ICS 11.040.99

CCS C30/49

中华人民共和国国家标准

GB/TXXXXX—XXXX

`

人工智能医疗器械 质量要求和评价 第4部分：可追溯性

Artificial intelligence medical device—Quality requirements and evaluation—Part 4: Traceability

立项草案稿

XXXX-XX-XX 发布　　XXXX-XX-XX 实施

目　　次

前　言

本文件按照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。

本文件是GB/T XXXX《人工智能医疗器械 质量要求和评价》的第2部分。GB/T XXXX已经发布了以下部分：

——第1部分：术语；

——第2部分：数据集通用要求；

——第3部分：数据标注通用要求；

——第4部分：可追溯性；

——第5部分：预训练模型。

本文件的附录A为规范性附录，附录B为资料性附录。

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

本文件由国家药品监督管理局提出并归口。

本文件起草单位：

本文件主要起草人：

引　言

近年来，人工智能医疗器械不断发展，成为医疗器械标准化领域的一个新兴方向。我国已初步建立人工智能医疗器械标准体系。在该标准体系中，GB/T XXXX《人工智能医疗器械 质量要求和评价》是基础通用标准，为开展细分领域的标准化活动提供指导，拟由八个部分组成。

——第1部分：术语。目的在于为人工智能医疗器械的质量评价活动提供术语。

——第2部分：数据集通用要求。目的在于提出数据集的通用质量要求与评价方法。

——第3部分：数据标注通用要求。目的在于提出数据标注环节的质量要求与评价方法。

——第4部分：可追溯性。目的在于明确人工智能医疗器械的可追溯性通用要求与评价方法。

——第5部分：预训练模型。目的在于规范人工智能医疗器械采用的预训练模型质量。

——第6部分：合成数据。目的在于规范人工智能医疗器械采用的合成数据质量要求与评价方法。

——第7部分：安装验证。目的在于加强人工智能医疗器械安装验证环节的质量控制。

——第8部分：伦理要求。目的在于从技术层面实现人工智能伦理的要求，保护人的权益。

本部分旨在加强人工智能医疗器械的可追溯性，对于规范人工智能医疗器械质量体系、指导人工智能医疗器械全生命周期健康发展具有重要意义。鉴于现阶段人工智能医疗器械产品的形态均为独立软件或软件组件，本部分在参照软件可追溯思想的基础上强调人工智能可追溯的特殊性。由于本领域的应用范围和技术尚在不断发展中，充分考虑适用性和风险后提出的新的质量要求和评价方法不受本部分的限制。同时，生产环节的可追溯性在现有《医疗器械生产质量管理规范》中已有规定，本文件不额外进行规范。

人工智能医疗器械 质量要求和评价 第4部分：可追溯性

1 范围

本文件规定了人工智能医疗器械的可追溯性通用要求和评价方法。本文件适用于人工智能医疗器械设计开发过程、生产过程、使用过程和更新过程的可追溯性活动。本文件不适用于人工智能医疗器械的流通环节。

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 42062 医疗器械 风险管理对医疗器械的应用

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

YY/T 1833.2-2022 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

3 术语和定义

YY/T 1833.1界定的以及下列术语和定义适用于本文件。

3.1

可追溯性 traceability

人工智能医疗器械的设计开发过程、生产过程、使用过程和更新过程能被记录的程度。

3.2

可追溯性矩阵 traceability matrix

记录两个或多个任务之间关系的矩阵。

[来源：GB/T 11457-2006, 定义2.1753，有修改]

示例：记录给定的算法需求和算法设计之间关系的矩阵。

4 可追溯性要求

4.1 概述

人工智能医疗器械的可追溯性是支撑人工智能自身透明度、可责性、可审计性的关键质量特性。人工智能医疗器械可追溯性的实现需要对设计开发过程、生产过程、使用过程和更新过程四个阶段的活动提出要求，确保人工智能医疗器械生命周期内需求、设计、验证与确认的一致性。

