者构建开发一款集成患者随访问诊、医疗咨询、为医生提供随访记录和风险评估、整合患者医疗信息的应用程序（APP），并基于此尝试构建通用型的慢性病患者随访管理系统，主要分为三个模块。

模块 1：与患者沟通询问近期情况

一、核心依赖

1、chatGPT

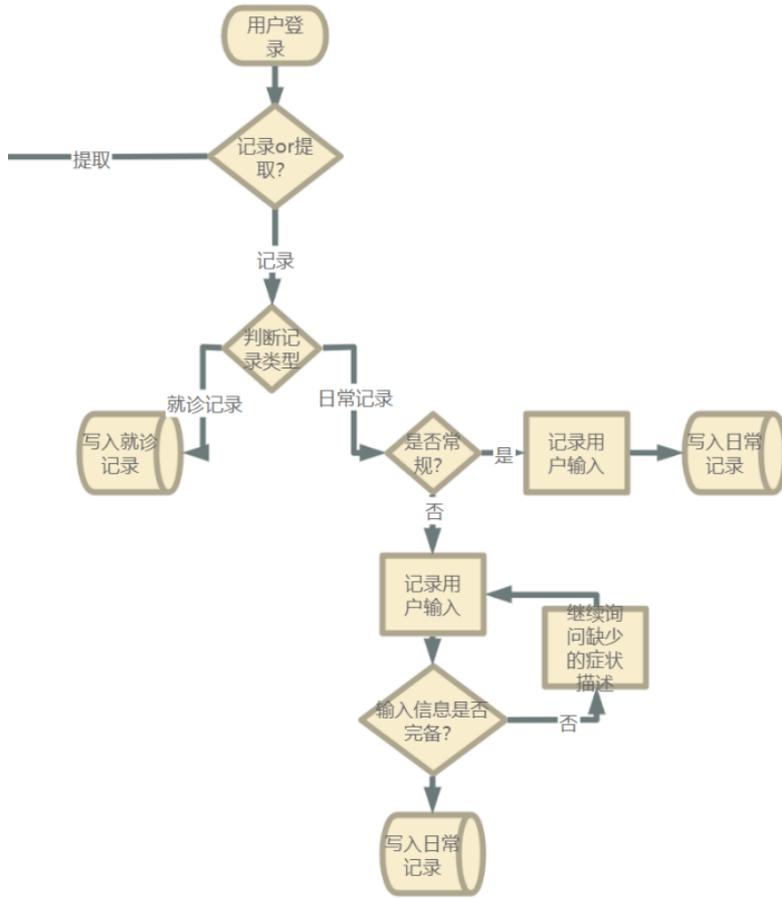
ChatGPT 是一种大型自然语言处理模型，是由 OpenAI 开发的基于深度学习的人工智能系统。它基于生成对抗网络 (GAN) 和自注意力机制 (transformer) 的模型架构，并使用海量语料数据进行预训练，可以生成人类类似的自然语言对话、生成文本和理解语言中的语义。ChatGPT 在 NLP 的多个任务中展现出了出色的表现，在翻译、问答、文本生成等领域都取得了显著成果。

2、MySQL 数据库

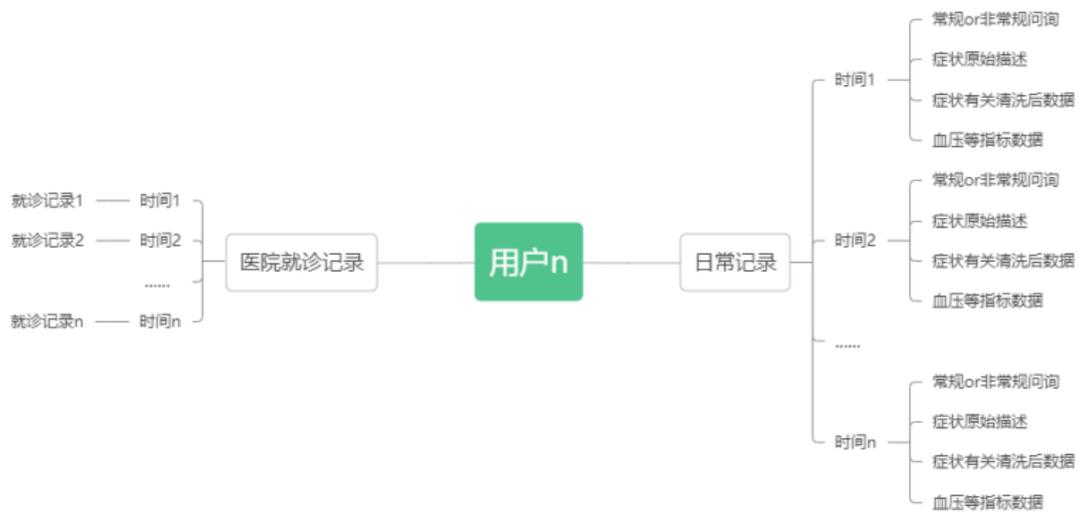
MySQL 是一种常见的关系型数据库，它可以在 Windows、Linux 和 macOS 等操作系统上运行。关系型数据库是一种基于关系模型的数据库，它将数据组织为表格形式，并使用关系来描述数据之间的联系。每个表格都包含若干个列和行，每个列代表一个属性，每个行代表一个记录。关系型数据库使用 SQL (Structured Query Language) 语言进行数据查询和操作。

