“人工智能医学信息系统软件审评指导体系构建”课题

——人工智能医学软件网络安全技术审评指导原则

(征求意见稿)

本指导原则是人工智能医学信息系统软件审评指导体系构建的组成部分，基于人工智能医疗器械审评指导原则和医疗器械网络安全注册审查指导原则的通用要求，细化了人工智能医学软件网络安全的一般要求。

本指导原则旨在指导注册人规范人工智能医学软件网络安全生存周期过程和为技术审评提供参考。不涉及相关行政审批事项，亦不作为法规强制执行，应在遵循相关法规的前提下使用本指导原则。

本指导原则是在现行法规和标准体系以及当前认知水平下制定的，随着法规和标准的不断完善，以及科学技术的不断发展，在使用过程中应对相关内容适时进行调整。

本指导原则是专门针对人工智能医疗软件特定的网络安全进行描述，通用要求请参考《医疗器械网络安全审查指导原则》。

一．适用范围

本指导原则适用于人工智能医学软件的网络安全的产品注册，人工智能医学软件指采用人工智能技术（AI）实现其预期用途的医学软件，包括人工智能医学独立软件和人工智

能医学软件组件。如采用机器学习、模式识别、规则推理等技术实现医疗用途的独立软件和软件组件。

二. 主要概念

（一）人工智能医学软件网络安全

通过采取必要措施、防范对数据、模型等攻击、侵入、干扰、破坏和非法使用以及意外事故，使软件设备处于稳定可靠运行的状态，以及保障数据、模型等的完整性、保密性、可得性的能力。此外，人工智能医学软件是基于海量医学数据和高算力算法的软件，医疗数据包含个人标识、健康状况以及医疗情况等相关信息，尽管信息安全、网络安全、数据安全的定义和范围各有侧重，即有联系又有区别，但本指导原则对三者不做严格区分，统一采用网络安全进行表述，综合考虑软件的信息安全和数据安全。

（二）网络安全特性

1.保密性

数据不被未授权实体（含产品、服务、个人、组织）获得或知悉的特性，即人工智能医学软件相关数据仅可由授权用户在授权时间以授权方式进行访问和使用。

2.完整性

数据的创建、传输、存储、显示未以非授权方式进行更改（含删除、添加）的特性，即人工智能医学软件相关数据是

准确和完整的，且未被篡改。

3.可得性

数据可根据授权实体要求进行访问和使用的特性，即人工智能医学软件自身和相关数据能以预期方式适时进行访问和使用。

除保密性、完整性、可得性三个基本特性外，人工智能医学软件产品网络安全还需考虑真实性、抗抵赖性、可核查性、可靠性等特性。注册人应结合人工智能医学软件产品的预期用途、使用场景、核心功能进行综合考量，从而确定人工智能医学软件产品网络安全特性的具体要求。

(三) 网络安全的风险级别

人工智能医学软件网络安全风险与软件风险存在差异，但是网络安全风险作为软件风险的重要组成部分，其风险级别亦可参照软件采用安全性级别进行表述 。在通常情形下，医疗器械网络安全的安全性级别与所属医疗器械软件的安全性级别相同；在特殊情形下，网络安全的安全性级别可低于软件的安全性级别，此时需详述理由并按网络安全的安全性级别提交相应注册申报资料。

三. 网络安全风险管理

（一）概述

随着人工智能技术的发展，越来越多医学软件产品使用人工智能技术实现辅助诊断、辅助分析等功能，大部分软件

具备网络连接功能以实现电子数据交换或远程控制，在提升医疗服务质量与效率的同时面临着网络攻击的威胁。人工智能医学软件是基于海量医学数据和高算力算法的软件，医疗数据包含个人标识、健康状况以及医疗情况等相关信息，鉴于医疗数据的特殊性，这些信息如被泄露、篡改或滥用，会影响健康护理、医学治疗以及科学研究效果，在更严重的情况下会导致医疗事故发生。另一方面，医疗数据大量涉及个人信息，数据的泄露、滥用和不正当披露会对个人信息安全造成侵害，甚至可能影响个人正常生活。因此，对于人工智能医学软件将带来更多的网络安全方面的考量，为保护个人健康医疗数据需要采取合理和适当的管理和技术保证措施，以达到以下目标:

a)保护医疗数据使用和披露过程中的保密性 、完整性和可得性 ；

b)确保医疗数据使用和披露过程的隐私性 ，保护个人隐私、个人权益。

（二）风险分析

人工智能医学软件除考虑软件自身网络安全能力建设外，还应当在软件全生命周期过程中考虑网络与数据安全过程控制要求，包括上市前设计开发阶段和上市后使用阶段。

上市前设计开发阶段包括算法数据构建阶段、算法训练生

成阶段。

算法数据构建阶段包括两个过程，分别是数据获取和数据整理：

1.数据获取部分是获取用于训练/更新算法的数据，其中数据包括自身私有数据、供应链数据、标准数据集、模型运行反馈数据等从多渠道获取的数据。

2.数据整理部分是为将获取的数据整合成为能够作为训练模型使用的数据，包括数据预处理、数据标注、数据集构建等过程及数据存储。

人工智能医疗软件，区别一般的医疗器械软件，人工智能医学软件依靠海量数据训练，所以应识别该过程中网络与数据安全风险并进行有效控制。该阶段的具体风险如表 1 所示（包括但不限于）：

表 1

风险点	阶段	风险描述及危害
数据投毒	数据获取	训练数据植入被修改的样本，使模型效果不佳，或定向使得某些实例被检测为指定的结果
模型后门注入	数据获取	模型植入部分具有特定模式特征的样本，训练后使模型产生后门
数据滥用	数据获取	可能使用未经用户授权的数据，或者用户要求“遗忘”的数据，造成法律风险
不安全的数据存储	数据整理	不安全的数据存储环境或配置管理缺陷，可能导致数据泄露，造成数据隐私泄露

算法训练与生成阶段包括模型选型实现及模型训练测试两个流程：

1.模型选型实现阶段是参照涉及与规范，选取适当的开发框架、预训练模型等，对模型进行编码实现。

2.模型训练测试阶段中训练过程一般划分多个批次进行迭代优化和验证，直到满足性能要求或达到最优结果。测试过程通常依赖大量的现实/模拟数据对模型表现进行评估，还包括对抗测试来确保模型抗鲁棒性。

模型训练与生成阶段具体风险如表 2 所示（包括但不限于）：

表 2

风险	阶段	风险描述及危害
预训练模型后门	模型选型实现	采用的预训练模型可能引入后门，进行 Fine-Tune 后后门可能继续存在，造成模型推理时触发错误预测
联邦学习恶意参与者投毒	模型训练测试	联邦学习恶意的参与者可能对数据投毒造成整个模型训练表现下降，或者植入后门
联邦学习恶意参与者模型隐私窃取	模型训练测试	联邦学习恶意的参与者可能利用共享的非加密梯度信息恢复相邻参与者的数据
开发框架&依赖库漏洞	模型选型实现	AI 模型开发依赖的框架、组件可能存在漏洞，造成模型预测时执行恶意程序
不可靠第三方环	模型训练测试	目前许多企业采用公有云训练环境，可

境风险		能由于公有云安全问题，造成数据及模型泄露，或被植入后门

上市后产品部署应用阶段包括产品部署使用和模型更新两个流程：

1.产品部署使用阶段是将算法部署到相应的业务流程、系统环境中，根据输入计算推理结果。算法使用可能涉及到从物理世界向数字平面转换等数据采集、预处理过程，也涉及到与系统、软件、硬件及第三方环境的集成和数据联动。

2.模型在线更新时，通常模型处理的数据分布可能发生变化，模型更新复用上市前流程，优化模型表现或修复模型缺陷。

产品部署应用阶段的具体风险如表 3 所示（举例如下）：

表 3

风险点	阶段	风险描述及危害
数字对抗样本	模型部署应用	在数字形态的模型输入中添加定向的噪声，欺骗模型，做出错误的预测
物理对抗攻击	模型部署应用	将数字形态对抗样本在物理世界中实现，造成模型错误预测，也可以定向添加噪声
模型后门触发	模型部署应用	输入数据中具有特定的模式，促使模型给出指定的结果，与正常推理相违背
模型窃取	模型部署应用	模型参数被窃取，AI 模型被逆向获取，造成企业核心竞争力 AI 资产损害
模型逆向	模型部署应用	构造特殊的输入及查询模型，通过输出结果来还原输入数据(可为训练集中样本)，从而造成数据