人工智能医疗器械设计开发过程重点考虑人工智能算法需求、算法设计、算法实现、验证与确认、风险管理、数据集六大要素及其内在关联，以过程文档和可追溯性矩阵的形式呈现。

生产过程参考医疗器械软件生产过程的通用要求，本文件不给出特殊要求。

使用过程考虑软件功能、软件界面和临床部署要求。

更新过程考虑产品在各个阶段的更新活动。

4.2 设计开发过程的可追溯性要求

4.2.1 通则

人工智能算法是人工智能医疗器械的核心，本条重点对人工智能算法设计开发可追溯性做出要求。

人工智能算法设计开发过程的可追溯性应分析算法需求、算法设计、算法实现、算法验证与确认、风险管理、数据集的关系，形成算法可追溯性分析报告。采用适当的可追溯性工具实现可追溯性的分析。

注1：软件设计开发过程中其他环节的可追溯性按照现有相关规定执行。

注2：附录A给出设计开发过程的算法可追溯性分析示例，作为参考。

注3：算法可追溯性分析报告可能独立于软件可追溯性分析报告，亦可能属于软件可追溯性分析报告的一部分。

4.2.2 算法需求可追溯性

算法需求可追溯应分析算法需求与风险分析之间的关系，确保需求的垂直可追溯性和水平可追溯性。需求层级可根据产品情况确定，下级需求应继承上级需求的要求。

算法需求可追溯性分析宜依托需求数据库进行，制定需求命名规则，对需求的垂直可追溯性和水平可追溯性进行跟踪。

注1：垂直可追溯性指的是贯穿软件开发文档到编码的需求可追溯性，指软件开发过程中前后阶段的可追溯，是需求的分解追溯。以典型的V模型为例，需求需要覆盖用户需求、架构设计覆盖软件需求、详细设计覆盖架构设计、代码覆盖详细设计等。

注2：水平可追溯性指的是指定测试级别的需求和相应测试文档之间的可追溯性，通过测试级别过程实现，是需求的验证确认追溯。以软件测试为例，测试条件覆盖对应的测试依据（例如：软件需求）、测试用例覆盖测试条件、测试规程覆盖测试用例，软件缺陷对应相关的测试规程。假如对应的测试用例执行通过，说明已经实现了对应的软件需求。

4.2.3 算法设计可追溯性

算法设计可追溯性应包含每个算法版本以及对应的训练集、测试集、调优集版本信息以及更新历史。

算法设计可追溯性应分析算法需求与算法设计、算法设计与源代码、算法设计与算法测试之间的关系。

注：若无单独的算法版本命名规则，也可以在软件版本命名规则中进行制定。

4.2.4 算法实现可追溯性

算法实现可追溯性应分析算法设计和源代码、源代码和测试用例之间的关系，确保算法设计规范中的每个要素得到实现。

4.2.5 算法验证与确认可追溯性

算法验证应针对人工智能算法或算法组合开展独立测试，宜使用测试用例测试。

算法验证可追溯性应分析算法设计与算法测试、风险管理与算法测试之间的关系。

算法确认可追溯性应分析算法需求与算法确认的关系。

4.2.6 风险管理可追溯性

风险管理可追溯性文档应按照GB/T 42062的要求编写，根据算法预期用途宜考虑算法风险、数据集风险、场景风险。风险管理可追溯性文档应分析风险管理与算法需求的关系。

