• 医学信息技术 •

基于预训练模型的医学习题解析半自动生成方法研究

孙月萍[1]　王　娟[1]　董良广[2]　刘　燕[1]　杨　丽[1]　李　姣[1]　侯　丽[1]

（[1] 中国医学科学院/北京协和医学院医学信息研究所　北京　100020
[2] 人民卫生出版社有限公司　北京　100021）

〔摘要〕　**目的/意义**　探索基于预训练语言模型的半自动化解决方案，提高医学习题解析生成效率与质量。**方法/过程**　引入基于MC-BERT的混合智能增强框架，先自动完成题目结构识别、知识点抽取，生成初步解析，再通过人工校验，严格把控内容的准确性与规范性，形成可追溯的解析语料。**结果/结论**　该方法能够显著提高解析生成效率，降低人工成本，同时保障解析内容与医学教学大纲、教材知识体系的一致性和可追溯性，为医学教育智能化提供了可行路径。

〔关键词〕　预训练模型；医学习题解析；半自动化生成；解析推荐
〔中图分类号〕R-058　　〔文献标识码〕A　　〔DOI〕10.3969/j.issn.1673-6036.2026.02.011

Study on a Semi-automatic Generation Method for Medical Exercise Answer Explanations Based on the Pre-trained Model

SUN Yueping[1], WANG Juan[1], DONG Liangguang[2], LIU Yan[1], YANG Li[1], LI Jiao[1], HOU Li[1]

[1] Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China;
[2] People's Medical Publishing House Co. Ltd., Beijing 100021, China

〔**Abstract**〕　**Purpose/Significance**　To explore a semi-automated approach based on pre-trained language models, and to enhance both the efficiency and quality of medical exercise answer explanation generation. **Method/Process**　A hybrid intelligence framework based on MC-BERT is introduced. Firstly, the question structure recognition, key knowledge point extraction and draft explanation formulation are automatically completed. Then, through manual verification, the accuracy and standardization of the content are strictly controlled to form a traceable reference explanation corpus. **Result/Conclusion**　The method significantly improves generation efficiency, reduces manual intervention costs, while ensuring the consistency and traceability of the explanation content with the medical teaching syllabus and textbook knowledge system, providing a feasible path for the intelligentization of medical education.

〔**Keywords**〕　pre-trained model; medical exercise answer explanation; semi-automatic generation; explanation recommendation

〔修回日期〕　2026-01-15

〔作者简介〕　孙月萍，副研究员，发表论文20余篇；通信作者：侯丽，研究员。

〔基金项目〕　国家社会科学基金青年项目（项目编号：18CTQ024，22CTQ024）；医学融合出版知识技术重点实验室项目（项目编号：23RW1201）。

1 引言

在医学教育中，高质量的习题解析有助于评估学习效果、巩固知识体系。此类解析对准确性、规范性与可追溯性要求较高，传统人工编写模式存在效率低、成本高且一致性难保证等固有局限。基于BERT等预训练语言模型的“预训练-微调”范式[1-2]，在生物医学领域文本表示和知识挖掘等任务中表现较好[3-4]，但其在医学解析生成领域的应用仍面临严峻挑战，这主要源于医学文本中密集的专业术语和严苛的逻辑严谨性要求。为此，本研究提出一种基于预训练模型的医学习题解析推荐框架，在提升解析生成效率与准确性的同时，降低对人工标注的依赖。