风险点	阶段	风险描述及危害
		隐私泄露
成员推理攻击	模型部署应用	利用模型输出置信度分布特征，判断某一样本数据是否在训练数据集中，造成训练集中数据隐私泄露
对抗资源消耗攻击	模型部署应用	利用对抗手段分析模型推理资源消耗，构造样本造成模型推理损耗提高，造成拒绝服务
输入预处理攻击	模型部署应用	攻击者对输入预处理方法（例如图片的预处理缩放算法）进行攻击，向其中定向插入扰动，使得输入发生顶下扰动，欺骗模型做出错误预测
反馈更新投毒	模型更新	很多模型引入了用户反馈对模型进行更新，攻击者利用投毒反馈数据来对模型进行“诱导”，为攻击谋利
AI 硬件后门	模型部署应用	硬件环境中植入后门，在模型部署后使用该硬件执行时触发后门
模型部署文件篡改	模型部署应用	模型部署环境可能存在漏洞，导致模型文件被恶意篡改，植入后门或者窃取信息
AI 模型侧信号窃取	模型部署应用	利用 AI 模型运行时 CPU、内存等资源进行侧信道分析， 能够实现对模型的关键信息窃取
软件&系统漏洞	模型部署应用	模型所属业务系统，与传统软件&系统安全面临的威胁相同，可能导致模型泄露或执行恶意程序
AI 模型滥用	模型部署应用	AI 模型服务被攻击者利用做恶意用途，可能造成危害

产品制造商按照产品特点对产品本身的风险进行分析，提供风险管理文档。基于风险分析，应在全生命周期中持续关注网络安全问题，包括但不限于设计开发、生产、分销、部署、更新维护、上市后监测等。重点关注数据采集、数据集

构建、算法学习、云计算、远程控制和外部软件环境等网络安全风险点。基于保密性、完整性、可得性等网络安全特性，确定人工智能医学软件网络安全能力建设要求。

四. 网络安全能力

人工智能医学软件产品网络安全的保障，是用户、网络设施提供方与注册人共同参与的结果。注册人可以参考医疗器械网络安全相关标准和技术报告，识别人工智能医学软件产品的网络安全能力，进行网络安全风险控制。注册人在对这些网络安全能力进行配置以配合用户进行网络安全风险管理时，应综合考虑人工智能医学软件产品的预期用途与使用场景限制，对于网络安全威胁应必要的识别、保护能力和适当的探测、响应、恢复能力，通用能力详见《医疗器械网络安全审查指导原则》中 22 项网络安全能力，此外针对人工智能医学软件提出 2 项新增网络安全能力。开发商应该决策网络安全能力：

1.训练数据保护验证

产品确保在数据构建阶段有效保护训练数据的能力，如提供安全的存储环境、完备的安全配置、确保数据不被恶意修改等，并确保训练数据得到用户授权。

2.模型训练生成阶段依赖的环境和库的安全

确保有训练环境、开发框架、组件、依赖库安全性能的验

证能力

基于以上所述网络安全能力，申请人应逐项分析所申报人工智能医学软件对于该项网络安全能力的适用性，若适用详述网络安全能力的实现方法和相应的防御措施，反之说明不适用的理由。

五、网络安全技术考量

（一）软件全生命周期

1.数据采集

采集的数据应进行数据脱敏以保护患者隐私，数据脱敏应描述用于保护受试者隐私的技术手段，如数据去标识化、数据匿名化等。数据转移应明确转移方法，转移方法应保证数据的安全，如对数据进行加密压缩、通过硬盘拷贝等方法。