二、具体操作

流程如下：



1、建立以用户个人 id 为标识的数据库。数据库结构如下：



其中，就诊记录包括病例记录以及本次检查获得的报告单。

- 2、与患者进行交流，交流方式分为两种：一是患者主动进行交流，例如糖尿病患者甲某日觉得头晕，主动向软件描述相关症状，用户在进行主动描述后，系统根据所需信息，主动向用户询问缺少的信息，例如“您今天具体进食情况怎么样？”完善相关情况的记录。二是软件定时向用户进行主动询问，根据不同病症设定不同的询问指标，并通过 GPT 转换为易于患者理解的询问语句，询问后获取用户的自然语言描述。同时，也可提前设定时间提醒患者服药，并记录服药时间。
- 3、获得用户回应之后，先将用户的自然语言完整存储下来，以便为后续进行检查核验提供原始依据。
- 4、对用户传入的自然语言进行数据清洗，转换成表格，例如：

8 请将“我今天有些头晕，觉得伤口有些疼”这句话提取有用的信息并转换成表格

8 如果将“有些”定义为2级程度呢

信息	描述
主体	我
状态	有些头晕，觉得伤口有些疼
时间	今天
症状/疼痛部位	头部和伤口
程度	2级

- 5、同时，通过外接设备传入的数据，将血压、血糖、血脂、心率等相关数据也添加至数据库中。

模块 2：处理信息反馈给医生的模块，同时生成标准化的医疗记录

一、核心依赖

1、Numpy 和 Pandas

NumPy (Numerical Python) 是一个开源 Python 库，几乎用于科学和工程的每个领域。它是在 Python 中处理数值数据的通用标准，也是科学 Python 和 PyData 生态系统的核心。NumPy 用户包括从初级编码员到从事最先进的科学和工业研发的经验丰富的研究人员的所有人。NumPy API 广泛用于 Pandas、SciPy、Matplotlib、scikit-learn、scikit-image 和大多数其他数据科学和科学 Python 包。

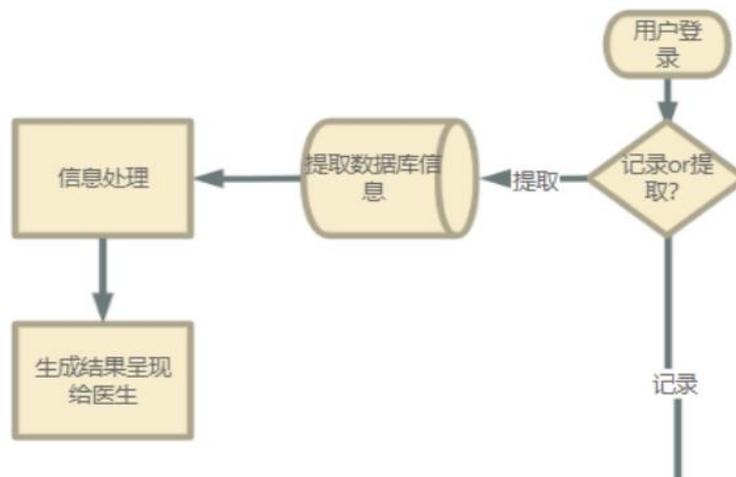
Pandas 是一个 Python 数据处理和数据分析库，提供了用于快速、灵活、易于操作和可理解的数据结构，可以处理各种数据类型（如：关系型和标签型数据）。Pandas 内置了许多处理数据的函数和方法，可以帮助用户进行数据清洗、数据重塑、数据切片、排序、聚合和统计等操作。

