附件1：

人工智能医疗器械注册审查指导原则

（征求意见稿）

前言

一、适用范围

二、主要概念

（一）人工智能医疗器械

（二）人工智能医疗器械类型

（三）人工智能算法更新

三、基本原则

（一）基于算法特性

（二）风险导向

（三）全生命周期管理

四、人工智能医疗器械生存周期过程

（一）需求分析

（二）数据收集

（三）算法设计

（四）验证与确认

（五）更新控制

五、技术考量

（一）注册单元与检测单元

（二）网络安全与数据安全

（三）移动计算与云计算

（四）人因设计与可用性

（五）压力测试

（六）对抗测试

（七）第三方数据库

（八）基于模型的算法

（九）迁移学习

（十）强化学习

（十一）联邦学习

（十二）生成对抗网络

（十三）基于数据的无监督学习

（十四）人工智能算法框架

（十五）人工智能芯片

六、算法研究资料

（一）算法研究报告

（二）算法更新研究报告

七、注册申报资料补充说明

（一）产品注册

（二）许可事项变更

（三）延续注册

八、编写单位

九、参考文献

人工智能医疗器械注册审查指导原则

（征求意见稿）

本指导原则旨在指导注册人建立人工智能医疗器械生存周期过程和准备人工智能医疗器械注册申报资料，同时规范人工智能医疗器械技术审评要求。

本指导原则是对人工智能医疗器械的一般要求。注册人应依据产品具体特性和风险程度确定本指导原则具体内容的适用性，若不适用应详述理由。注册人也可采用其他满足法规要求的替代方法，但应提供详尽的支持资料。

本指导原则是在现行法规、强制性标准体系以及当前科技能力、认知水平下制定的，随着法规、强制性标准体系的不断完善以及科技能力、认知水平的不断发展，本指导原则相关内容也将适时调整。

本指导原则是供注册人、审评人员和检查人员使用的指导文件，不涉及行政审批事项，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。

本指导原则作为数字医疗指导原则体系的重要组成部分，遵循医疗器械软件、医疗器械网络安全、移动医疗器械、医疗器械人因设计、医疗器械独立软件生产质量现场检查等数字医疗相关指导原则要求。

本指导原则是人工智能医疗器械的通用指导原则，其他含有人工智能技术的医疗器械指导原则可在本指导原则基础上结合具体情况进行有针对性的调整、修改和完善。采用人工智能技术进行医疗器械的设计、生产，亦可参考本指导原则适用要求。

一、适用范围

本指导原则适用于人工智能医疗器械（含体外诊断医疗器械）的注册申报，包括第二类、第三类人工智能独立软件和含有人工智能软件组件的医疗器械。

本指导原则也可用作人工智能医疗器械体系核查的参考。

二、主要概念

（一）人工智能医疗器械

本指导原则所述人工智能医疗器械是指基于“医疗器械数据”，采用人工智能技术实现其预期用途的医疗器械。

医疗器械数据是指医疗器械产生的客观医疗数据，如医学影像设备产生的医学图像数据（如X射线、CT、MRI、超声等图像）、医用电子设备产生的生理参数数据（如心电、脑电、血压、无创血糖等波形数据）、体外诊断设备产生的体外诊断数据（如病理图像、显微图像、有创血糖波形数据等）；通用设备产生的用于医疗用途的客观数据亦属于医疗器械数据，如数码相机拍摄的用于皮肤疾病诊断的皮肤照片、健康电子产品采集的用于心脏疾病预警的心电数据等。基于医疗器械数据包括单独使用医疗器械数据，或者以医疗器械数据为主联合使用非医疗器械数据（如电子病历、医学检查报告的结果文本等）。

人工智能是指机器表现出与人类智能相关行为的能力，通常是指通过感知周围环境做出合理行动以达到预期目标的计算机软件或系统。机器学习是指与人类学习行为相关的人工智能，通常是指通过整理现有数据和/或获取新数据以提升性能的计算机软件或系统。机器学习虽是人工智能的子集，但却为人工智能的核心领域，二者当前对于医疗器械而言含义基本相同，故本指导原则对二者不做严格区分，统一采用人工智能进行表述。

基于非医疗器械数据的医学人工智能产品，或者采用人工智能技术实现非医疗用途的医疗器械均非人工智能医疗器械。医学人工智能产品是否属于人工智能医疗器械，必要时可申请医疗器械分类界定。

（二）人工智能医疗器械类型

从医疗器械软件角度，人工智能医疗器械可分为人工智能独立软件和人工智能软件组件，故其类型划分可参考医疗器械软件指导原则相关维度。

人工智能医疗器械从用途角度可分为辅助决策类和非辅助决策类。其中，辅助决策是指通过提供诊疗活动建议辅助医务人员进行临床决策，如通过异常识别、病灶性质判定、自动制定手术计划进行辅助分诊、辅助检测、辅助诊断、辅助治疗等，相当于医务人员的“助手”。反之即为非辅助决策，包括流程优化、诊疗驱动，如诊疗流程简化、自动测量、三维重建等，相当于医务人员的“工具”。

人工智能医疗器械从功能角度大体上可分为控制功能、处理功能，处理功能又可分为前处理功能和后处理功能。其中，控制功能是指控制/驱动医疗器械硬件运行的功能，如控制机械臂运动等。前处理功能是指采集人体解剖、生理信息生成医疗器械数据过程的处理功能，如成像质量改善、成像流程简化等。后处理功能是指利用医疗器械数据生成诊疗信息过程的处理功能，如诊疗流程简化、自动测量、三维重建、异常识别、病灶性质判定、自动制定手术计划等。

人工智能医疗器械从算法角度具有多种类型划分维度。从学习策略角度可分为有监督学习和无监督学习，前者需要对训练数据进行标注，如回归、分类等典型算法，后者无需对训练数据进行标注，如聚类、降维等典型算法，前者对于数据标注的要求高于后者。从学习方法角度可分为基于模型的算法和基于数据的算法，前者采用统计模型、规则推理等方法，后者主要基于大数据方法，前者对于训练数据量的要求低于后者。从可解释性角度可分为白盒算法和黑盒算法，前者特征提取需要人为干预，可与现有医学知识建立关联，后者自动完成特征提取，难与现有医学知识建立关联，前者可解释性优于后者。

