

团 体 标 准

T/CCSA001—2024

医疗领域大模型应用数据安全规范

Specification for Data Security of Large Model Applications in Healthcare

2024 年 6 月 26 日发布

2024 年 7 月 1 日实施

成都市生物医学信息学会 发布

目 录

前 言 II

引 言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 3

5 概述 3

 5.1 医疗大模型概述 3

 5.2 安全要求 4

 5.3 模型框架 4

6 基本原则 5

 6.1 法律和伦理规定 5

 6.2 数据安全和隐私保护 5

 6.3 数据可用性 6

7 数据收集和预处理 6

 7.1 数据来源和采集方式 6

 7.2 数据隐私和脱敏处理 7

 7.3 数据标准化和编码 7

 7.4 数据处理和分析 7

 7.5 数据开放和共享 8

8 医疗大模型开发和验证安全 8

 8.1 训练数据的安全 8

 8.2 大模型架构选择 8

 8.3 大模型训练和优化 9

9 医疗大模型推理安全 9

 9.1 推理数据的隐私安全 9

 9.2 医疗大模型参数的隐私安全 10

10 医疗大模型安全评估 10

11 医疗大模型应用安全管理 10

附录 A（资料性）大模型在医疗领域的应用 12

附录 B（资料性）医疗大模型的分类 13

附录 C（资料性）医疗数据的分类 14

前 言

本规范按照 GB/T 1.1-2020《标准化工作导则第 1 部分：标准化文件的结构和起草规则》的规定起草。

请注意本规范的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本规范由成都市生物医学信息学会提出并归口。

本规范起草单位：中国信息通信研究院、北京数牍科技有限公司、北京壹永科技有限公司。

本规范主要起草人：郑威、凌霞、张学阳、蔡超超、单进勇、金银玉、刘晓华、张程割、陈小梅。

本规范为首次发布。

引 言

医疗数据是一类极其重要的数据，包含患者病历、影像数据、生理参数、基因数据等敏感信息，不仅事关个人隐私，还涉及社会公共利益和国家安全。近期大模型研究取得一系列突破性进展，参数规模达到千亿级别，在文本生成和理解方面具有超强的能力。大模型广泛应用的同时也衍生出一系列隐私保护和数据安全问题。在医疗领域，大模型在改善医疗诊断、药物研发、临床决策等方面具有巨大应用前景。由于大模型的训练和应用均涉及海量医疗数据，其不当使用可能带来的个人隐私泄露、伦理道德、歧视偏见等风险需要得到广泛关注。本规范适用于具备医疗大模型开发能力的供应商或制造企业等，可指导医疗机构部署医疗大模型产品/解决方案，也可作为医疗机构选择或评价大模型安全应用的依据，旨在规范医疗领域大模型数据应用的行为，明确相应的数据隐私保护措施和模型安全评估要求，确保医疗数据的隐私和安全得到妥善保护，推动医疗领域大模型技术创新与发展。

医疗领域大模型应用数据安全规范

1 范围

本文件规定了医疗领域大模型应用的数据安全管理要求，涵盖大模型数据的收集、存储、处理、共享和应用等各个阶段。

本文件适用于开发、部署和使用医疗大模型的医疗机构、科技公司、研究机构、服务提供商及监管部门等开展数据安全管理或评估评价。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 35273—2020 信息安全技术 个人信息安全规范

GB/T 41867—2022 信息技术 人工智能 术语

3 术语和定义

GB/T 35273—2020、GB/T 41867—2022界定的以及下列术语和定义适用于本文件。

3.1

大模型 large model

具有大规模参数和复杂计算结构的机器学习模型。这些模型通常由深度神经网络构建而成，拥有数十亿甚至数千亿个参数，通过训练海量数据来获得。

3.2

大模型应用 large model application

在实际场景中利用大模型来解决各种问题或提供服务的过程，包括但不限于自然语言处理、计算机视觉、语音识别和推荐系统等。

3.3

人工智能 artificial intelligence

人工智能系统相关机制和应用的研究和开发，其中人工智能系统是针对人类定义的给定目标，产生诸如内容、预测、推荐或决策等输出的一类工程系统。

[来源：GB/T 41867—2022，3.1]