2 相关研究

2.1 医学预训练语言模型

近年来医学预训练模型在自然语言处理（natural language processing，NLP）和多模态学习领域取得了显著进展。其通过在PubMed文献、临床电子病历等大规模生物医学文本数据上开展预训练，再针对下游任务微调，显著提升了医疗信息处理、辅助诊断和科研效率。基于Transformer的领域自适应预训练模型，如BioBERT[5]和ClinicalBERT[6]，能更精准地捕捉医学术语的复杂语境与同义关系，在命名实体识别（named entity recognition，NER）和关系抽取（relation extraction，RE）等任务中表现优于通用领域嵌入模型。随着研究深入，训练策略与数据规模得到进一步拓展。PubMedBERT[7]专攻PubMed抽象文本，在多项医学NLP任务（如NER、RE）达到了最先进的性能。BioGPT[8]基于生成式架构，在医学文本生成和问答中表现突出。Med-PaLM[9]和Med-PaLM 2[10]在美国医师执照考试问答上达到“专家”水平。LLaMA-Med[11]基于LLaMA在生物医学文献上进行继续预训练和指令微调，提高了模型遵循医学指令和进行推理的能力。此外，Zhang N等[12]提出的MC-BERT通过全实体遮蔽和医学知识图谱增强，显著提升了中文医学文本的语义表征能力，在NER、问答等任务上优于通用BERT模型。类似地，赵一鸣等[13]提出的MQ-BERT通过增量预训练，融合医学知识，在意图识别任务上的$F1$值达到95.34%，较MC-BERT提升7.84个百分点。此外，MedBert[14]参照BioBERT的思路，在医学文本（如临床病历、医学文献、医学百科）上继续预训练。SMedBERT[15]通过在相关实体的邻接节点中融入深度结构化语义知识来增强模型对医学语义的理解。

目前医学预训练模型（如BioBERT、PubMedBERT）多采用标准随机掩码语言建模，随机遮蔽单个词汇进行预测。这种策略虽然有效，但在医学解析生成任务中存在明显局限：医学概念（如“冠状动脉粥样硬化”）作为完整语义单元，随机遮蔽会破坏概念完整性，导致模型学习的是词汇层面的共现关系而非概念层面的逻辑关联。MC-BERT采用全实体遮蔽策略，通过实体链接识别完整医学概念后统一遮蔽，迫使模型从上下文整体语义推理被遮蔽实体间的医学逻辑关系。这种设计直接适配医学解析生成所需的概念完整性保持能力。

通用医学预训练模型（如Med-PaLM）主要从大规模文本中隐式学习知识关联，属于“统计驱动”模式，然而医学解析生成要求严格遵循教材知识体系的逻辑严谨性。MC-BERT通过知识图谱增强，在预训练中显式融入结构化医学知识，使模型不仅学习语言模式，更内化了医学知识体系的结构化关系，与教材知识体系深度绑定。因此，相较于其他医学预训练模型，MC-BERT的预训练任务与问题-答案自然语言推理（question-answering natural language inference，QNLI）结合度更高。

2.2 半自动语义标注

半自动标注方法结合机器学习模型与人工校验，能够显著提升效率。近年来基于Transformer架构的预训练语言模型及其在生物医学领域适配的预

训练模型（如 BioBERT、ClinicalBERT）的崛起，极大地推动了半自动标注的效率和精度。

预训练模型能够有效减少人工标注员在烦琐、重复性工作上的时间消耗。例如，Lee J 等[5]在构建生物医学命名实体识别基准时，利用 BioBERT 模型对 PubMed 摘要进行初步实体标注，人工标注员仅需专注于补充或修正模型遗漏或错误的实体。研究表明，这种方法比从零开始的全人工标注效率更高，同时因为模型提供了一个相对一致的标注基线，也显著减少了不同标注员之间的主观偏差。领域特定的预训练模型如 ClinicalBERT，通过在 MIMIC-III 等大型临床电子病历语料库上继续预训练，学习临床记录中的语言风格和术语分布。在临床概念（如药物、剂量、症状等）提取任务中，这类模型能够更准确地识别复杂信息（如“每天两次口服 500mg 阿司匹林”），为半自动标注系统提供了更可靠的候选答案，大幅降低了人工修正负担。