上述类型划分维度相互交叉，例如：前处理和后处理均可采用不同类型的人工智能算法实现辅助决策、非辅助决策用途。同时，同一维度亦不存在严格的划分界线，例如：在用途方面，通过图像识别技术进行流程优化则需考虑诊疗驱动相关要求，自动测量结果若为临床决策指标则属于辅助决策范畴；在功能方面，控制功能可与处理功能相结合，前处理过程可包含后处理功能；在算法方面，有监督学习和无监督学习可结合为半监督学习，统计模型可基于大数据，白盒算法和黑盒算法可结合为灰盒算法。

人工智能医疗器械可同时采用多种、多个人工智能算法，在前处理、后处理过程中实现辅助决策、非辅助决策用途。因此，应结合人工智能医疗器械的预期用途、使用场景、核心功能以及所用算法的类型特点、技术特征、组合形式开展相应产品质控工作，以保证产品的安全有效性。

（三）人工智能算法更新

人工智能算法特别是基于数据的算法，具有快速迭代更新的特性。人工智能算法更新属于软件更新范畴，故遵循软件更新的一般原则及要求：人工智能算法更新若影响到人工智能医疗器械的安全性或有效性则属于重大软件更新，应申请许可事项变更；反之，人工智能算法更新若未影响到人工智能医疗器械的安全性和有效性则属于轻微软件更新，通过质量管理体系进行控制，无需申请许可事项变更。

人工智能算法更新可分为算法驱动型更新和数据驱动型更新。其中，算法驱动型更新是指人工智能医疗器械所用算法、算法结构、算法流程、算法框架（详见后文）、输入输出等发生改变，通常属于重大软件更新。算法重新训练即弃用原有训练数据而采用全新训练数据进行算法训练，亦属于算法驱动型更新。

数据驱动型更新是指仅由训练数据量增加而发生的算法更新。数据驱动型更新是否属于重大软件更新原则上以算法性能评估结果（基于相同测试集）为准，算法性能评估结果若发生显著性改变则属于重大软件更新，即算法性能评估结果与前次注册相比存在统计学差异，反之属于轻微软件更新。

软件版本命名规则应涵盖算法驱动型更新和数据驱动型更新，明确并区分重大软件更新和轻微软件更新，其中重大软件更新列举全部典型情况。

人工智能医疗器械其他类型的算法更新、软件更新以及重大软件更新判定原则详见医疗器械软件指导原则、医疗器械网络安全指导原则。软件版本命名规则的一般要求详见医疗器械软件指导原则。

三、基本原则

（一）基于算法特性

人工智能技术从发展驱动要素角度实为基于模型/数据和算力的算法，其中模型/数据是人工智能技术的基础，算力是人工智能技术的保证，算法是人工智能技术的核心。

由于算力所用计算资源本身不属于监管对象，计算资源的监管要求取决于其所属的计算平台类型（参见后文人工智能芯片要求）。故从监管角度出发，人工智能医疗器械安全有效性评价以算法特性为核心重点关注其泛化能力，以模型/数据为基础重点关注其质控情况，同时兼顾算力不足与失效的影响及风险。

人工智能算法的类型不同，其算法特性、适用场景也不同，评价重点亦有所侧重；同时，不同类型的人工智能算法可组合使用，需结合各算法特性和算法组合形式进行整体评价。因此，应结合人工智能医疗器械的预期用途、使用场景、核心功能选择与之相适宜的人工智能算法或算法组合，基于算法特性并结合风险管理开展相应验证与确认工作。

以深度学习为例，其是指通过训练具有多个隐层的神经网络而获得输入输出映射关系的人工智能算法，实为基于海量数据和高算力的黑盒算法。因此，对于采用深度学习技术的人工智能医疗器械，重点关注其算法泛化能力、数据质控、可解释性等问题，同时，深度学习若与其他类型的人工智能算法组合使用，还应基于各算法特性重点关注算法组合的整体评价问题。

人工智能新算法研究处于深入发展阶段，如基于小样本数据、基于弱标注数据、基于非结构化数据的算法。人工智能医疗器械若使用人工智能新算法，亦需基于算法特性并结合风险管理开展相应验证与确认工作，以保证产品的安全有效性。

（二）风险导向

人工智能医疗器械的风险水平亦可用软件安全性级别进行表述，软件安全性级别越高，其生存周期质控要求越严格，注册申报资料越详尽，具体要求详见医疗器械软件指导原则。

人工智能医疗器械的软件安全性级别可基于产品的预期用途、使用场景、核心功能进行综合判定，其中预期用途主要考虑用途类型、重要程度、紧迫程度等因素，使用场景主要考虑使用场合、疾病特征、适用人群、目标用户等因素，核心功能主要考虑功能类型、核心算法、输入输出、接口等因素。亦可根据风险管理所确定的风险等级进行判定，软件安全性级别与风险等级的分级可以不同，但二者存在对应关系，因此可根据风险等级来判定软件安全性级别，但应在采取风险缓解措施之前进行判定。

人工智能医疗器械的主要风险从算法角度包括过拟合和欠拟合，其中过拟合是指算法对于训练数据过度学习而将非普遍规律作为重要特征，欠拟合是算法对于训练数据学习不充分而遗漏重要特征，均会降低算法泛化能力。从用途角度，辅助决策主要包括假阴性和假阳性，其中假阴性即漏诊，可能导致后续诊疗活动延误，特别是要考虑快速进展疾病的诊疗活动延误风险，而假阳性即误诊，可能导致后续不必要的诊疗活动；非辅助决策从算法设计目标能否得以实现角度，亦可参考辅助决策分为假阴性和假阳性。此外，进口人工智能医疗器械还需考虑中外差异风险，如中外在人种、流行病学特征、临床诊疗规范等差异。

人工智能医疗器械应结合产品的预期用途、使用场景、核心功能开展相应风险管理活动，采取有效的风险控制措施将风险降至可接受水平，并贯穿于人工智能医疗器械全生命周期过程。

（三）全生命周期管理

注册人应结合质量管理体系要求，参考软件、人工智能相关标准和良好工程实践，建立人工智能医疗器械生存周期过程，开展与软件安全性级别相匹配的产品质量保证工作，将风险管理、可追溯分析贯穿于生存周期全程，形成记录以供体系核查。

