·专题报道——公共卫生信息化建设与人工智能·

呼吸道传染病疫情流行病学调查数据智能采集策略研究

孟玲[1]，牛艳[1]，向妮娟[1]，任小凡[2]，马永胜[3]，李峰[4]，黄鑫[5]，余涛[5]，鲁飞洋[5]，陈雪瑞[6]，李群[7]

1. 中国疾病预防控制中心卫生应急中心 传染病溯源预警与智能决策全国重点实验室，北京 102206；
2. 虹口区疾病预防控制中心，上海 200082；
3. 济南市疾病预防控制中心，济南 250021；
4. 深圳市腾讯科技（深圳）有限公司，深圳 518057；
5. 讯飞医疗科技股份有限公司，合肥 230088；
6. 北京辰安科技股份有限公司，北京 100094；
7. 中国疾病预防控制中心 传染病溯源预警与智能决策全国重点实验室，北京 102206
通信作者：李群，Email：liqun@chinacdc.cn

【摘　要】目的 分析呼吸道传染病疫情流行病学调查场景模式，设计针对不同流行病学调查数据采集目的的呼吸道传染病疫情流行病学调查数据智能采集策略。**方法** 系统回顾和梳理我国典型呼吸道传染病疫情的发生场景，采用文献研究、焦点小组访谈、专家咨询等方法，归纳呼吸道传染病疫情流行病学调查场景模式结构与模式特征。结合当前主要应用的数据智能采集技术和工具，设计呼吸道传染病疫情流行病学调查数据智能采集策略，设置流行病学调查数据智能采集模块，并分析整理各模块的具体采集内容和采集模式。**结果** 呼吸道传染病疫情流行病学调查场景模式包括病例个案流行病学调查和涉疫场所流行病学调查。针对流行病学调查数据采集目的确定智能采集策略，包括对少量特异风险人群的详细精准病例个案流行病学调查数据采集、对大量未知风险人群的简单快速病例个案流行病学调查数据采集以及对涉疫场所的流行病学调查数据采集。设置 7 个流行病学调查数据智能采集模块，整理流行病学调查采集变量共 1 012 个。**结论** 本研究基于呼吸道传染病疫情的传播场景以及不同传播场景下的流行病学调查数据采集目的，提出针对性的呼吸道传染病疫情流行病学调查数据智能采集策略，设置智能采集模块，确定采集内容和采集模式，为智能流行病学调查系统的设计和开发奠定了基础。

【关键词】 呼吸道传染病；现场流行病学调查；信息采集；人工智能；人机交互

Intelligent data acquisition strategies for epidemiological investigations of respiratory infectious disease outbreaks

MENG Ling[1], NIU Yan[1], XIANG Nijuan[1], REN Xiaofan[2], MA Yongsheng[3], LI Feng[4], HUANG Xin[5], YU Tao[5], LU Feiyang[5], CHEN Xuerui[6], LI Qun[7] (*1. National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases*, *Public Health Emergency Center*, *Chinese Center for Disease Control and Prevention*, *Beijing* 102206, *China*; *2. Hongkou District Center for Disease Control and Prevention*, *Shanghai* 200082, *China*; *3. Jinan Center for Disease Control and Prevention*, *Jinan* 250021, *China*; *4. Tencent Technology (Shenzhen) Company Limited*, *Shenzhen* 518057, *China*; *5. Xunfei Healthcare Technology Co.*, *Ltd*, *Hefei* 230088, *China*; *6. Beijing GS Technology Co.*, *Ltd*, *Beijing* 100094, *China*; *7. National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases*, *Chinese Center for Disease Control and Prevention*, *Beijing* 102206, *China*)
Corresponding author: LI Qun, Email: liqun@chinacdc.cn

【Abstract】 Objective To analyze the epidemiological investigation scenario patterns of respiratory infectious disease outbreaks and design intelligent data collection strategies for different epidemiological investigation data collection objectives. **Methods** We systematically reviewed and sorted the occurrence scenarios of typical respiratory infectious disease outbreaks in China. Using literature research, focus group interviews, and expert consultations, we summarized the structural and characteristic patterns of epidemiological investigation scenarios for respiratory infectious diseases. Combined with currently used intelligent data collection technologies and tools, we designed intelligent data collection strategies for epidemiological investigations of respiratory infectious disease outbreaks, established intelligent data collection modules for epidemiological investigations, and analyzed and organized the specific collection content and modes of each module. **Results** Epidemiological investigation scenario patterns of respiratory infectious disease outbreaks include case-based epidemiological investigations and epidemiological investigations of affected sites. Intelligent collection strategies were determined based on the data collection objectives of the epidemiological