3.4

生成式人工智能 artificial intelligence generated content

基于人工智能通过已有数据寻找规律，并自动生成内容的生产方式，如文本生成、图像生成、视频生成、音频生成、游戏生成、代码生成等。

3.5

个人信息 personal information

以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人身份或者反映特定自然人活动情况的各种信息。

注1：个人信息包括姓名、出生日期、身份证号码、个人生物识别信息、住址、通信通讯联系方式、通信记录和内容、账号密码、财产信息、征信信息、行踪轨迹、住宿信息、健康生理信息、交易信息等。

注2：个人信息控制者通过个人信息或其他信息加工处理后形成的信息，例如，用户画像或特征标签，能够单独或者与其他信息结合识别特定自然人身份或者反映特定自然人活动情况的，属于个人信息。

[来源：GB/T 35273—2020，3.1]

3.6

个人敏感信息 personal sensitive information

一旦泄露、非法提供或滥用可能危害人身和财产安全，极易导致个人名誉、身心健康受到损害或歧视性待遇的个人信息。

注1：个人敏感信息包括身份证件号码、个人生物识别信息、银行账户、通信记录和内容、财产信息、征信信息、行踪轨迹、住宿信息、健康生理信息、交易信息、14岁以下（含）儿童的个人信息等。

注2：个人信息控制者通过个人信息或其他信息加工处理后形成的信息，如一旦泄露、非法提供或滥用可能危害人身和财产安全，极易导致个人名誉、身心健康受到损害或歧视性待遇等的，属于个人敏感信息。

[来源：GB/T 35273—2020，3.2]

3.7

授权同意 consent

个人信息主体对其个人信息进行特定处理作出明确授权的行为。

注：包括通过积极的行为做出授权（即明示同意），或者通过消极的不作为而作出授权（如信息采集区域内的个人信息主体在被告知信息收集行为后没有离开该区域）。

[来源：GB/T 35273—2020，3.7]

3.8

隐私计算 privacy-preserving computation

在保证数据提供方不泄露原始数据的前提下，对数据进行分析计算的一系列信息技术，保障数据在流通与融合过程中的“可用不可见”。隐私计算技术主要包括多方安全计算、联邦学习、可信执行环境等主流技术。

3.9

差分隐私 differential privacy

一种隐私保护的数据共享手段，通过添加噪声的方式（如Laplace噪声、指数噪声等），实现仅分享可以描述数据库的一些统计特征、而减少识别其单个个体数据记录的机会。

3.10

联邦学习 federated learning

由两个或以上参与方，在保证各自原始私有数据不出其定义的私有边界的前提下，协作构建并使用机器学习模型的技术架构。

[来源：IEEE P3652.1- 2020，有修改]

3.11

匿名化 anonymization

通过对个人信息的技术处理，使得个人信息主体无法被识别或者关联，且处理后的信息不能被复原的过程。

注：个人信息经匿名化处理后所得的信息不属于个人信息。

[来源：GB/T 35273—2020，3.14]

3.12

去标识化 de-identification

通过对个人信息的技术处理，使其在不借助额外信息的情况下，无法识别或者关联个人信息主体的过程。

注：去标识化建立在个体基础之上，保留了个体颗粒度，采用假名、加密、哈希函数等技术手段替代对个人信息的标识。

[来源：GB/T 35273—2020，3.15]

3.13

脱敏 data masking

一种数据保护技术，用于在数据共享、分析等场景下，对敏感信息进行变换或隐藏，以降低数据的隐私风险。脱敏处理旨在保护数据的隐私，使敏感信息无法被直接识别，如个人身份信息、金融数据、健康数据等，同时保留数据的结构和可用性。脱敏处理通常采用替换、泛化、屏蔽、加噪声等技术手段。

4 缩略语

下列缩略语适用于本文件。

AIGC	生成式人工智能	Artificial Intelligence Generated Content
API	应用程序接口	Application Programming Interface
CNN	卷积神经网络	Convolutional Neural Networks
GPT	生成式预训练Transformer模型	Generative Pre-trained Transformer
LLM	大规模语言模型	Large Language Models
RLHF	人类反馈的强化学习	Reinforcement Learning from Human Feedback
RNN	递归神经网络	Recurrent Neural Network
SFT	有监督微调	Supervised Fine-Tuning