注1：常见的算法风险包括决策失误（假阴、假阳）、不可预测、不可重复、过拟合、欠拟合等。

注2：常见的数据集风险包括敏感信息泄露、数据污染、数据偏倚、数据不完整等。

注3：常见的场景风险包括由临床部署场景导致的算法性能下降、效率下降等风险。

4.2.7 数据集可追溯性

数据集可追溯性的记录应按照YY/T 1833.2-2022条款5.2.11建立。

4.2.8 设计开发过程的算法可追溯性矩阵要求

人工智能医疗器械的算法可追溯性矩阵应呈现算法需求、算法设计、算法实现、验证与确认、风险管理、数据集等要素之间的关系，算法可追溯性矩阵宜依托以下文件形成闭环：

1）算法需求规范；

2）算法风险分析；

3）算法设计规范；

4）算法训练和测试记录；

5）算法验证报告；

6）临床评价报告或基于测评数据库的确认报告；

7）数据集说明文档和可追溯记录。

可追溯性矩阵的已识别关系应满足如下要求：

——正确性：确保映射关系成立；

——一致性：各文档内容不存在矛盾；

——完备性：各文档之间的映射关系全面；

——准确性：各文档的内容准确、清晰。

4.3 使用过程的可追溯性要求

4.3.1 软件功能可追溯性要求

为了支持使用过程的可追溯性，人工智能医疗器械宜记录人工智能算法的数据输入、决策过程、输出结论，例如以下适用的情形：

1）输入样本的标识信息；

2）用于预处理的算法组件名称和版本号；

3）用于人工智能辅助决策的算法框架、算法组件名称和版本号；

4）算法输入样本、输出结果的时间；

5）人工智能辅助决策的原始结果、预测概率、阈值；

6）使用者信息；

7）使用者对人工智能辅助决策的处理，如接受、修改、驳回；

8）使用者通过产品给出的最终决策。

人工智能医疗器械宜保护人工智能辅助决策过程记录的完整性和可得性，防止未授权的访问、修改、删除、覆盖。

4.3.2 软件界面可追溯性要求

软件界面宜提供可视化的方式区分人工智能辅助决策、人工决策产生的结果。

4.3.3 临床部署可追溯性要求

制造商宜提供技术手段，记录人工智能医疗器械临床部署环境和使用反馈。

注1：临床环境包括人工智能医疗器械运行的软件环境、硬件配置、网络资源等。

注2：使用反馈包括对人工智能医疗器械性能、安全等方面的反馈。

4.4 更新过程的可追溯要求

更新过程可追溯应记录更新相关的风险评估、更新需求分析、软件更新使用的训练集、调优集和测试集及其测试结果。制造商可参考4.2的要求，建立可追溯性矩阵。

5 评价方法

5.1 设计开发过程的可追溯性

检查可追溯性矩阵和过程文件，可识别的关系应体现正确性、准确性、一致性、完备性，给出结论，可用评价量表形式体现，结果应符合4.2的要求。

注：可追溯性报告示例见附录A。

5.2 使用过程的可追溯性

编写测试用例，检查软件功能，例如运行产生的日志，开展操作验证，应符合4.3.1的要求。

检查软件界面，开展操作验证，应符合4.3.2的要求。

检查临床使用反馈记录，编写测试用例，验证产品临床部署环境、使用反馈能否被授权用户记录，应符合4.3.3的要求。

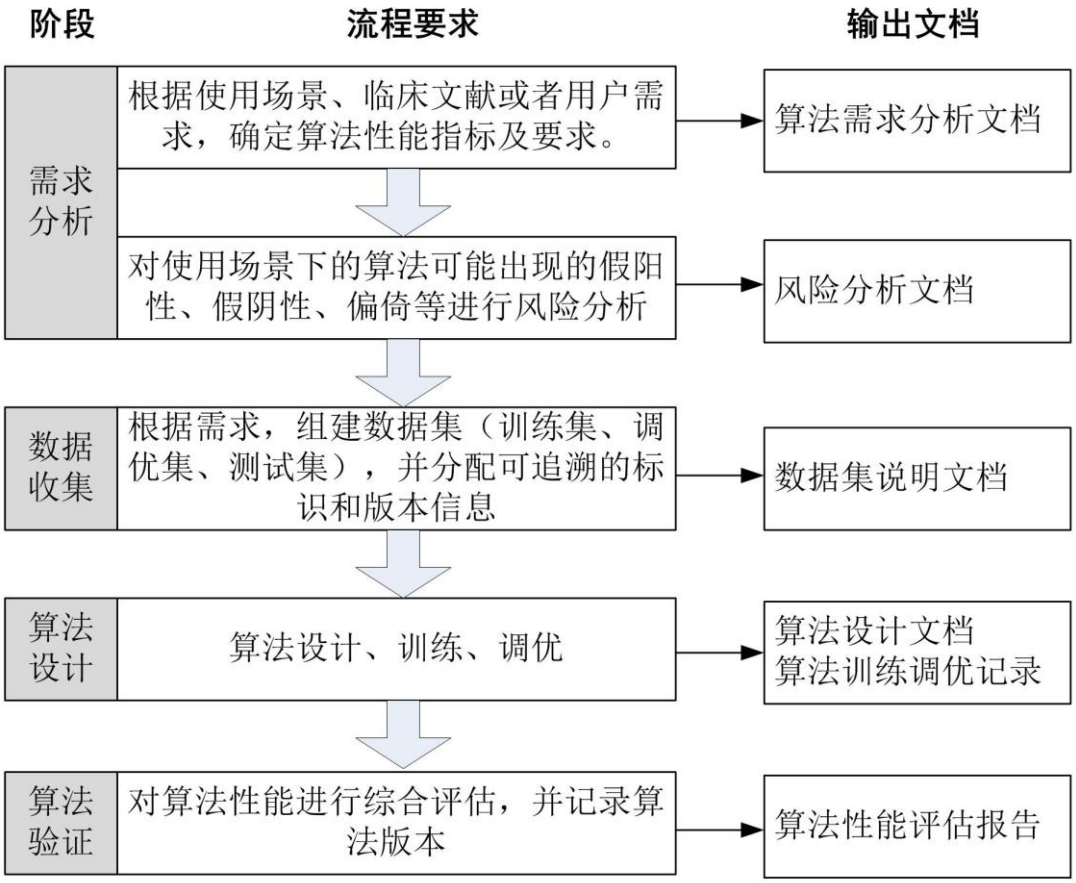