2.数据整理

需在封闭或受控的网络环境下开展以防止数据污染。

3.数据标注

应有防止数据污染的防护措施，特别是在开放网络环境下。如标注软件部署在云服务器，应提供云服务商的名称、住所和资质，明确云服务的服务模式、部署模式、确保云服务器的网络安全能力。

4.数据集构建

需在封闭或受控的网络环境下开展以防止数据污染，描述

数据集存储方式与存储路径、安全控制、备份（方法、频次）、恢复方式，若数据集使用云服务存储，应提供云服务商的名称、住所和资质、访问路径、使用权限等。

5.算法训练

需在封闭或受控的网络环境下开展以防止数据污染，建议离线训练，避免训练数据被污染。

6.算法性能评估

需在封闭或受控的网络环境下开展以防止数据污染。若基于第三方数据库开展算法性能评估，应保证第三方测评数据库的网络安全与数据安全。

7.软件验证

需在封闭或受控的网络环境下开展以防止数据污染，应描述网络能力的验证，。

8.软件确认

软件确认一般需要模拟实际情况在开放网络下进行，应有防止数据污染的防护措施，应描述网络能力的确认。

9.上市后使用

上市后产品在医疗机构内部环境下运行，需要考虑医疗机构关于网络安全与数据安全的接口要求。产品使用时应有用户访问控制措施、数据传输过程应有加密措施等来保证网络安全。应描述上市后安全保障、环境保障，安全风险应急方

案。

（二）其他考量

1.软件运行环境

应满足软件运行需要，包括 CPU、GPU、RAM、ROM、网络、操作系统、数据库等要求。

2. 网络通信要求

2.1 网络架构

应保证网络架构的处理能力和带宽满足业务高峰的需要，软件的部署位置需要有边界防护措施，且通信线路、网络设备做好硬件冗余，保证软件的可用性。

2.2 通信传输

应采用校验码技术或密码技术保证通信过程中数据的完整性和敏感信息的保密性。

2.3 访问控制

不允许非授权设备或程序联到人工智能医学软件所在设备，不允许人工智能医学软件非授权用户访问并使用软件。跨网访问人工智能医学软件需要设置访问控制策略，对源地址、目的地址、源端口、目的端口和协议等进行检查，以允许/拒绝访问。在信息交互过程中应具备内容过滤能力，实现对内容的访问控制。

2.4 入侵防范

应具备检测、防止或限制从内、外部发起的网络攻击能力。当检测到有攻击行为时，能够记录攻击源 IP、攻击类型、攻击目的、攻击时间，并及时告警。具备对恶意代码进行检测和清除的能力，维护恶意代码防护机制的升级和更新。

2.5 安全审计

网络安全审计包括网络事件的日期和时间、用户、事件类型、事件是否成功及其他与审计相关的信息。建立审计信息的备份机制，确保信息最少保存 1 年。

3.主机安全要求

3.1 访问控制

应对登录主机的用户进行身份标识和鉴别，身份标识具有唯一性和不可抵赖性，对用户进行角色设置，授予所需的最小权限，制定登录处理策略，包括限制失败次数、连接超时、访问终端等。定期梳理用户帐号，及时删除或停用多余的、过期的和弱口令帐号。

3.2 入侵防范

应遵循最小安装的原则，仅安装需要的组件和应用程序，关闭不需要的系统服务、默认共享和高危端口，定期对人工智能医学软件及涉及到的主机、操作系统、中间件进行漏洞扫描、安全检测和恶意代码检测，并及时进行修补。应对主机进行实时安全监控，能够识别违规操作或攻击行为，并及

时告警。

3.3 安全审计

应具备安全审计能力，审计覆盖到每个主机用户。审计记录包括事件的日期和时间、用户、事件类型、事件是否成功及其他与审计相关的信息，并对审计信息进行备份，备份保留最少 1 年。

4.软件和数据安全要求

4.1 身份鉴别

应对人工智能医学软件的用户进行身份标识和鉴别，且具有唯一性和不可抵赖性，限制登录失败次数和连接时长，并对口令进行强制要求，规避弱口令情况。用户帐号信息丢失或遗忘时，应采用技术措施确保重置过程的安全，并对其进行二次验证，验证手段至少使用动态口令、手机验证码、密码技术、生物特征技术中的一种实现。