在标注资源极其稀缺的特定医学子领域（如罕见病），预训练模型的少样本学习能力尤为重要。通过提示学习或适配器等参数高效微调技术，模型可以快速适应新定义的实体类型。研究者通常将其与主动学习循环结合：模型对未标注数据进行预测并计算不确定性，筛选出最“困惑”或最具信息量的样本优先交由专家标注，之后将这些新增的标注数据加入训练集以迭代优化模型[7]。这种策略实现了标注资源的最优分配，以最小的人工成本获得最大化的模型性能提升。另有研究[16]针对医学领域涉及大量专业术语和复杂表述方式，传统匹配模型难以达到较高准确率的问题，提出语义召回加精准排序的两阶段模型，以提升医学术语标准化效果。

在多模态与知识增强方面，有研究开始探索超越纯文本的预训练模型。例如，CheXbert[17]结合胸部 X 线报告的图像编码器和文本编码器，半自动标注放射学报告中的观察结果（如“肺不张”“心脏肥大”）。这种多模态方法确保了文本标注与影像表现的一致性。此外，如 Bio-LM 等模型尝试将医学知识图谱（如 UMLS）融入预训练过程，使模型生成的标注结果不仅在文本层面准确，而且更符合医学知识逻辑，进一步提升了半自动标注输出的质量。此外，部分研究[18]探索利用跨模态信息相互增强的策略。

医学预训练模型在语义理解与知识推理方面展现出显著潜力，半自动语义标注方法为高质量医学文本处理提供了可行技术路径，但既有研究仍存在明显局限性。其多聚焦基础性信息抽取与术语标准化任务，尚未充分探索如何将预训练模型的深层语义表征能力与半自动标注的高效性相结合，以完成医学习题解析生成这一更具复杂性的任务。因此，本研究提出一种基于预训练模型的半自动医学习题参考答案解析推荐方法，通过融合知识增强的语义编码机制与交互式标注策略，协同优化解析内容在准确性、深层语义理解和生成效率 3 个维度的性能。

3 基于预训练模型的医学习题参考解析推荐框架

3.1 总体思路

针对医学解析生成对概念完整性和逻辑严谨性的特殊要求，MC-BERT 的全实体遮蔽策略将医学概念作为完整单元处理，能够确保模型学习概念层面的逻辑关系；同时通过知识图谱增强显式融入结构化医学知识，使模型学习教材知识体系的内在逻辑。提出基于 MC-BERT 模型的医学参考答案解析推荐框架，结合预训练语言模型与自然语言推理任务，实现医学习题解析的半自动化生成，见图 1。MC-BERT 模型使用多源中文生物医学语料进行自监督预训练，学习领域特异性词嵌入。微调阶段：以社区问答和医学教科书等语料对模型进行领域适配微调，优化语义表示。QNLI 任务构建阶段：将医学习题的问题-答案对与从教科书抽取的候选解析组合为处理单元，通过模型计算语义匹配得分，筛选高评分解析形成最终答案。经人工标注，将上述输出构建为结构化数据集 CMedBookQNLI。

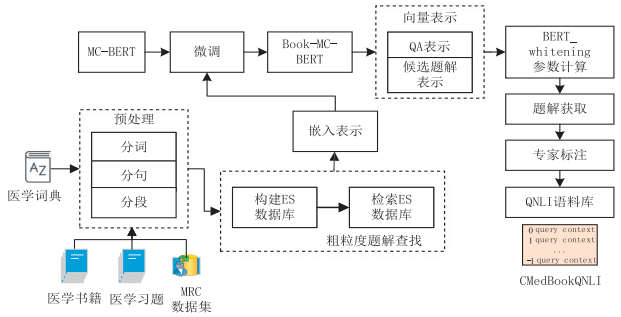

图 1　基于预训练模型的医学参考答案解析推荐框架

3.2　任务定义

给定一对输入句子$\{X,Y\}$，$X=\{x_1,x_2,\cdots,x_N\}$，$Y=\{y_1,y_2,\cdots,y_M\}$，其中X代表由问题和候选答案构建的查询，Y代表从教科书中检索到的候选解析上下文，x_i是查询的第i个词/标记，y_i是上下文的第i个词/标记，N和M分别代表查询和上下文的长度。分类的目标是判断X和Y是否彼此相关。将该任务视为机器阅读理解任务，从而将标注风格的数据集转换为一组（查询，上下文，答案）三元组。其中，查询是给定的输入句子X，上下文是根据句子X抽取的候选解析句子，而答案是查询与上下文之间的目标关系。