5.3 更新过程的可追溯性

检查软件更新的过程记录，根据正确性、准确性、一致性、完备性的要求，检查制造商提供的可追溯矩阵，应符合4.4的要求。

附 录 A
（资料性）
算法可追溯性分析示例

本附录主要分析某人工智能医疗器械的算法需求、算法设计、算法实现、算法验证与确认、风险管理，数据集的关系，分析并识别关系的正确性、一致性、完整性、准确性。

以需求为导向建立算法可追溯性映射关系，其过程举例如下：对某个预期用于在计算机断层扫描（computed tomography，简称CT）影像上检出肺结节的人工智能算法，首先在需求分析阶段根据使用场景、临床文献或用户需求描述，确定性能指标及要求，写入算法需求分析；依据算法需求分析，对该使用场景下算法可能出现的假阳性、假阴性、偏倚等风险进行分析，写入算法风险分析文档；根据需求分析、风险分析和流行病学特征，组建训练集、调优集、测试集，形成数据集说明文档、风险管理文档、可追溯记录；各数据集分配可追溯的标识和版本信息；开展算法训练与调优，形成算法训练调优记录。下一步，算法进行单元测试，形成测试记录；对算法性能进行综合性评估，通过测试的算法在发布时明确其版本信息。上述各个步骤的产出分别对应表A.1第二行的各个要素。图A.1描述了上述过程和输出文档，帮助读者理解。



图A.1 算法可追溯性流程图

表A.1给出了算法可追溯性分析矩阵的示例，用于表述各个文档之间的总体映射关系。表A.1中的算法测试用例识别号（identification,简称ID）属于统称，对应算法验证、确认环节的一系列测试用例，包括单元测试ID、集成测试ID、系统测试ID、用户测试ID。实际操作时，表A.1可能进一步展开为一系列表格的组合，共同支撑可追溯性分析，例如表A.2～表A.6。

表A.1 算法可追溯性分析矩阵

算法需求ID	算法需求简述	算法设计ID	源代码	算法测试用例ID	算法版本	数据集标识与版本	数据集说明文档
SRS001	算法检出性能	DS001	Script001	TC001	V1.0	Dataset001	DM001
注：表格第二行的各个要素（第二列除外）属于对文档、版本名称的举例，实际命名由制造商决定。							

表A.2作为示例,为梳理算法风险到算法需求之间的追溯关系提供参考,与具体的算法风险相对应，风险简述的内容考虑影响算法性能、安全的各种要素；适当时，也包含网络安全、数据安全风险。

表A.3作为示例，为梳理用户需求到算法需求的追溯关系提供参考。

表A.4作为示例，为梳理算法需求到算法设计的追溯关系提供参考。

表A.5作为示例，为梳理算法需求到算法验证的追溯关系提供参考。

表A.6作为示例，为梳理用户需求到算法确认的追溯关系提供参考。

表A.2 算法风险到算法需求的追溯

算法风险ID	风险简述	算法需求ID
RISK001	假阳性	SRS001
注：表格第二行的各个要素（第二列除外）属于文档名称的举例，实际命名由制造商决定。		

表A.3 用户需求到软件需求的追溯

用户需求ID	需求简述	算法需求ID
URS001	CT肺结节检出	SRS001
注：表格第二行的各个要素（第二列除外）属于文档名称的举例，实际命名由制造商决定。		