开放获取：CC BY-NC-ND 4.0　DOI：10.11847/zgggws1146411
基金项目：新一代人工智能国家科技重大专项（2021ZD113901）；公共卫生人才培养支持项目
第一作者：孟玲（1978—），副研究员，硕士，研究方向：传染病预防与控制。
收稿日期：2024-11-12　修回日期：2024-11-22　录用日期：2024-11-28
利益冲突：不存在　伦理审查：不需要　出版授权：全体作者已与编辑部签署作者贡献声明及版权转让协议

investigations, including detailed and precise case-based epidemiological investigation data collection for a small number of specific high-risk populations, simple and rapid case-based epidemiological investigation data collection for a large number of unknown risk populations, and epidemiological investigation data collection for affected sites. Seven intelligent data collection modules for epidemiological investigations were established, and a total of 1 012 epidemiological investigation collection variables were organized. **Conclusions** Based on the transmission scenarios of respiratory infectious disease outbreaks and the data collection objectives of epidemiological investigations under different transmission scenarios, this study proposes targeted intelligent data collection strategies for epidemiological investigations of respiratory infectious disease outbreaks, establishes intelligent collection modules, and determines the collection content and modes, laying the foundation for the design and development of intelligent epidemiological investigation systems.

【**Keywords**】 respiratory infectious diseases; field epidemiological investigation; information acquisition; artificial intelligence; human-computer interaction

急性呼吸道传染病发病急、传播快，尤其是新发突发急性呼吸道传染病，一旦发生大流行，可对健康、政治、经济和社会产生巨大危害[1–2]。通过流行病学调查开展疫情溯源、传播链分析、密接人员的判定和追踪是我国开展呼吸道传染病疫情主动防控策略的关键环节，在我国成功实施并取得了显著成效[3–5]。流行病学调查数据的采集是后续开展数据分析、形势研判、风险管理等防控措施的基础和前提。然而，传统流行病学调查模式的数据采集环节存在采集效率低、工作负荷大、防护要求高、快速采集难等问题[6–9]。数据采集相关的人工智能技术一直在发展演进，通过先进的图像识别、语音识别、人机对话、自然语言处理和智能推理等技术，采用流行病学调查终端辅助流调人员记录、或机器人电话、APP、网页直接与流调对象进行交互等形式，开展人员、地点、环境、活动数据的采集和智能分析，为实现智能化流行病学调查提供了可能，在呼吸道传染病疫情应对中得到了初步应用[10–20]。利用高效、稳定、无接触式的多模态流行病学调查数据智能采集技术，提升流行病学调查工作效率，是流行病学调查智能化发展的主要方向。为分析呼吸道传染病疫情流行病学调查场景模式，设计针对不同流行病学调查数据采集目的的传染病疫情流行病学调查数据智能采集策略，特进行本研究。

1 资料与方法

1.1 典型呼吸道传染病疫情发生场景分析 采用文献研究法，系统回顾和梳理 2019—2021 年我国 40 起典型呼吸道传染病疫情的发生场景，分析疫情发生规模和主要发生场所。

1.2 流调场景模式结构与模式特征归纳 采用文献研究、焦点小组访谈、专家咨询等方法，基于对一线流调人员呼吸道传染病疫情流行病学调查场景以及流行病学调查数据采集需求的系统梳理，归纳呼吸道传染病疫情流行病学调查场景模式结构与模式特征。

1.3 流调数据智能采集策略设计 结合 2021—2024 年主要应用的 OCR 识别、语音识别、人机对话、自然语言处理、关键要素抽取、关联推荐、地图交互选址等数据智能采集技术，以及电脑、平板、智能外呼电话、APP、网页等数据智能采集工具，设计针对不同流行病学调查数据采集目的的呼吸道传染病疫情流行病学调查数据智能采集策略。