5 概述

5.1 医疗大模型概述

医疗大模型是一种应用于医疗健康领域的大型机器学习模型，它结合了深度学习、自然语言处理、计算机视觉等先进技术，通过海量的医疗数据进行训练，旨在解决医学诊断、治疗决策支持、疾病预测、药物研发、患者管理和公共卫生分析等方面的问题，参见附录A。这些模型通常具有高度复杂的结构和

大量的参数，可以对医疗文本、图像、基因组学等多种类型的数据进行综合分析，以提供更为精准、全面的医疗服务。另外医疗大模型按照技术类型、训练目标、应用场景等不同的维度进行分类，参见附录B。

相较于其他领域的大模型，医疗大模型有以下显著区别：

——数据质量与安全要求极高。医疗行业的特殊性决定了其数据具有高度敏感性和法律合规要求，因此在数据收集、存储和使用过程中必须遵循严格的医疗数据隐私保护法规。同时，由于医疗决策直接关系到生命健康，所以医疗大模型所用的数据必须具备极高的准确性和完整性，容错率较低。

——领域专业知识依赖性强。医疗大模型不仅需要强大的算法和计算能力，更需要深厚的医学知识背景支撑。模型不仅要理解并应用复杂的医学概念和术语，还要能够根据最新的临床指南和研究进展不断更新迭代。

——应用场景专业且细分：医疗大模型的应用更加垂直和专业化，如针对特定疾病的诊断模型、药物发现模型以及个性化治疗方案推荐系统等，每个细分领域都可能需要定制化的模型设计和训练。

——安全风险及伦理考量：除了技术层面的安全性，医疗大模型还涉及伦理和社会责任问题，比如公平性和透明度，确保模型不引入偏见，解释结果可追溯，并尊重患者的知情权和选择权。

5.2 安全要求

医疗大模型涉及数据提供者、服务提供者、技术支持者、服务使用者等几个角色，可以由不同的实体承担不同的角色，不同实体之间应保护数据和模型的隐私安全；涉及数据收集与处理、大模型训练、模型安全评估和大模型推理等主要步骤的安全要求，见图1。

- 数据提供者：为医疗大模型训练提供数据或标注的组织、个人。
- 服务提供者：提供医疗大模型服务的组织、个人。
- 技术支持者：为医疗大模型提供技术支持的组织、个人。
- 服务使用者：使用医疗大模型服务的组织、个人。

注：相关角色参考《互联网信息服务深度合成管理规定》。

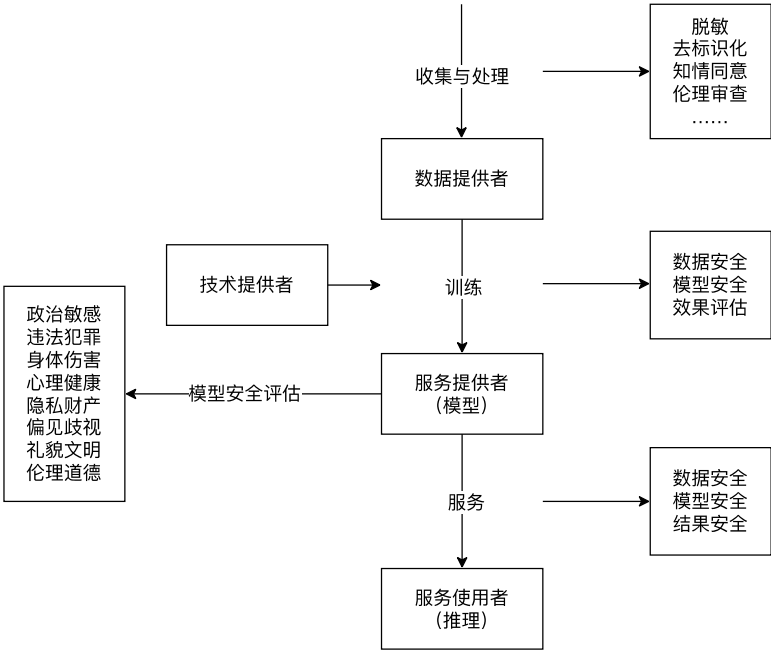


图 1 医疗大模型安全要求

5.3 模型框架

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/235114024204011233>