3.3　模型微调

采用基于BERT的模型架构，并以专门针对生物医学领域预训练的MC-BERT参数进行初始化。使用MC-BERT在机器阅读理解（machine reading comprehension，MRC）框架中实现分类任务，见图2。首先，连接查询X与上下文Y，形成组合序列，其中[CLS]和[SEP]是用于标记开始和分隔的特殊标记。其次，将组合序列输入MC-BERT，主要通过公式（1）—（3）定义。其中L表示MC-BERT的总层数，l是当前层的编号，Trm表示Transformer块，包括多头注意力层、全连接层和归一化层。此外，H是MC-BERT的输出，$N+M$是查询和上下文的总长度。最后，执行一个二元分类器进行模型预测。K表示分类类别数，z_k表示第k个类的逻辑值（从H中获得的相似度值）。

$$h_i^0 = W_e t_i + W_b \tag{1}$$

$$h_i^l = Trm(h_i^{l-1}) \tag{2}$$

$$H = [h_1^L, h_2^L, \cdots, h_{N+M}^L] \tag{3}$$

$$Softmax(z_k) = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \tag{4}$$

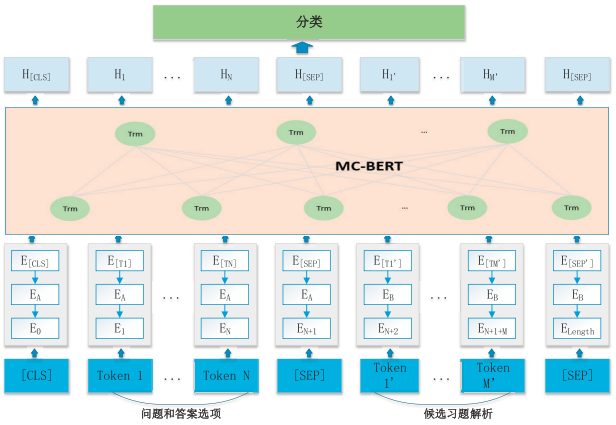

图 2　基于MC-BERT的QNLI任务模型结构

注：T代表词元Token，E代表嵌入式表示Embedding，Trm代表Transformer。

微调训练过程具体如下：给定初始化的模型参数θ及预设的学习率λ，先对输入数据进行预处理，将医学教科书及问答文本输入MC-BERT编码器，得到语义向量$H=[h_1,h_2,\cdots,h_n]$，作为新的输入X。随后计算损失函数$L_E(\theta)=-y\log p(y|x;\theta)$，$P(y|x)$代表预测标签为$y$的概率。最后基于该损失函数对参数$\theta$进行迭代更新。经过多轮训练后，最终获得面向目标任务的最优参数集。

3.4　实验设置

以人民卫生出版社结核病学、儿科学教材，以及与教材配套的标准化习题集为数据来源，确保知识点覆盖一致性。此外，微调中还使用了CMed-MRC数据集，其数据来源于2021知识图谱评测[19]。实验中模型的超参数包括训练批次（epoch）、学习率（learning rate）、文本最大长度（max_len）、批量大小（batch_size）。由于显存限制，max_len设置为2 800，文本超出部分将截断，batch_size设置为2。参照BERT类模型微调的普遍实践设置模型微调的学习率，并在小范围（2e-5，3e-5，5e-5）内初步验证其收敛稳定性。采用网格搜索法，结果表明