1.4 流调数据智能采集需求定义 设置流行病学调查数据智能采集模块，分析整理各流行病学调查数据智能采集模块的具体采集内容和采集模式。

1.5 研究技术路线（图 1）

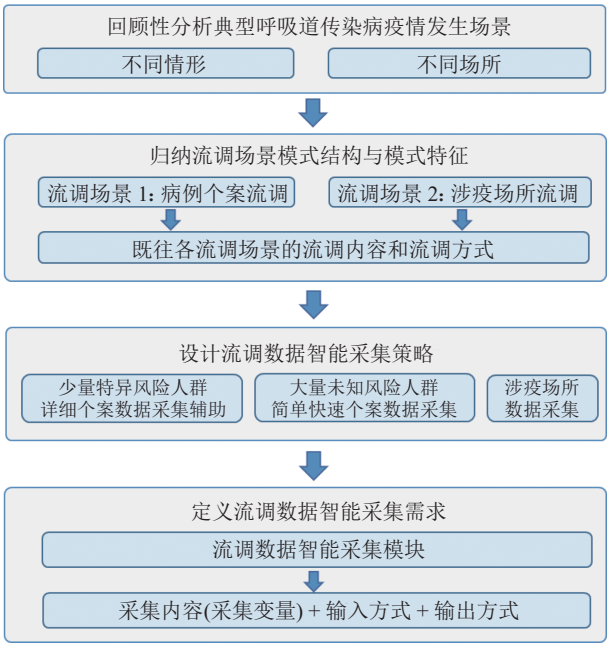


图 1 呼吸道传染病疫情流调数据智能采集策略研究路线

Fig. 1 Research roadmap for intelligent collection strategies of respiratory infectious disease epidemiological investigation data

2 结 果

2.1 传染病疫情流行病学调查场景模式

2.1.1 传染病疫情流行病学调查场景 流行病学

调查场景主要包括病例个案流行病学调查、涉疫场所流行病学调查。当发现呼吸道传染病病例时，流调人员通过电话调查、面对面调查、“三公（工）”协同多部门技术手段和大数据信息支撑等方式对病例个案进行流行病学调查，通过电话调查、现场勘查、“三公（工）”协同多部门技术手段和大数据信息支撑等方式对涉疫场所进行流行病学调查。

2.1.2 个案流行病学调查模式　病例个案流行病学调查通常依照个案调查表进行。病例个案调查表的主要内容包括病例基本情况、病例发现与就诊、疫苗接种、危险因素、流行病学史（包括活动轨迹、接触史、暴露史、旅居史）、密切接触者、标本采集及检测等信息。

2.1.3 涉疫场所流行病学调查模式（表 1）　本研究对国内典型呼吸道传染病疫情场景实例的分析显示，涉疫重点场所类型主要包括：家庭，工作单位，医疗机构，餐饮场所，商场超市便利店，交通工具和场站，农贸、集贸市场，工厂、加工厂，娱乐休闲健身场所，养老机构，监狱、看守所、戒毒所等管控场所，学校、幼托、培训机构，工地，隔离点，采样点，办事点或营业网点，酒店等。上述涉疫重点场所的流行病学调查主要内容通常均包括场所基本情况（场所概况、日常防控、人员管控等）、场所中病例个案轨迹、场所中病例个案密切接触者情况、场所标本采集及检测等，但具体流行病学调查内容根据场所类型和特征不同而有所差异。

表 1 涉疫重点场所流行病学调查主要内容
Table 1 Main contents of epidemiological investigations in key places involved in the epidemic

场所类型	流行病学调查主要内容
病例家庭	病家所在小区/村居基本情况、病家基本情况（含病例在家主要活动情况）等； 病例家庭密切接触者情况； 病例家庭标本采集及检测情况。
工作单位	工作单位基本情况、工作岗位情况（含病例在工作单位主要活动情况）、工作单位日常防控情况、工作单位人员管控情况； 工作单位中密切接触者情况； 工作单位标本采集及检测情况。
医疗机构，餐饮场所，商场超市便利店，交通工具和场站，农贸、集贸市场，工厂、加工厂，娱乐休闲健身场所，养老机构，监狱、看守所、戒毒所等管控场所，学校、幼托、培训机构，工地，隔离点，采样点，办事点或营业网点，酒店等	场所基本情况、日常防控情况、场所人员管控情况等； 场所中病例个案轨迹； 场所中密切接触者情况； 场所标本采集及检测情况。