上市前应开展充分、适宜、有效的验证与确认活动，保证算法泛化能力满足临床需求，识别可预见风险并将其降至可接受水平，明确产品使用限制（含技术限制，下同）和必要警示提示信息。上市后持续开展算法泛化能力研究，结合用户投诉、不良事件和召回等情况，识别前期未预见风险并采取有效的风险控制措施将风险降至可接受水平，同时根据产品更新需求，经评估后实施更新活动，开展与之相适宜的验证与确认活动，保证算法泛化能力持续满足临床需求。

四、人工智能医疗器械生存周期过程

人工智能医疗器械生存周期（即生命周期）过程可基于软件生存周期过程予以建立，具体要求详见医疗器械独立软件生产质量管理规范及其现场检查指导原则。

考虑到有监督深度学习是当前人工智能医疗器械的主流算法，故本指导原则以有监督深度学习为例详述人工智能医疗器械生存周期过程质控要求，主要包括需求分析、数据收集、算法设计、验证与确认、更新控制等阶段。其他类型的人工智能算法可参照执行，不适用内容详述理由并予以记录。

（一）需求分析

需求分析以临床需求与风险为导向，结合产品的预期用途、使用场景、核心功能，综合考虑法规、标准、用户、产品、数据、功能、性能、接口、用户界面、网络安全、警示提示等需求，重点考虑数据收集、算法性能、使用限制等要求。

数据收集需考虑数据来源的合规性和多样性、数据分布的科学性和合理性、数据质控的充分性和有效性。数据来源在合规性基础上保证多样性，以提高算法泛化能力，例如：尽可能来源于多家、多地域、多层级的代表性临床机构，尽可能来源于多家、多种、多参数的代表性采集设备。数据分布的科学性和合理性需结合目标疾病流行病学特征予以考虑，包括但不限于疾病构成（如分型、分级、分期）、人群分布（如健康、患者，性别、年龄、职业、地域、生活方式）、统计指标（如发病率、患病率、治愈率、死亡率、生存率）等情况，以及目标疾病并发症与类似疾病的影响情况。数据质控的充分性和有效性需结合数据采集与数据标注的人员、设备、过程等影响因素予以考虑，具体要求详见后文。

算法性能需结合临床实际和产品定位，综合考虑假阴性与假阳性、重复性与再现性、鲁棒性/健壮性等性能指标要求，兼顾不同性能指标的制约关系，如假阴性与假阳性等。

使用限制需考虑产品禁用、慎用等场景，准确表述产品使用场景，提供必要警示提示信息。

（二）数据收集

数据收集基于合规性要求，主要考虑数据采集、数据整理、数据标注、数据集构建等活动的质控要求，以保证数据质量和算法训练效果。

1.数据采集

数据采集需考虑采集设备、采集过程、数据脱敏等质控要求，并建立数据采集操作规范。数据采集亦可使用历史数据，需结合样本规模、采集难度等影响因素合理选择数据采集方式。

采集设备质控包括采集设备的兼容性和采集特征等要求。兼容性基于数据生成方式（直接生成、间接生成）考虑采集设备的兼容性要求，如采集设备的名称、型号规格、制造商、性能指标等要求，若无需考虑兼容性要求详述理由并予以记录。采集特征考虑采集设备的采集方式（如常规成像、增强成像）、采集协议（如MRI成像序列）、采集参数（如CT加载电压、加载电流、加载时间、层厚）、采集精度（如分辨率、采样率）等要求。

采集过程质控包括人员管理、采集流程、采集质量评估等要求。人员管理考虑采集人员、审核人员的选拔（如职称、工作年限、工作经验、所在机构，若有国外人员明确其资质要求）、培训（如培训材料、培训方案）、考核（如方法、频次、指标、通过准则、一致性）等要求。采集流程考虑人员职责、采集步骤、结果审核等要求。采集质量评估考虑评估人员、评估方法、评估指标、通过准则等要求，并记录评估结果。

数据采集若使用历史数据，需列明采集设备及采集特征要求，并开展数据采集质量评估工作。

采集的数据应进行数据脱敏以保护患者隐私，数据脱敏需明确脱敏的类型（静态、动态）、规则、程度、方法。

2.数据整理

脱敏数据汇总形成原始数据库，不同模态的数据在原始数据库中需加以区分（下同）。

数据整理基于原始数据库考虑数据清洗、数据预处理的质控要求。数据清洗需明确清洗的规则、方法、结果，数据预处理需明确处理的方法（如滤波、增强、重采样、尺寸裁剪、均一化等）、结果。数据整理所用软件工具均需明确名称、型号规格、完整版本、制造商、运行环境，并进行软件确认。

数据经整理后形成基础数据库，需明确样本类型、样本量、样本分布等信息。样本类型以适用人群为单位可分为单一数据、数据序列（由多个单一数据组成，如结构序列、功能序列、时间序列）。样本量需考虑样本规模的充分性，明确样本总量确定依据。样本分布需考虑样本的多样性和偏倚控制问题，依据疾病构成、适用人群、数据来源机构、采集设备、样本类型等因素明确数据分布情况。

3.数据标注

数据标注作为有监督学习数据质控的关键环节，应建立数据标注操作规范，明确标注资源管理、标注过程质控、标注质量评估等要求。

标注资源管理包括人员管理和基础设施管理。人员管理考虑标注人员、审核人员和仲裁人员的选拔（如职称、工作年限、工作经验、所在机构，若有国外人员明确其资质要求）、培训（如培训材料、培训方案）、考核（如方法、频次、指标、通过准则、一致性）等要求。基础设施管理考虑标注场所（真实场所或模拟场所，模拟场所可根据产品实际情况调整模拟程度，详述调整理由并予以记录）、标注环境条件（如空间、照明、温度、湿度、气压）、标注软件（名称、型号规格、完整版本、制造商、运行环境、软件确认）等要求。

标注过程质控包括人员职责（如资质、数量、职责）、标注规则（如临床指南、专家共识、专家评议、文献分析）、标注流程（如标注对象、标注形式、标注轮次、标注步骤、结果审核）、分歧处理（如仲裁人员、仲裁方式）、可追溯性（如数据、操作）等要求。

标注质量评估包括评估人员、评估方法、评估指标、通过准则等要求，并记录评估结果。

数据经标注后形成标注数据库，其样本类型可分为数据块（如图像区域、数据片段）、单一数据（由多个数据块组成）、数据序列（由多个单一数据组成）。样本量、样本分布等要求及风险考量与基础数据库相同。