learning rate 设定为 5e-5，epoch 设定为 25 时模型效果最优。

为验证 MC-BERT 模型效果，选择 3 类基线进行对比分析：基于潜在语义索引（latent semantic indexing，LSI）的传统参考信息推荐模型，面向通用中文优化的预训练模型 chinese-roberta-wwm-ext（CRWE），以及针对生物医学领域预训练的 PubMedBERT 模型。其中，考虑到 PubMedBERT 无法理解中文医学术语的内部结构，对该模型进行中文词表扩展。此外，为了提高习题解析搜索的准确性，在所有预训练模型中使用 Elasticsearch-7.17.3（ES）工具构建索引并对 QA 相关段落进行粗筛。Elasticsearch 是一个开源搜索引擎，通过简单连贯的 RESTful API 使全文搜索变得简单，隐藏了 Lucene 的复杂性，并提供分布式的实时文件存储。

以解析推荐精确率（precision，P）、召回率（recall，R）和 $F1$ 值为主要评估指标，判断模型推荐答案与专家标注标准答案的一致性。其中，TP 表示真正例，FP 表示假正例，FN 表示假负例。

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{7}$$

3.5 人工标注

人工标注环节分别针对习题解析推荐模型生成的参考答案与 QNLI 语料库进行系统标注。标注工作由两名医学专业研究生独立完成，并由一名医学信息学专业博士进行最终校验，以确保标注质量与一致性。在习题解析推荐模型生成答案标注中，主要判断所推荐的习题解析是否与给定的问答对具有相关性。在 QNLI 语料库的标注中，则依据标准 QNLI 分类体系，对参考解析与问答对之间的语义关系进行细粒度标注，具体分为 3 类。一是蕴含，解析内容与医学问答对相关，且能够逻辑推导出答案。二是矛盾，解析内容与问答对所陈述信息存在冲突。三是无关，解析内容与医学问答对之间不具备语义关联。所有经人工标注的数据将整合为独立语料库，作为后续模型优化与实验评估的基础资源。

4 实验结果与讨论

4.1 数据集构建

数据集包括 1 008 篇 CMedMRC 医学文本片段、约 2 万个问答对、41 本结核病学电子书和 1 本儿科学电子书，见表 1。考虑到儿科学习题涵盖的疾病种类分布更为分散，除了新生儿与新生儿疾病，还包含心血管系统疾病、呼吸系统疾病、营养和营养障碍性疾病等一系列常见疾病种类，使用前期标注的结核病学用于微调训练，随机抽取 150 道儿科学习题数据作用于测试。

表 1 数据集概况

数据项	规模
CMedMRC 医学文本片段	1 008 篇
CMedMRC 问答对	20 161 个
结核病学电子书	41 本
结核病学章节数	4 746 个
儿科学电子书	1 本
儿科学章节数	135 个
儿科学习题	10 236 道
测试集（Q-C 对）	2 250 个

4.2 对比分析

以基于 LSI 的参考信息推荐模型（LSI 模型）作为比较模型，验证 Book-MC-BERT 模型的性能。在不区分题型（案例类与事实类）的情况下，精确率为 48.5%。分析参考信息推荐错误的案例发现，问题的主题分类（尤其是第 3 级分类）与定位到的章节目录信息有很强的相关性。去除案例类实体，仅研究事实类问题，且加入主题分类相似度计算时，改进的 LSI 模型精确率为 64.5%。CRWE 模型和 Book-MC-BERT 模型精确率分别达到 71.3% 和 74.7%，说明了基于深度学习的问答模型的有效性。PubMedBERT 模型的性能仍显著低于 CRWE 模型。尽管进行了中文词表扩展，但其预训练目标（随机