4.2 访问控制

人工智能医学软件应具备访问控制功能，对用户能够分配角色和权限，权限需要达到功能级。定期开展帐号梳理工作，及时删除或停用多余的、过期的帐号。

4.3 数据安全

应采用密码技术保证重要数据在传输和存储过程中的安全性，包括但不限于重要业务数据、重要审计数据、重要配

置数据、重要视频数据和重要个人信息等，避免未授权用户访问数据。密码技术的使用应遵循国家相关要求和规定。建立数据备份和恢复机制，定期开展演练，保证软件业务的连续性。

4.4 安全审计

人工智能医学软件应具备安全审计能力，审计覆盖到每个用户。审计记录包括操作的日期和时间、用户、操作类型、操作结果及其他与审计相关的信息，并对审计信息进行备份，备份保留最少 1 年。

5.软件运行对云安全的基本要求

云服务商的名称、住所和资质，明确云服务的服务模式、部署模式、确保云服务器的网络安全能力。

六. 网络安全管理要求

（一）安全策略和管理制度

1、安全策略

应制定网络安全工作的总体方针和安全策略，说明安全工作的总体目标、范围、原则和安全框架等。

2、管理制度

应对安全管理活动中的各类管理内容建立安全管理制度，对要求管理人员或操作人员执行的日常管理操作建立操作规程，并形成由安全策略、管理制度、操作规程、记录表单等构成的全面的网络安全管理制度体系。

3、制定和发布

应指定或授权专门的部门或人员负责安全管理制度的制定，通过正式、有效的方式发布，并进行版本控制。

4、评审和修订

应定期对安全管理制度的合理性和适用性进行论证和审定，对存在不足或需要改进的安全管理制度进行修订。

（二）安全建设管理

1、定级和备案

人工智能医学软件正式投入使用前应进行定级和备案，组织相关部门和有关安全技术专家对定级结果的合理性和正确性进行论证和审定，备案材料需报主管部门和相应公安机关备案。

2、安全方案设计

根据定级结果，结合实际环境，进行安全整体规划和安全方案设计，组织有关安全专家对方案进行审定，明确需要配套的安全设备和手段，以及实施过程。

3、相关软件的采购和使用

应确保所采用的安全产品符合国家的有关规定，密码产品与服务的采购和使用需符合国家密码管理主管部门的要求。

4、自行软件开发

应确保开发环境与实际运行环境物理分开，测试数据和测

试结果受到控制。通过建立开发管理制度和代码规范，约束开发人员的行为。开发过程所产生的文档应妥善保管。软件开发完成后，应进行漏洞扫描、代码审计、安全评估，保障安全性。

5、软件采购

软件采购前应对供应商进行审查，确认其是否具有完善的安全开发体系。软件正式交付前应进行漏洞扫描、代码审计和安全评估，并要求软件供应商提供软件设计文档和使用指南。在交付过程中应确立监理角色，确保软件本身及配套的安全产品符合预期要求。

6、软件供应商选择

应确保软件供应商的选择符合国家的有关规定，签订相关协议，明确整个建设过程各方需履行的职责和义务。定期监视、评审和审核供应商提供的服务和产品，并对其变更服务内容加以控制。

（三）安全运维管理

1、环境管理

应指定专门的部门或人员负责机房安全，对机房出入进行管理，定期对机房供配电、空调、温湿度控制、消防等设施进行维护管理，并对物品带进和带出进行相关要求和管控。

2、资产管理

应编制并维护与人工智能医学软件相关的资产清单，包括软件本身、主机、数据、第三方组件、中间件、数据库等，实时掌握资产的变化情况。

3、介质管理

应确保介质存放在安全的环境中，对各类介质进行控制和保护，介质在物理传输过程中的人员选择、打包、交付等情况进行控制，并对介质的归档和查询等进行登记记录。

4、设备维护管理

应对各种设备（包括备份和冗余设备）、线路等指定专门的部门或人员定期进行维护管理。通过建立设备和软硬件维护方面的管理制度，对其维护进行有效的管理，确保设备必须经过审批才能带离机房或办公地点，含有存储介质的设备带出工作环境时其中重要数据必须加密。对于含有存储介质的设备在报废或重用前，应进行完全清除或被安全覆盖，确保该设备上的敏感数据和授权软件无法被恢复重用。