2、Matplotlib

Matplotlib 是一个 Python 数据可视化库，能够通过各种图表和图形方式展示和呈现数据。它提供了一系列用于绘制图形的函数和工具，例如：线型图、散点图、直方图、条形图、饼图等。Matplotlib 的可视化工具简单易用，能够满足数据可视化的各种需求，能够通过设置不同的参数和样式，对图表进行完全自定义的设计和调整，支持将图表以多种文件类型（如 png、svg、pdf 等）进行输出和存储。Matplotlib 是数据科学领域必不可少的一个工具库，它可以帮助数据科学家将数据转化为更直观、更易于理解、更具可表达力的可视化形式。

二、具体操作

流程：



- 1、读取数据库所有信息。
- 2、对于血压、血糖、血脂、心率等需要转换成数据的常规指标，将各指标数据做成直观的图表，并对超出正常范围的进行特殊标记。
- 3、对于文字表达的数据，可将其按照时间顺序，展示在图表数据之后，对以程度较重的症状进行特别标注。
- 4、对以往的就诊记录进行查询，并且按照就诊时间顺序展示医生结论，当前医生可以选择查看过往病例记录的电子版。

模块 3：结合专业的风险预测模型，预测风险并对患者提出复诊建议，给医生作为病情评估的参考

一、核心依赖

1. 逻辑回归模型

逻辑回归是一种用于分类问题的机器学习模型，通过将线性回归模型的输出经过 sigmoid 函数映射到 0 和 1 之间的概率值，进行二分类预测。它基于最大似然估计，通过最小化对数损失函数来学习模型参数。逻辑回归模型简单、高效，并且可解释性强，常用于二分类任务，如疾病预测、信用评分等。

2. ARIMA 模型

ARIMA 模型（自回归移动平均模型）是一种经典的时间序列模型，用于分析和预测具有一定规律性的数据。它基于时间序列数据的自相关和移动平均性质，通过差分操作将非平稳数据转化为平稳数据，然后使用自回归和移动平均的组合来建立模型。ARIMA 模型在疾病监测方面可以应用于分析病情的趋势、周期性变化等，提供疾病的预测和预警功能。

二、具体操作

以糖尿病预测和监测为例，下面介绍具体流程：

1. 糖尿病的预测

- (1) 数据收集和准备：收集包含糖尿病相关特征和标签（是否患有糖尿病）的数据集。确保数据集的质量和完整性，处理缺失值和异常值，并进行数据预处理步骤，如特征缩放和归一化。
- (2) 数据拆分：将数据集划分为训练集和测试集，通常采用随机拆分的方式。训练集用于模型训练，测试集用于评估模型性能。
- (3) 特征选择和提取：对于逻辑回归模型，选择适当的特征对预测效果至关重要。可以使用统计方法、领域知识或特征重要性评估等方法进行特征选择。另外，还可以进行特征工程，从原始特征中提取更有信息量的特征。
- (4) 模型训练：使用训练集进行逻辑回归模型的训练。在训练过程中，通过最大化似然函数或最小化对数损失函数，调整模型的权重和偏差，以使模型能够更好地拟合训练数据。
- (5) 模型评估：使用测试集评估模型的性能。常用的评估指标包括准确率、精确率、召回率、F1 值等。这些指标可以帮助评估模型的分类能力和预测准确性。
- (6) 结果解释：分析模型的权重和系数，解释每个特征对糖尿病预测的贡献程度。这有助于理解哪些特征对预测糖尿病起着重要作用。