掩码语言建模）无法适应医学概念完整性要求，且英文语料训练导致的语序差异进一步影响中文医学文本理解。增加分类信息后，Book-MC-BERT模型精确率升高至78.0%，比基础模型LSI模型提高了20.9%。通过对比结果，本研究方法取得了最优结果，见表2。

表2 模型性能与对比实验结果(%)

模型	P	R	*F*1
LSI（不区分题型）	48.5	85.2	61.8
LSI（区分题型，增加分类）	64.5	87.1	74.1
CRWE（Chinese-roberta-wwm-ext）	71.3	68.5	69.9
PubMedBERT	61.2	52.3	56.4
Book-MC-BERT（未增加分类）	74.7	84.5	79.3
Book-MC-BERT+分类（本研究方法）	78.0	83.2	80.5

4.3 人工标注及结果分析

使用最优模型，经过人工标注，最终形成可支持QNLI任务的语料库CMedBookQNLI，包含9 246条记录。标注范例，见表3。标签分为3类，1代表蕴含，基于候选解析可推理出答案；-1代表矛盾，基于候选解析可以排除答案；0代表不相关，候选解析与答案无关。每一个标注范例均可对应到原始问题标号及解析对应的具体章节，方便信息回溯。

表3 标注范例

标签	问题-答案对	候选解析
1	动脉导管完全闭合绝大多数发生于：3个月内	胎儿期动脉导管开放是血液循环的重要通道，出生后，大约15小时即发生功能性关闭，80%在出生后3个月解剖性关闭
0	营养不良患儿，当发生腹泻、呕吐时，均可导致以下的水、电解质紊乱，但除了：低血钾症	营养不良患儿患腹泻时易迁延不愈，持续腹泻又加重了营养不良，两者互为因果，形成恶性循环，最终导致多脏器功能异常
-1	下列哪项不是法洛四联症的主要临床表现：青紫	法洛四联症临床表现：青紫、蹲踞症状、杵状指（趾）、阵发性缺氧发作

基于预训练模型的医学参考答案解析推荐框架显著提高了标注效率。一方面其通过提供高质量的解析初稿，将人工任务从解析编写转变为校验与修正。依据既往相似规模全人工解析编写项目的工时记录（平均每人投入约20个工作日），本次标注任务总耗时约5个工作日，效率显著提升。另一方面，系统生成的解析中，有68%的内容置信度高（按标注标签为1且人工最终确认为1统计），可直接采纳。尽管未进行严格的随机对照耗时测量，但上述基于历史基准与详细过程数据的分析，表明该方法具备一定效率提升潜力。

训练MC-BERT模型时，由于MC-BERT本身的训练语料与电子书有较大区别，仅使用现有的结核病主题电子书和儿科学电子书，对模型改善有限，后期应纳入更多医学电子书资源，进一步改进Book-MC-BERT模型。未来应选取多个模型的参考信息推荐结果，根据相似度等指标进行重排序，以进一步提高解析推荐的准确率。此外，未来模型可通过结合自监督学习与主动学习，分析未标注的医学习题数据，自动挖掘潜在关联并生成高质量伪标签，进一步减少对人工标注的依赖。

5 结语

本研究针对医学解析在准确性、规范性和可解释性3方面要求高，以及既有研究仍存在错误累积、领域偏见，且未能充分利用多样、深层的文本信息等问题，提出基于预训练模型的医学参考答案解析推荐框架。通过MC-BERT模型融合领域知识，结合医学习题特有的分类信息，实现医学习题解析的半自动化生成，并通过对比实验验证改进策略的有效性与必要性。尽管未纳入BioBERT、Med-PaLM等全部相关模型，但对比实验体系已能有效阐明MC-BERT在解析抽取与QNLI任务标注中的独特价值。相较于通用领域预训练模型及传统方法，MC-BERT通过全实体遮蔽和知识图谱增强的专门化设计，显著提升了对医学概念完整性与逻辑严谨性的建模能力，这恰恰是医学习题解析生成任务的核心需求。本研究为半自动医学习题解析生成提供了有益补充，其性能仍受限于人工标注，未来将引入自监督学习与主动学习，以大幅减少人工标注需求，并基于生成的优质语料，进一步开展自然语言推理