2.2 基于流行病学调查数据采集目的的传染病疫情流行病学调查数据智能采集策略

2.2.1 对少量特异风险人群的详细精准病例个案流行病学调查数据采集

2.2.1.1 低水平流行期或重点时期、重点病例　对发现早、病例数少、未发生持续社区传播的疫情，或对疫情早期病例、收尾阶段的新发病例、重点病例（如物资保障人员、快递人员、志愿者、社区服务人员等封管控区域之外人员中出现的病例），需对病例个案开展详细精准流行病学调查，收集病例基本情况、病例发病与就诊、疫苗接种、危险因素、流行病学史（包括活动轨迹、接触史、暴露史、旅居史）、密切接触者、标本采集及检测等信息。该情形下，采用人工流调员辅以智慧流行病学调查系统的方式进行采集，采集模式为流调人员全程直接进行流行病学调查并持电脑/平板终端记录。

2.2.1.2 中水平流行期　在病例个案数较多且仍需要进行详细精准个案流行病学调查数据采集的情形下，也需要收集病例基本情况、病例发病与就诊、疫苗接种、危险因素、流行病学史（包括活动轨迹、接触史、暴露史、旅居史）、密切接触者、标本采集及检测等信息。该情形下，可先通过个案–APP 交互、个案–网页交互、个案–机器人电话交互的模式对流行病学调查核心信息所包括的关键要素进行智能采集，再由流调人员核查、补充调查、并持电脑/平板终端记录人机交互未完成的流行病学调查数据采集内容。

2.2.2 对大量未知风险人群的简单快速病例个案流行病学调查数据采集　在高水平流行期，当疫情出现社区持续传播，病例数明显增多时，对病例开展简单快速流行病学调查，重点收集病例的基本信息、发病时间、检测阳性时间、危险因素等信息，用于疫情发展态势分析。简单快速个案

流行病学调查数据采集时，主要通过个案–APP交互、个案–网页交互、个案–机器人电话交互的模式进行数据智能采集。

2.2.3 对涉疫场所的流行病学调查数据采集 根据疫情防控需要，对涉疫重点场所进行流行病学调查，重点收集场所基本情况（场所概况、日常防控、人员管控等）、场所中病例个案轨迹、场所中病例个案密切接触者情况、场所标本采集及检测等信息。对涉疫场所的流行病学调查数据采集，主要由流调人员全程直接进行流行病学调查并持电脑/平板终端记录。

2.3 流行病学调查数据智能采集模块设置（表 2） 根据本研究对流行病学调查场景模式结构和模式特征的分析结果以及对流行病学调查数据智能采集策略的设计，将流行病学调查数据智能采集内容设置为 7 个流行病学调查数据智能采集模块，分别为模块 1 病例个案流调基础模块、模块 2 病例个案轨迹基础模块、模块 3 病例个案轨迹重点场所模块，模块 4 密切接触者情况模块、模块 5 病例个案标本采集及检测模块、模块 6 重点场所标本采集及检测模块、模块 7 重点场所流调基础模块。模块 1 至 4 的流调对象主要为病例个案，其中模块 3 和 4 可根据涉疫重点场所流调进行补充；模块 5 和模块 6 的流调对象为检测机构，模块 7 的流调对象为涉疫重点场所的联系人或负责人。各流行病学调查数据智能采集模块的具体采集内容以及各采集模式下的采集变量数见表 2。全部流行病学调查采集变量共 1 012 个。在全部采集变量中，选取流调关键核心信息相关变量，同时考量不同人机交互方式的便利性和可接受度、数据结构化的便利性，确定不同人机交互模式下的采集变量，具体为：从病例个案流行病学调查的模块 1 至 4 中选取 86 个核心变量，可通过个案–APP 交互、个案–网页交互的模式采集；从病例个案流行病学调查的模块 1 中选取 24 个变量，可通过个案–机器人电话交互的模式采集。

表 2 呼吸道传染病疫情流行病学调查数据智能采集模块

Table 2 Intelligent collection module of epidemiological investigation data of respiratory infectious diseases

模块编号	模块名称	采集内容	全部采集变量数/个	APP或网页采集变量数/个	机器人电话采集变量数/个
1	病例个案流调基础模块	病例基本情况、病例发现与就诊、病例疫苗接种情况、流行病学史、标本采集及检测情况	115	31	24
2	病例个案轨迹基础模块	轨迹日期类型、轨迹日期、时间段、场所	4	4	—
3	病例个案轨迹重点场所模块	病例个案轨迹–病例家庭	13	6	—
		病例个案轨迹–病例工作单位	17	9	—
		病例个案轨迹–医疗机构	21	8	—
		病例个案轨迹–交通工具	18	10	—
		病例个案轨迹–农贸、集贸市场	17	—	—
		病例个案轨迹–工厂、加工厂	35	—	—
		病例个案轨迹–娱乐休闲健身场所	17	—	—
		病例个案轨迹–隔离点	17	—	—
		病例个案轨迹–采样点	11	—	—
		病例个案轨迹–办事点或营业网点	20	—	—
		病例个案轨迹–其他场所通用	17	7	—
4	密切接触者情况模块	病例接触者/暴露者的个人基本信息、发现途径、接触情况	21	11	—
5	病例个案标本采集及检测模块	病例标本采集情况、检测项目、检测方法、检测结果	28	—	—
6	重点场所标本采集及检测模块	重点场所标本采集情况、检测项目、检测方法、检测结果	18	—	—