数据标注可使用自动标注软件，但自动标注结果不得直接使用，应由标注人员审核后方可使用；同时，自动标注软件亦需明确名称、型号规格、完整版本、制造商、运行环境等信息，并进行软件确认。

4.数据集构建

基于标注数据库构建训练集（用于算法训练）、调优集[[1]](#footnote-1)（若有，用于算法超参数调优）、测试集（用于算法性能评估），明确训练集、调优集、测试集的划分方法、划分依据、数据分配比例。训练集需保证样本分布具有均衡性，测试集、调优集原则上需保证样本分布符合临床实际情况，训练集、调优集、测试集的样本应两两无交集并通过查重予以验证。

为解决样本分布不满足预期的问题，可对训练集、调优集小样本量数据进行扩增，原则上不得对测试集进行数据扩增，对抗测试（详见后文）除外。数据扩增需明确扩增的对象、方式（离线、在线）、方法（如翻转、旋转、镜像、平移、缩放、滤波、生成对抗网络等）、倍数，扩增倍数过大应考虑数据偏倚的影响及风险。若采用生成对抗网络（详见后文）进行数据扩增，需明确算法基本信息以及算法选用依据。

数据经扩增后形成扩增数据库，需列表对比扩增数据库与标注数据库在样本量、样本分布（注明扩增倍数）等差异，以证实扩增数据库样本量的充分性以及样本分布的合理性。

（三）算法设计

人工智能算法作为人工智能医疗器械的核心，其设计主要考虑算法选择、算法训练、算法性能评估等要求。对于黑盒算法，算法设计应开展算法性能影响因素分析，同时建议与现有医学知识[[2]](#footnote-2)建立关联，以提升算法可解释性。

1.算法选择

算法选择提供所用算法的名称、类型、结构（如层数、参数规模）、输入输出、流程图、算法框架（含自研框架、现成框架）、运行环境等基本信息，并明确算法选用依据，包括选用的理由和基本原则。

若组合使用迁移学习（详见后文）、强化学习（详见后文），亦需提供算法基本信息以及算法选用依据。

2.算法训练

算法训练需基于训练集、调优集进行训练和调优，考虑评估指标、训练方法、训练目标、调优方法、训练数据量-评估指标曲线等要求。

评估指标建议根据临床需求进行选择，辅助决策可选择敏感性、特异性等指标，非辅助决策可选择图像质量、测量准确性等指标。训练方法包括但不限于留出法和交叉验证法，若组合使用联邦学习（详见后文）亦需明确算法选用依据，并提供算法基本信息。训练目标根据临床情况详述目标确定依据，提供ROC曲线等证据予以证实。调优方法明确优化策略和实现方法。训练数据量-评估指标曲线用于证实算法训练的充分性和有效性。

3.算法性能评估

算法性能评估作为软件验证的重要组成部分，需基于测试集对算法设计结果进行评估，综合考虑假阴性与假阳性、重复性与再现性、鲁棒性/健壮性等评估要求，以证实算法性能满足算法设计目标，并作为软件验证、软件确认的基础。亦可基于第三方数据库（详见后文）开展算法性能评估。

同时，开展算法性能影响因素分析以提升算法可解释性，详述影响算法性能的主要因素及其影响程度，如采集设备、采集参数、疾病构成、病变特征等因素，基于分析结果明确产品使用限制和必要警示提示信息。

此外，根据产品实际情况开展压力测试（详见后文）、对抗测试等测试，以全面深入评估算法性能。

（四）验证与确认

1.软件验证

软件验证是指通过提供客观证据认定软件开发、软件更新某一阶段的输出满足输入要求，包括软件验证测试（单元测试、集成测试、系统测试）、设计评审等系列活动。

软件验证通过法规、标准、用户、产品、数据、功能、性能、接口、用户界面、网络安全、警示提示等测试，验证软件的安全有效性，并作为软件确认的基础。

2.软件确认

软件确认是指通过提供客观证据认定软件满足用户需求和预期目的，包括软件确认测试（用户测试）、临床评价、设计评审等系列活动。根据产品实际情况，软件确认方式可单一使用，亦可组合使用。

软件确认测试基于软件需求在真实或模拟使用场景下予以开展，亦可基于测评数据库（详见后文）予以开展。

临床评价基本原则详见医疗器械软件指导原则，对于非辅助决策类功能，可基于核心功能开展同品种医疗器械比对，全新的功能、算法和用途原则上均需开展临床评价；对于辅助决策类功能，需基于核心算法开展同品种医疗器械比对，所选同品种医疗器械的临床证据需基于临床试验（可使用历史数据），全新的功能、算法和用途原则上均需开展临床试验。具体要求详见临床评价、临床试验设计、境外临床试验数据接受等通用指导原则，以及人工智能医疗器械临床评价等专用指导原则。

同时，开展算法性能比较分析，详述各类测试场景（含临床评价）算法性能变异度较大的原因，基于分析结果明确产品使用限制和必要警示提示信息。

最后，结合算法训练、算法性能评估、临床评价等结果开展算法性能综合评价，针对训练样本量和测试样本量过少、测试结果明显低于算法设计目标、算法性能变异度过大等情况，对产品的适用范围、使用场景、核心功能进行必要限制。

（五）更新控制

人工智能医疗器械若发生算法更新、软件更新，均应当按照质量管理体系的要求，开展与算法更新、软件更新的类型、内容和程度相适宜的验证与确认活动，将风险管理、可追溯分析贯穿于更新全程，形成记录以供体系核查。

对于算法更新，无论算法驱动型更新还是数据驱动型更新，均应开展算法性能评估、临床评价等验证与确认活动，以保证算法更新的安全有效性。对于软件更新，具体要求详见医疗器械软件指导原则、医疗器械独立软件生产质量现场检查指导原则。

总之，人工智能医疗器械所含的每个人工智能算法，均应独立开展需求分析、数据收集、算法设计、验证与确认、更新控制等活动，同时考虑人工智能算法组合的整体评价要求，以保证产品的安全有效性。

前期已开发的人工智能医疗器械若不满足本指导原则的适用要求，应开展差距分析并予以记录，明确产品使用限制和必要警示提示信息。必要时采取补救措施，以满足质量管理体系要求，并保证产品的安全有效性。