表A.4 算法需求到算法设计的追溯

算法需求ID	需求简述	架构设计ID	单元设计
SRS001	CT肺结节检出	ARCH001	UNIT001
注：表格第二行的各个要素（第二列除外）属于文档名称的举例，实际命名由制造商决定。			

表A.5 算法需求到算法验证的追溯

算法需求ID	需求简述	验证测试(单元+集成+系统)用例ID	验证结果
SRS001	CT肺结节检出	VERT001（UT001+SIT001+ST001）	通过/不通过
注：表格第二行的各个要素（第二列除外）属于文档名称的举例，VERT001为验证测试用例ID的示例，UT001为单元测试用例ID的示例，SIT001为集成测试用例ID的示例，ST001为系统测试用例ID的示例，实际命名由制造商决定			

表A.6 用户需求到算法确认的追溯

用户需求ID	需求简述	用户测试用例ID	确认结果
URS001	CT肺结节检出	VALT001	通过/不通过
注：表格第二行的各个要素（第二列除外）属于文档名称的举例，实际命名由制造商决定。			

对于算法迭代更新的情形，可追溯性分析过程在参考表A.1的基础上还需要考虑新旧模型的对比，具体形式由制造商决定。

表A.7给出算法可追溯性分析量表的一般示例，为正文5.1开展设计开发过程的可追溯性评价提供参考。在实际过程中，根据需要把表A.7进一步展开成多个量表开展可追溯性评价，例如采用表A.8的形式对算法验证、确认环节的可追溯性进行评价。

表A.7 算法可追溯性评价量表的一般举例

算法设计是否继承算法需求	□符合 □不符合
算法设计是否追溯到源代码	□符合 □不符合
算法的详细设计是否通过单元测试验证	□符合 □不符合
算法架构设计是否通过系统集成测试验证	□符合 □不符合
算法需求是否具有相应的测试用例	□符合 □不符合
用户需求是否具有相应的测试用例	□符合 □不符合

注1：算法需求的范围一般比软件需求小；软件需求包括用户、产品、法规标准、数据、功能、接口、网络安全、风险分析、软件性能等需求。

注2：人工智能算法可追溯性分析文档可能作为软件可追溯性分析文档的一部分。

注3：算法验证记录是否体现算法需求分析中的性能指标要求宜结合算法实现过程的关键证据评判，关键证据包括训练、调优记录等。

表A.8 验证与确认环节的可追溯性评价量表示例

评价项目	评价结果
算法验证记录是否体现算法需求分析中的性能指标要求	□符合 □不符合
算法风险分析是否考虑了算法需求分析中的性能指标要求	□符合 □不符合
数据集的偏倚控制能否响应算法风险分析	□符合 □不符合
算法设计使用的训练集、调优集是否具有标识和版本	□符合 □不符合
算法设计能否追溯到源代码	□符合 □不符合
算法性能指标通过了基于独立测试集的验证	□符合 □不符合
测试集具有标识、版本信息	□符合 □不符合

测试集的可追溯性评价主要依托数据集全生命周期的可追溯记录进行，记录的具体格式由数据集制造责任方确定。

参 考 文 献

[1] GB/T 11457-2006 信息技术 软件工程术语

[2] GB/T 36061-2018 电子商务交易产品可追溯性通用规范

[3] YY/T 0664-2020 医疗器械软件 软件生存周期过程.

[4] 《深度学习辅助决策医疗器械软件审评要点》[Z]. 国家药品监督管理局医疗器械技术审评中心，2019.

[5] 《人工智能医疗器械注册审查指导原则》[Z]. 国家药品监督管理局医疗器械技术审评中心，2022.

[6] 《医疗器械软件注册审查指导原则(2022年修订版)》[Z]. 国家药品监督管理局医疗器械技术审评中心，2022.

[7] 《医疗器械生产质量管理规范》[Z].原国家食品药品监督管理总局，2015.

[8] 《医疗器械生产质量管理规范附录独立软件》[Z].国家药品监督管理局，2021.

[9]World Health Organization. ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH: WHO Guidance[Z]. Geneva: World Health Organization; 2021.

[10]Joshua A. Kroll. Outlining Traceability: A Principle for Operationalizing Accountability in Computing Systems[C]. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 2021: 758–771. https://doi.org/10.1145/3442188.3445937.