续表 2
Table 2 Continued

模块编号	模块名称	采集内容	全部采集变量数/个	APP或网页采集变量数/个	机器人电话采集变量数/个
7	重点场所流调基础模块	重点场所情况–病例家庭	30	—	—
		重点场所情况–病例工作单位	36	—	—
		重点场所情况–医疗机构	25	—	—
		重点场所情况–餐饮场所	30	—	—
		重点场所情况–商场超市便利店	30	—	—
		重点场所情况–交通工具和场站	48	—	—
		重点场所情况–农贸、集贸市场	66	—	—
		重点场所情况–工厂、加工厂	44	—	—
		重点场所情况–娱乐休闲健身场所	32	—	—
		重点场所情况–养老机构	45	—	—
		重点场所情况–监狱、看守所、戒毒所等管控场所	32	—	—
		重点场所情况–学校、幼托、培训机构	31	—	—
		重点场所情况–工地	32	—	—
		重点场所情况–隔离点	32	—	—
		重点场所情况–采样点	27	—	—
		重点场所情况–办事点或营业网点	28	—	—
		重点场所情况–酒店	28	—	—
		重点场所情况–其他场所通用	27	—	—
合计			1 012	86	24

3 讨　论

本研究以呼吸道传染病疫情流行病学调查为例，基于呼吸道传染病疫情的传播场景以及不同传播场景下的流行病学调查数据采集目的，提出针对性的传染病疫情流行病学调查数据智能采集策略，设置流行病学调查数据智能采集模块，确定流行病学调查数据智能采集内容和采集模式，为智能流行病学调查系统的设计和开发奠定了基础。

在对重点场所中病例个案轨迹、重点场所情况进行流行病学调查时，具体调查内容根据场所类型和特征不同而有所差异。因此，为了增强调查内容的适配性，本研究在模块 3 病例个案轨迹重点场所模块、模块 7 重点场所流调基础模块，针对各个重点场所的特异性，分别梳理了各个场所内需要采集的变量。

本研究所设计的流行病学调查数据智能采集策略中，包含在中高水平流行期病例个案数较多的情形下，通过任务型人机对话，以个案–APP 交互、个案–网页交互、个案–机器人电话交互的模式，完成部分流行病学调查数据的采集。目前人机交互的核心技术——人机对话技术还处于发展阶段，在自然性、逻辑性、流畅性等方面与人类有一定差距，在低资源驱动、多领域切换、知识推理和复杂任务等场景下，对话效果和用户体验不具有稳定性，任务型人机对话系统受限于场景的难易程度，还有很大进步空间[21–22]。在设计任务型人机对话系统时，需要平衡问题数与用户体验之间的关系，过多的问题可能会增加用户的负担和不满，从而导致调查中断或调查质量下降。流行病

学调查信息密度大，问答情形复杂，目前的人机对话智能技术尚无法替代有经验的流调人员完成全部变量的采集。为了确保人机对话数据采集的效果，鉴于 APP 和网页可以采用文本、语音、OCR、拍照、地图交互选址、GPS 定位等多种人机交互方式进行信息采集，采集效率相对较高，本研究将个案–APP 交互、个案–网页交互的采集变量数限定为 86 个核心变量；而机器人电话仅采用语音方式进行信息采集，采集模式相对单一，不宜纳入过多变量，故本研究将个案–机器人电话的采集变量数限定为 24 个变量。

致谢 感谢清华大学工程物理系（公共安全研究院）陈涛老师在本文撰写中提供宝贵意见

参考文献

[1] 马旺, 霍翔, 周明浩. 急性呼吸道传染病疾病负担及评价方法研究进展 [J]. 中国公共卫生, 2017, 33(2): 266-270.