五、技术考量

（一）注册单元与检测单元

人工智能医疗器械的注册单元与检测单元划分原则详见医疗器械软件指导原则，人工智能独立软件、人工智能软件组件分别参照独立软件、软件组件的要求。

若软件核心功能相同但核心算法类型不同，则每类核心算法均需在产品技术要求中注明并检测。

（二）网络安全与数据安全

人工智能医疗器械应结合预期用途、使用场景、核心功能，基于保密性、完整性、可得性等网络安全特性，确定网络安全能力建设要求，以应对网络攻击和数据窃取等网络威胁，如算法框架漏洞攻击、数据污染等。具体要求详见医疗器械网络安全指导原则。

除此之外，还应在人工智能医疗器械全生命周期过程中考虑数据安全问题，包括上市前设计开发阶段和上市后使用阶段。数据转移需明确转移方法、数据污染防护措施、数据销毁等要求。数据整理、数据集构建、算法训练、算法性能评估、软件验证等内部活动需在封闭网络环境下开展以防止数据污染。数据标注、软件确认等涉及外方的活动需明确数据污染防护措施，特别是在开放网络环境下。各数据库（集）均需进行数据备份，明确备份的方法、频次、数据恢复方法。数据采集、上市后使用需考虑临床机构关于网络安全与数据安全的接口要求。

（三）移动计算与云计算

人工智能医疗器械若使用移动计算、云计算等技术，遵循相关指导原则要求。具体而言，移动计算相关要求详见移动医疗器械指导原则，云计算相关要求详见医疗器械软件指导原则，移动计算与云计算的网络安全相关要求详见医疗器械网络安全指导原则。

（四）人因设计与可用性

人工智能医疗器械建议加强用户接口人因设计以提升可用性，并将用户错误使用的风险降至可接受水平，特别是对软件用户界面。具体要求详见医疗器械人因设计相关指导原则。

（五）压力测试

本指导原则所述压力测试[[3]](#footnote-3)是指采用罕见或特殊临床真实数据样本开展的算法性能测试，侧重于评估算法泛化能力的极限。

注册人应根据产品实际情况开展压力测试，以全面深入评估算法性能。若未开展相应测试或测试结果不佳，均需对产品的适用范围、使用场景、核心功能进行必要限制，并在说明书中明确产品使用限制和必要警示提示信息。

（六）对抗测试

本指导原则所述对抗测试是指基于临床真实数据样本利用数据扰动、生成对抗网络等技术产生对抗样本，并采用对抗样本开展的算法性能测试，侧重于评估算法的鲁棒性/健壮性。

若条件具备，建议注册人开展对抗测试，以全面深入评估算法性能。若未开展相应测试或测试结果不佳，均需明确产品使用限制和必要警示提示信息。

（七）第三方数据库

第三方数据库可用于算法性能评估，但其类型、用途等情况各不相同，未必能够完全满足软件确认测试的要求。因此，使用第三方数据库进行软件确认测试，应评估其满足软件确认测试的充分性、适宜性和有效性。

可用于软件确认测试的第三方数据库即为测评数据库。测评数据库除满足数据库通用要求（如数据管理、网络安全与数据安全、可扩展性等）外，还应满足以下专用要求。

1.权威性：测评数据库的数据采集、数据标注及其质控工作应由相应临床专业领域的权威机构（如国家临床医学研究中心）负责，以保证数据准确性；标注人员、审核人员、仲裁人员需具备与其岗位职责相匹配的临床实践经验。

2.科学性：测评数据库的数据样本均应为临床真实数据，不得进行数据扩增；样本分布符合目标疾病流行病学特征情况，样本总量基于目标疾病流行病学统计指标、测试抽样误差控制要求，通过统计学计算予以确定；单次测试原则上根据测评数据库样本分布情况等比例随机抽取测试样本，且其样本量满足测试抽样误差控制要求。

3.规范性：测评数据库的数据采集、数据脱敏、数据清洗、数据预处理、数据标注、数据更换、数据管理、数据安全保证等数据治理活动以及测评活动均应建立质控程序文件，并满足可追溯性要求。

4.多样性：测评数据库的样本应来源于多家、多地域、多层级的代表性临床机构以及多家、多种、多参数的代表性采集设备，以保证数据多样性能够满足算法泛化能力评估要求；若条件具备，测评数据库可包含适当比例的罕见或特殊临床真实数据样本，以便开展压力测试深入评估算法泛化能力的极限。

5.封闭性：测评数据库应封闭管理，样本总量需远大于单次测试样本量，测评活动亦需封闭管理，以保证算法性能评价的客观性、公正性、公平性。

6.动态性：测评数据库应定期更换一定比例的数据样本，以保证其具备持续的多样性和封闭性；数据更换的频率、比例需有确定依据，并满足规范性要求；更换出库的数据样本由测评数据库责任方自行确定其处理方案。

基于测评数据库，利用数据扰动、生成对抗网络等技术，可衍生出对抗测评数据库。若条件具备，建议基于测评数据库建设对抗测评数据库，以全面深入评价算法性能。

第三方数据库经评估符合要求后可成为测评数据库（含对抗测评数据库），测评数据库名录另行制定并动态调整。

公开数据库因不具备封闭性而不能用作测评数据库，但可用于算法性能评估，若用于算法训练应重新进行质量评估。使用公开数据库应根据其使用情况开展必要评估工作，并在注册申报资料中予以说明。

其他类型第三方数据库可参照测评数据库和公开数据库的要求予以使用，同时考虑其适用场景、数据质量评估、时效性等要求。

（八）基于模型的算法

基于模型的算法对于训练数据的数量要求相对较低，通常为白盒算法，特征提取需要人为干预，可与现有医学知识建立关联，可解释性相对较高。

对于此类算法，有监督学习的模型/数据质控要求参照有监督深度学习执行，不适用内容详述理由并予以记录。同理，无监督学习的模型/数据质控要求亦参照有监督深度学习执行，其中数据标注不适用，其余不适用内容详述理由并予以记录。

此类算法无论有监督学习还是无监督学习，均需明确特征信息，包括但不限于特征分类（如人口统计学、生物学、形态学）、特征属性（如形态、纹理、性质、尺寸、边界）、特征展现方式（如形状、尺寸、边界、颜色、数量）等信息。