5、漏洞和风险管理

应采取必要的措施识别安全漏洞和隐患，定期开展安全测评和恶意代码检测，对发现的安全漏洞和隐患及时进行修补或评估可能的影响后进行修补；

6、配置管理

应记录和保存基本配置信息，包括网络拓扑结构、各个设

备安装的软件组件、软件组件的版本和补丁信息、各个设备或软件组件的配置参数等，对于变更应进行控制，并及时更新基本配置信息库。

7、密码管理

应采用国家密码管理主管部门批准使用的密码算法和产品，并对密码的使用过程进行监控和审计。

8、变更管理

应明确变更需求，变更前根据变更需求制定变更方案和回退方案，变更方案经过评审、审批后方可实施，实施时应记录整个过程。

9、备份与恢复管理

对于需要定期备份的重要业务信息、系统数据及软件等应规定备份信息的备份方式、备份频度、存储介质、保存期等。

10、安全事件处置

应及时向安全管理部门报告所发现的安全弱点和可疑事件，并制定安全事件报告和处置管理制度，明确不同安全事件的报告、处置和响应流程，规定安全事件的现场处理、事件报告和后期恢复的管理职责等。

11、应急预案管理

应急预案包括启动预案的条件、应急组织构成、应急资源保障、事后教育和培训等内容，定期对应急预案重新评估，

修订完善。

七、注册申报相关要求

申请人根据《人工智能医疗器械注册审查指导原则》、《医疗器械网络安全技术审查指导原则》、《医疗器械软件技术审查指导原则》、《移动医疗器械技术审查指导原则》等相关指导原则要求，提交相应注册申报资料。

申请人应依据《医疗器械网络安全注册技术审查指导原则》提交网络安全描述文档。其中，基本信息应围绕数据类型进行描述；风险管理、验证与确认应基于 24 项网络安全能力进行分析和实施，不适用项详述理由；可追溯性分析报告应追溯网络安全需求、设计、测试、风险管理的相互关系；维护计划应包含网络安全日常维护计划、网络安全事件应急响应预案。

若使用云计算服务、移动计算终端，生产企业应依据《移动医疗器械注册技术审查指导原则》提交相应研究资料。其中，使用云计算服务应明确服务模式、部署模式、核心功能、数据接口、网络安全能力、服务（质量）协议等要求；使用移动计算终端应结合终端的类型、特点、使用风险明确相应性能指标要求。

八、编写单位

广东省药品监督管理局审评认证中心、华为终端有限公

司、腾讯医疗健康（深圳）有限公司、深圳迈瑞生物医疗电子股份有限公司

九、参考文献

[1]医疗器械监督管理条例[Z].

[2]医疗器械注册与备案管理办法[Z].

[3]医疗器械软件注册技术审查指导原则（2022 年修订版）[Z].

[4]医疗器械网络安全注册技术审查指导原则[Z].

[5]人工智能医疗器械注册审查指导原则[Z]

[6] YY/T 0316,医疗器械 风险管理对医疗器械的应用[S].

[7] YY/T 1833.1,人工智能医疗器械 质量要求和评价 第 1 部分：术语[S].

[8] YY/T 1833.2,人工智能医疗器械 质量要求和评价 第 2 部分：数据集通用要求[S].

[9] YY/T 1833.3,人工智能医疗器械 质量要求和评价 第 3 部分：数据标注通用要求[S].

[10] YY/T 1843-2022 医用电气设备网络安全基本要求[S]

[11] N. Papernot A Marauder’ s Map of Security and Privacy in Machine Learning: https://arxiv.org/pdf/1811.01134.pdf

[12] A Taxonomy and Terminology of 3 Adversarial

402 Machine Learning." Draft NISTIR 8269