技术的探索。

作者贡献：孙月萍负责研究设计、数据实验、论文撰写；王娟负责完善研究方案、数据实验；董良广负责数据集构建、实验结果分析；刘燕、杨丽负责数据标注与分析；李姣负责完善研究方案、论文修订；侯丽负责提出选题、完善研究方案、论文修订。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 JACOB D, CHANG M, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]. Minneapolis: The 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019.

2 VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. [2025-07-25]. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

3 游至宇, 阳倩, 傅姿晴, 等. 基于Transformer的预训练语言模型在生物医学领域的应用[J]. 厦门大学学报(自然科学版), 2024, 63(5): 883-893.

4 张仪方, 李琛, 程雨飞, 等. 生物医学大模型研究进展[J]. 生命科学, 2025, 37(12): 1481-1492.

5 LEE J, YOON W, KIM S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining[J]. Bioinformatics, 2020, 36(4): 1234-1240.

6 ALSENTZER E, MURPHY J, BOAG W, et al. Publicly available clinical BERT embeddings[C]. Minneapolis: The 2nd Clinical Natural Language Processing Workshop, 2019.

7 GU Y, TINN R, CHENG H, et al. Domain-specific language model pretraining for biomedical natural language processing[J]. ACM transactions on computing for healthcare, 2021, 3(1): 1-23.

8 LUO R, SUN L, XIA Y, et al. BioGPT: generative pre - trained transformer for biomedical text generation and mining [J]. Briefings in bioinformatics, 2022, 23 (6): 409.

9 SINGHAL K, AZIZI S, TU T, et al. Large language models encode clinical knowledge [J]. Nature, 2023, 620 (7): 172-180.

10 SINGHAL K, TU T, GOTTWEIS J, et al. Toward expert-level medical question answering with large language models [J]. Nature medicine, 2025, 31 (1): 943-950.

11 WU C, LIN W, ZHANG X, et al. PMC-LLaMA: toward building open-source language models for medicine [J]. Journal of the American medical informatics association, 2024, 31(9): 1833-1843.

12 ZHANG N, JIA Q, YIN K, et al. Conceptualized representation learning for Chinese biomedical text mining [EB/OL]. [2025-08-25]. https://arxiv.org/abs/2008.10813.

13 赵一鸣, 潘沛, 毛进. 基于任务知识融合与文本数据增强的医学信息查询意图强度识别研究[J]. 数据分析与知识发现, 2023, 7(2): 38-47.

14 VASANTHARAJAN C, TUN K Z, THI-NGA H, et al. MedBERT: a pre-trained language model for biomedical named entity recognition[C]. Chiang Mai: 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2022.

15 ZHANG T, CAI Z, WANG C, et al. SMedBERT: a knowledge-enhanced pre-trained language model with structured semantics for medical text mining [EB/OL]. [2025-08-20]. https://arxiv.org/abs/2108.08983.

16 周景, 崔灿灿, 王梦迪, 等. 基于RoBERTa和T5的两阶段医学术语标准化[J]. 计算机系统应用, 2024, 33(1): 280-288.

17 SMIT A, JAIN S, RAJPURKAR P, et al. CheXbert: combining automatic labelers and expert annotations for accurate radiology report labeling using BERT[C]. Online: The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.

18 WANG B, XIE Q, PEI J, et al. Pre-trained language models in biomedical domain: a systematic survey [J]. ACM computing surveys, 2023, 56(3): 1-52.

19 ZHANG T, WANG C, QIU M, et al. Knowledge-empowered representation learning for Chinese medical reading comprehension: task, model and resources [C]. Virtual: Findings of the Association for Computational Linguistics: ACL-IJCNLP, 2021.