（九）迁移学习

迁移学习是指将在某领域或任务学习到的模型应用于不同但相关的领域或任务的人工智能算法，如将在自然图像领域学习形成的模型应用于医学图像领域。迁移学习主要包括基于特征和基于模型的算法，并可与深度学习、联邦学习等人工智能算法组合使用。

人工智能医疗器械若使用迁移学习算法，注册申报资料需明确算法的名称、类型、输入输出、流程图、运行环境等基本信息以及算法选用依据，并根据迁移学习的类型及其算法特性提供预训练模型的数据集构建、验证与确认等资料。

（十）强化学习

强化学习是指基于行动与环境的交互，实现行动从环境所获累积奖励最大化的人工智能算法，采用“试错”的学习策略。强化学习从不同角度具有不同类型划分维度，如有模型与无模型的算法、基于策略和基于价值的算法等，并可与深度学习等人工智能算法组合使用。

人工智能医疗器械若使用强化学习算法，注册申报资料需明确算法的名称、类型、输入输出、流程图、运行环境等基本信息以及算法选用依据，并根据强化学习类型及其算法特性提供算法验证与确认资料。

（十一）联邦学习

联邦学习是基于多个数据源协同建立学习模型的人工智能算法，即每个数据源独立进行本地数据学习，通过交换数据模型共同建立学习模型，实为分布式算法训练技术，亦属于隐私计算技术。联邦学习主要包括横向联邦（基于样本）和纵向联邦（基于特征）等算法，并可与迁移学习等人工智能算法组合使用。

人工智能医疗器械若使用联邦学习算法等隐私计算技术进行算法训练，注册申报资料需明确算法的名称、类型、输入输出、流程图、运行环境等基本信息以及算法选用依据，并根据联邦学习等隐私计算技术的算法类型及其算法特性提供算法验证与确认资料。

（十二）生成对抗网络

生成对抗网络是通过生成器和判别器的互相博弈学习而产生与临床真实数据类似样本的人工智能算法，主要用于数据扩增或者对抗测试的样本生成，并可与深度学习等人工智能算法组合使用。

人工智能医疗器械若使用生成对抗网络进行数据扩增，注册申报资料需明确算法的名称、类型、输入输出、流程图、运行环境等基本信息以及算法选用依据，并根据生成对抗网络类型及其算法特性提供算法验证与确认资料。

（十三）基于数据的无监督学习

基于数据的无监督学习具备自学习的能力，在临床机构部署后可通过持续学习该临床机构的数据而进行产品快速更新。此时，临床机构亦成为产品开发者，与注册人共同承担产品质量责任和法律责任；同时，此种更新对于产品安全有效性的影响具有高度不确定性，特别是人工智能黑盒算法。

因此，在当前法律法规体系和技术水平条件下，基于数据的无监督学习应关闭自学习功能，或者虽开放自学习功能但不得投入临床使用，即临床始终使用产品原有功能，自学习功能仅用于算法训练或临床科研。

注册人应按照质量管理体系要求，对自学习功能所产生的产品更新的安全有效性进行验证与确认，若适用还应根据法律法规要求申请许可事项变更，待批准后方能将自学习功能所产生的产品更新投入临床使用。

（十四）人工智能算法框架

人工智能算法框架从开发者角度可分为自研算法框架和现成算法框架，其中自研算法框架即注册人自行研发的算法框架，属于自研软件范畴；现成算法框架是指注册人所使用的由第三方开发的算法框架（如Tensorflow、PyTorch等），视为现成软件，现成算法框架开发者视为医疗器械供应商。

人工智能算法框架更新包括产品更新（即更换算法框架）、版本更新、补丁更新，其中产品更新、非效率型版本更新对于人工智能医疗器械而言属于重大软件更新，效率型版本更新（即运算效率单纯提高）、补丁更新对于人工智能医疗器械而言一般属于轻微软件更新，除非影响到人工智能医疗器械的安全性或有效性。人工智能算法框架同时发生多种类型的更新，同样遵循风险从高原则。

人工智能算法框架可参照自研软件、现成软件相关要求提交相应注册申报资料，均需明确算法框架的名称、类型（自研框架、现成框架）、型号规格、完整版本、制造商等信息。

（十五）人工智能芯片

人工智能芯片作为计算资源的组成部分，本身不属于监管对象，需要根据其所属的计算平台类型考虑监管要求。对于人工智能独立软件，人工智能芯片属于通用计算平台的组成部分，不属于产品结构组成，作为软件运行环境予以考虑。对于人工智能软件组件，人工智能芯片属于医用计算平台的组成部分，属于产品结构组成，需与产品进行整体评价。

无论何种情况，使用人工智能芯片均应在注册申报资料中明确其名称、型号规格、制造商、性能指标等信息。

六、算法研究资料

（一）算法研究报告

算法研究报告适用于人工智能算法或算法组合的初次发布和再次发布，包括算法基本信息、算法风险管理、算法需求规范、数据质控、算法训练、算法验证与确认、算法可追溯性分析、结论等内容，不适用内容详述理由。

1.算法基本信息

明确算法的名称、类型、结构、输入输出、流程图、算法框架、运行环境等基本信息以及算法选用依据。

其中，算法类型从学习策略、学习方法、可解释性等角度明确算法特性。算法结构明确算法的层数、参数规模等超参数信息。算法框架明确所用人工智能算法框架的基本信息，包括名称、类型（自研框架、现成框架）、型号规格、完整版本、制造商等信息。运行环境明确算法正常运行所需的典型运行环境，包括硬件配置、外部软件环境、网络条件；若使用人工智能芯片需明确其名称、型号规格、制造商、性能指标等信息。算法选用依据详述人工智能算法或算法组合选用的理由和基本原则。

2.算法风险管理

明确算法的软件安全性级别（轻微、中等、严重）并详述判定理由。提供算法风险管理资料，明确过拟合与欠拟合、假阴性与假阳性、数据扩增与数据偏倚等风险的控制措施。若无单独文档可提供软件风险管理资料，并注明算法风险管理所在位置。