[2] 杨津, 冯录召, 赖圣杰, 等. 急性呼吸道传染病症状监测及预警技术的现状与展望 [J]. 中华流行病学杂志, 2023, 44(1): 60-66.

[3] 中华人民共和国国家卫生健康委员会. 2020 年 2 月 23 日新闻发布会文字实录 [EB/OL]. [2023-12-21]. http://www.nhc.gov.cn/xcs/yqfkdt/202002/23d396a092224c51a8599b492a44a81c.shtml.

[4] 中华人民共和国国家卫生健康委员会. 2020 年 5 月 6 日新闻发布会文字实录 [EB/OL]. [2023-12-21]. http://www.nhc.gov.cn/xcs/fkdt/202005/e960f8dda94b46358872f18f7e09e0b3.shtml.

[5] 中华人民共和国国家卫生健康委员会. 国务院联防联控机制 2021 年 9 月 29 日新闻发布会文字实录 [EB/OL]. [2023-12-21]. http://www.nhc.gov.cn/xcs/yqfkdt/202109/14271f3eb33145058725fa9bd50af537.shtml.

[6] 万源, 罗俊, 刘普林, 等. 运用信息技术辅助流行病学调查工作的思考 [J]. 中国现代医生, 2022, 60(20): 85-88.

[7] 朱文杰, 金武, 黄友, 等. 新冠肺炎流行病学调查工作的实践与体会 [J]. 人民军医, 2021, 64(12): 1313-1316.

[8] 刘榆, 向于, 田应桥, 等. 基层疾控机构新冠肺炎疫情期间流行病学调查实践 [J]. 江苏预防医学, 2021, 32(1): 110-111, 117.

[9] 李亚品, 宋文静, 高东旗, 等. 新型冠状病毒肺炎疫情现场流行病学调查技术及调查实案探讨 [J]. 解放军医学院学报, 2020, 41(2): 103-107.

[10] 王维笑, 贾露荣, 陶焜, 等. 基于人工智能的新型冠状病毒肺炎流行病学信息采集系统的开发 [J]. 中国医学装备, 2020, 17(11): 143-146.

[11] 栗圆, 高燕琳, 李刚. 新冠肺炎疫情现场流行病学调查系统的建设与应用 [J]. 中国卫生信息管理杂志, 2020, 17(5): 627-631.

[12] 科大讯飞. 科大讯飞: 智医助理电话机器人 [EB/OL]. [2023-12-21]. https://chima.org.cn/Html/News/Articles/6174.html.

[13] 周雨萌. 深圳现场流调处置系统投入使用 [EB/OL]. [2023-12-21]. http://sztqb.sznews.com/PC/content/202204/06/content_1182288.html.

[14] 唐斌, 凌思凯, 姜胜耀. 智能语音外呼系统在新冠肺炎疫情防控中的应用实践 [J]. 中国数字医学, 2020, 15(6): 63-65.

[15] 陈曦. 天津用科技实力“硬刚”奥密克戎 [EB/OL]. [2023-12-21]. http://www.stdaily.com/index/kejixinwen/2022-01/13/content_1245961.shtml.

[16] 李玉凤. 严防死守 以快制快 坚决打赢疫情阻击战 AI 智能“流调员” 跑出流调加速度 [EB/OL]. [2023-12-21]. https://kfqgw.beijing.gov.cn/zwgkkfq/ztzl/fkyqbjjkqzxdkfq/rdxwkfq/202207/t20220706_2765379.html.

[17] Amiri P, Karahanna E. Chatbot use cases in the Covid-19 public health response[J]. Journal of the American Medical Informatics Association, 2022, 29(5): 1000-1010,

[18] 周慎, 朱旭峰, 薛澜. 人工智能在突发公共卫生事件管理中的赋能效用研究——以全球新冠肺炎疫情防控为例 [J]. 中国行政管理, 2020(10): 35-43.

[19] 王石萌. 人工智能在抗疫中的应用 [J]. 河南科技, 2021, 40(10): 6-8.

[20] 赵杨, 曹文航. 人工智能技术在新冠病毒疫情防控中的应用与思考 [J]. 信息资源管理学报, 2020, 10(6): 20-27.

[21] 赵阳洋, 王振宇, 王佩, 等. 任务型对话系统研究综述 [J]. 计算机学报, 2020, 43(10): 1862-1896.

[22] 曹亚如, 张丽萍, 赵乐乐. 多轮任务型对话系统研究进展 [J]. 计算机应用研究, 2022, 39(2): 331-341.