附：

以下是一个简单的示例来展示逻辑回归模型的输入和输出：

输入特征：

- 年龄：45 岁
- 体重指数 (BMI)：29.5
- 血压：130/80
- 血糖：150 mg/dL
- 家族糖尿病史：有

输出标签：

- 是否患有糖尿病：是

在这个例子中，我们使用逻辑回归模型对这些输入特征进行预测，以确定该人是否患有糖尿病。模型将根据训练数据学习到的权重和偏差进行预测，并输出一个概率值。

预测输出：

- 预测患糖尿病的概率：0.78

根据模型的预测概率，我们可以将其与设定的阈值进行比较，例如 0.5。如果预测概率大于阈值，则判定为患糖尿病高风险；如果小于等于阈值，则判定为患糖尿病低风险。

上述仅为一个简单的示例，实际上，逻辑回归模型可能使用更多特征，并在更大规模的数据集上进行训练和预测，以取得更准确的结果。

2. 糖尿病患者的监测

- (1) 数据收集和准备：收集糖尿病患者的监测数据，包括时间戳和相关指标（如血糖水平）。确保数据的时间顺序和完整性。
- (2) 数据预处理：对数据进行必要的预处理，如缺失值处理、异常值检测和平滑处理。可以使用时间序列平滑技术（如移动平均、指数平滑）来减小数据的噪音和波动。
- (3) 模型训练：使用 ARIMA 模型进行训练。ARIMA 模型是一种基于时间序列的统计模型，包括自回归（AR）、差分（I）和移动平均（MA）三个部分。可以根据数据的特点和需求选择合适的 ARIMA 模型阶数。
- (4) 模型预测与评估：使用训练好的 ARIMA 模型进行糖尿病健康状况的预测。输入当前的监测数据序列，模型输出未来一段时间内的糖尿病状况预测。可以使用评估指标（如均方根误差、平均绝对误差）来评估模型的预测性能。
- (5) 提供建议与反馈：根据模型的预测结果，给出相应的建议和指导。例如，如果模型预测出糖尿病状况不稳定或存在风险，可以建议患者前去复诊、调整饮食或咨询医生。

需要注意的是，ARIMA 模型假设数据具有平稳性，对于非平稳的数据可能需要进行差分或其他预处理方法。

三、本项目执行计划

项目安排主要分为三个阶段，如下表所示：

时间	内容
2023. 5. 28-2023. 6. 15	依据项目框架，基于 Android 开发搭建软件框架，划分模块。
2023. 6. 15-2023. 7. 31	学习并利用提示工程技术，针对面向患者和面向医生两大应用场景优化与大语言模型的对话效果，并封装入相应模块。同时查阅资料开发糖尿病患者预后风险相

	关模型，利用开源数据库训练，最终封装入对应模块。并设计前端用户界面。
2023. 8. 1-2023. 9. 1	将相关模块整合，完成完整的应用程序开发。在内部进行测试与修改，然后进一步在小样本医生与患者中进行初步测试与进一步优化。

四、本项目预算（不超过 5 千元）

支出门类	金额（单位：元）	测算依据
文献检索	500	文献数据库获取相关文献资料
软件开发	2000	购买软件开发工具套件，进行项目软件的开发与调试
软件测试	2000	用于软件的小范围测试与优化

参考文献：

- [1] 王萍萍：人口总量略有下降 城镇化水平继续提高, 2023. http://www.stats.gov.cn/xxgk/jd/sjjd2020/202301/t20230118_1892285.html#:~:text=2022%E5%B9%B4%E6%9C%AB%EF%BC%8C%E5%85%A8%E5%9B%BD0%E2%80%9415%E5%B2%81%E4%BA%BA%E5%8F%A3%E4%B8%BA25615%E4%B8%87%E4%BA%BA%EF%BC%8C%E5%8D%A0%E5%85%A8%E5%9B%BD%E4%BA%BA%E5%8F%A3%E7%9A%8418.1%25%EF%BC%9B16%E2%80%9459%E5%B2%81%E5%8A%B3%E5%8A%A8%E5%B9%B4%E9%BE%84%E4%BA%BA%E5%8F%A3%E4%B8%BA87556%E4%B8%87%E4%BA%BA%EF%BC%8C%E5%8D%A062.0%25%EF%BC%9B60%E5%B2%81%E5%8F%8A%E4%BB%A5%E4%B8%8A%E4%BA%BA%E5%8F%A3%E4%B8%BA28004,%E4%B8%87%E4%BA%BA%EF%BC%8C%E5%8D%A019.8%25%EF%BC%8C%E5%85%B6%E4%B8%AD65%E5%B2%81%E5%8F%8A%E4%BB%A5%E4%B8%8A%E4%BA%BA%E5%8F%A3%E4%B8%BA20978%E4%B8%87%E4%BA%BA%EF%BC%8C%E5%8D%A014.9%25%E3%80%82.
- [2] 中国慢病管理现状与国内外管理模式探索, 2018. <https://www.cn-healthcare.com/articlewm/20181017/content-1036294.html>.
- [3] 国务院关于实施健康中国行动的意见, 2019. http://www.gov.cn/zhengce/content/2019-07/15/content_5409492.htm.
- [4] 中国健康管理协会. 慢性病健康管理规范 (T/CHAA 007-2019) [J]. 中华预防医学杂志, 2020, (1): 39-41.

- [5] LEE P, BUBECK S, PETRO J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine [J]. The New England Journal of Medicine, 2023, 388(13): 1233-9.
- [6] WAYNE, ZHOU K, LI J, et al. A Survey of Large Language Models [J]. arXiv pre-print server, 2023.
- [7] PENG Y, ROUSSEAU J F, SHORTLIFFE E H, et al. AI-generated text may have a role in evidence-based medicine [J]. Nat Med, 2023.
- [8] JOHN I. The Art of Asking ChatGPT for High-Quality Answers: A Complete Guide to Prompt Engineering Techniques [M]. 2023.