3.算法需求规范

提供算法需求规范，若无单独文档可提供软件需求规范，并注明算法需求所在位置。

4.数据质控

提供数据来源合规性声明，列明数据来源机构的名称、所在地域、数据收集量、伦理批件编号等信息。

提供数据采集操作规范，根据数据采集方式明确采集设备、采集过程、数据脱敏等质控要求。

提供数据标注操作规范，明确数据整理、标注资源管理、标注过程质控、标注质量评估、数据安全保证等要求。

若适用，提供数据扩增情况说明，明确扩增的对象、方式、方法、倍数等信息。

依据疾病构成、适用人群、数据来源机构、采集设备、样本类型等因素提供原始数据库、基础数据库、标注数据库、扩增数据库的数据分布情况。

若数据来自公开数据库，提供公开数据库的基本信息（如名称、创建者、数据总量等）和使用情况（如数据使用量、数据质量评估、数据分布等）。

5.算法训练

依据疾病构成、适用人群、数据来源机构、采集设备、样本类型等因素提供训练集、调优集的数据分布情况。

明确算法训练所用的评估指标、训练方法、训练目标、调优方法，提供ROC曲线证明训练目标满足临床要求，提供训练数据量-评估指标曲线以证实算法训练的充分性和有效性。

6.算法验证与确认

依据疾病构成、适用人群、数据来源机构、采集设备、样本类型等因素提供测试集的数据分布情况。

提供假阴性与假阳性、重复性与再现性、鲁棒性/健壮性等算法性能指标评估结果，以证明算法性能满足算法设计目标。

若使用第三方数据库开展算法性能评估，提供第三方数据库的基本信息（如名称、创建者、数据总量等）和使用情况（如测试数据样本量、评估指标、评估结果等）。

若适用，提供算法性能影响因素分析报告，明确影响算法性能的主要因素及其影响程度，以及产品使用限制和必要警示提示信息。

若适用，提供压力测试、对抗测试等测试报告。若未开展相应测试或测试结果不佳，均需明确产品使用限制和必要警示提示信息。

若基于测评数据库进行算法确认，提供测评数据库的基本信息（如名称、创建者、数据总量等）和使用情况（如评估指标、评估结果等）。若基于临床评价方式进行算法确认，指向临床评价资料即可。

提供上述各类测试场景（含临床评价）下的算法性能评估结果比较分析报告，明确产品使用限制和必要警示提示信息。

7.算法可追溯性分析

提供算法可追溯性分析报告，即追溯算法需求、算法设计、源代码（明确软件单元名称即可）、算法测试、算法风险管理的关系表。

若无单独文档可提供软件可追溯性分析报告，并注明算法可追溯性分析所在位置。

8.结论

简述算法性能综合评价结果，明确对产品的适用范围、使用场景、核心功能所做的必要限制，并判定人工智能算法或算法组合的安全有效性是否满足要求。

（二）算法更新研究报告

算法更新研究报告仅适用于人工智能算法或算法组合的再次发布，在算法研究报告相应内容中明确算法更新情况。

其中，算法基本信息描述申报算法情况，若适用详述与前次注册相比的变化情况；算法风险管理、算法需求规范、数据质控、算法训练、算法验证与确认、算法可追溯性分析描述算法更新的具体情况；结论简述算法性能综合评价结果，并判定人工智能算法或算法组合更新的安全有效性是否满足要求。

七、注册申报资料补充说明

注册申报资料应在符合医疗器械注册申报资料要求等文件要求基础上，满足医疗器械软件、医疗器械网络安全、移动医疗器械等相关指导原则要求，同时需要重点关注以下要求。

（一）产品注册

1. 申请表信息

（1）人工智能独立软件

产品名称应符合通用名称命名规范要求，通常体现输入数据（如CT图像、眼底照片）、目标疾病（含病变、疾病的属性）、预期用途（如辅助分诊、辅助检测、辅助诊断）等特征词。

适用范围基于预期用途、使用场景、核心功能予以规范，包括但不限于处理对象、目标疾病、临床用途、适用人群、目标用户、使用场所、采集设备要求、使用限制等内容。

（2）人工智能软件组件

人工智能软件组件通常无需在注册证载明信息中体现。其软件功能名称可参照人工智能独立软件要求。若有，适用范围体现辅助决策类软件功能的情况。

2. 算法研究资料

在软件研究资料中以算法为单位，提交每个人工智能算法的算法研究报告，具体要求详见第六章。

3. 产品技术要求

产品技术要求若含有基于测评数据库测试的性能指标，应明确测评数据库的基本信息（如名称、型号规格、完整版本、责任方等）。

基于其他类型第三方数据库测试的性能指标，原则上无需在产品技术要求中体现。

4. 说明书

根据算法性能综合评价结果，对产品的适用范围、使用场景、核心功能进行必要限制，并在说明书中明确产品使用限制和必要警示提示信息。若适用，明确数据采集设备和数据采集过程相关要求。

对于辅助决策类产品，说明书应明确人工智能算法的算法性能评估总结（测试集基本信息、评估指标与结果）、临床评价总结（临床数据基本信息、评价指标与结果）、决策指标定义（或提供决策指标定义所依据的临床指南、专家共识等参考文献）等信息。此时若采用基于数据的人工智能算法，说明书还应补充算法训练总结信息（训练集基本信息、训练指标与结果）。

若产品采用人工智能黑盒算法，根据算法影响因素分析报告，在说明书中明确产品使用限制和必要警示提示信息。

对于软件安全性级别为严重级别的产品，原则上需提供用户培训材料。

（二）许可事项变更

1. 算法研究资料

根据人工智能算法的更新情况，在软件研究资料中以算法为单位，提交每个人工智能算法的算法更新研究报告（或算法研究报告），具体要求详见第六章。

2. 产品技术要求

若适用，产品技术要求变更对比表需体现测评数据库的变化情况。

3. 说明书

若适用，提交用户培训材料。

（三）延续注册

延续注册通常无需提交算法相关研究资料。若适用，根据注册证“备注”所载明的要求提交相应算法研究资料。

八、编写单位

本指导原则编写单位为国家药品监督管理局医疗器械技术审评中心。

九、参考文献

[1] 国家药品监督管理局. 医疗器械注册管理办法（修订草案征求意见稿），2021.3

[2] 原国家食品药品监督管理总局. 医疗器械说明书和标签管理规定（总局令第6号），2014.7

[3] 国家药品监督管理局. 医疗器械注册申报资料要求和批准证明文件格式（修订草案征求意见稿），2021.5

[4] 国家药品监督管理局. 医疗器械主文档登记事项（2021年第36号公告），2021.3

[5] 国家药品监督管理局. 医疗器械通用名称命名指导原则（2019年第99号通告），2019.12

[6] 国家药品监督管理局. 医疗器械安全和性能的基本原则（2020年第18号通告），2020.3

[7] 国家药品监督管理局. 医疗器械临床评价技术指导原则（征求意见稿），2021.5

[8] 国家药品监督管理局医疗器械技术审评中心. 医疗器械软件技术审查指导原则（第二版）（征求意见稿），2020.6

[9] 国家药品监督管理局医疗器械技术审评中心. 医疗器械网络安全技术审查指导原则（第二版）（征求意见稿），2020.9

[10] 原国家食品药品监督管理总局. 移动医疗器械注册技术审查指导原则（2017年第222号通告），2017.12

[11] 国家药品监督管理局医疗器械技术审评中心. 医疗器械人因设计技术审查指导原则（报批稿），2020.10

[12] 国家药品监督管理局医疗器械技术审评中心. 人工智能医疗器械临床评价技术审查指导原则（草稿），2021.

[13] 国家药品监督管理局医疗器械技术审评中心. 肺结节CT影像辅助决策软件注册技术审查指导原则（草稿），2021.

[14] 国家药品监督管理局医疗器械技术审评中心. 肺结节CT影像辅助决策软件临床评价技术审查指导原则（草稿），2021.

[15] 国家药品监督管理局医疗器械技术审评中心. 眼底照片辅助决策软件注册技术审查指导原则（草稿），2021.

[16] 国家药品监督管理局医疗器械技术审评中心. 眼底照片辅助决策软件临床评价技术审查指导原则（草稿），2021.

[17] 国家药品监督管理局医疗器械技术审评中心. 深度学习辅助决策医疗器械软件审评要点（2019年第7号通告），2019.7

[18] 国家药品监督管理局医疗器械技术审评中心. 肺炎CT影像辅助分诊与评估软件审评要点（试行）（2020年第8号通告），2020.3

[19] 原国家食品药品监督管理总局. 医疗器械生产质量管理规范（2014年第64号公告），2014.12

[20] 国家药品监督管理局. 医疗器械生产质量管理规范附录独立软件（2019年第43号通告），2019.7

[21] 国家药品监督管理局. 医疗器械生产质量管理规范独立软件现场检查指导原则（药监综械管〔2020〕57号），2020.5

[22] 国家药品监督管理局标准管理中心. 人工智能类医用软件产品分类界定指导原则（征求意见稿），2021.4

[23] 国家药品监督管理局标准管理中心. 辅助决策类医用软件产品分类界定指导原则（征求意见稿），2021.6

[24] [国家卫生健康委员会](https://www.baidu.com/link?url=DSbvgbNwImYdnPd0Ay8PiPR2GdQgKCIBQaWjXzDH7pa&wd=&eqid=cbcc6bcd0001f11f0000000260b5f3ce). 人工智能辅助诊断技术管理规范（国卫办医发〔2017〕7号），2017.2

[25] [国家卫生健康委员会](https://www.baidu.com/link?url=DSbvgbNwImYdnPd0Ay8PiPR2GdQgKCIBQaWjXzDH7pa&wd=&eqid=cbcc6bcd0001f11f0000000260b5f3ce). 人工智能辅助诊断技术临床应用质量控制指标（国卫办医发〔2017〕7号），2017.2

[26] [国家卫生健康委员会](https://www.baidu.com/link?url=DSbvgbNwImYdnPd0Ay8PiPR2GdQgKCIBQaWjXzDH7pa&wd=&eqid=cbcc6bcd0001f11f0000000260b5f3ce). 人工智能辅助治疗技术管理规范（国卫办医发〔2017〕7号），2017.2

[27] [国家卫生健康委员会](https://www.baidu.com/link?url=DSbvgbNwImYdnPd0Ay8PiPR2GdQgKCIBQaWjXzDH7pa&wd=&eqid=cbcc6bcd0001f11f0000000260b5f3ce). 人工智能辅助治疗技术临床应用质量控制指标（国卫办医发〔2017〕7号），2017.2

[28] YY/T 0287-2017 医疗器械 质量管理体系 用于法规的要求

[29] YY/T 0316-2016 医疗器械 风险管理对医疗器械的应用

[30] YY/T 0664-2020 医疗器械软件 软件生存周期过程

[31] YY/T 1406.1-2016 医疗器械软件 第1部分：YY/T 0316应用于医疗器械软件的指南

[32] YY/T 人工智能医疗器械质量要求和评价 第1部分：术语（报批稿），2020.

[33] YY/T 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（报批稿），2020.

[34] YY/T 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（草稿），2021.

[35] YY/T 人工智能医疗器械 肺部影像辅助分析软件 算法性能测试方法（草稿），2021.

[36] AIMDICP-WG6-2020-001 基于眼底彩照的糖尿病视网膜病变辅助决策产品性能指标和测试方法，2020.7

[37] AIMDICP-WG6-2020-002 基于胸部CT的肺结节影像辅助决策产品性能指标和测试方法，2020.7

[38] AIMDICP-WG12-2021-001 医学人工智能技术学名词（草稿），2021.

[39] FDA. Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data, 2012.7

[40] FDA. Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data, 2012.7

[41] FDA. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML) - Based Software as a Medical Device (SaMD) Draft, 2019.5

[42] FDA. Artificial Intelligence and Machine Learning(AI/ ML) Software as a Medical Device(SaMD) Action Plan, 2021.1

[43] 人工智能医疗器械创新合作平台. www.aimd.org.cn

[44] IMDRF AIMD. www.imdrf.org/workitems/wi-aimd.asp

[45] ITU&WHO AI4H. www.itu.int/go/fgai4h

1. 人工智能领域称之为验证集（Validation set）。为避免与医疗器械领域所用术语验证（Verification）、确认（Validation）相混淆，本指导原则将其改称为调优集。 [↑](#footnote-ref-1)
2. 医学知识作为医疗器械的外部参考准则，其自身评价问题不属于医疗器械安全有效性评价范畴。 [↑](#footnote-ref-2)
3. 本指导原则所述压力测试与软件工程领域的压力测试含义不同。 [↑](#footnote-ref-3)