ICS 11.040.70
C30/49

YY

中华人民共和国医药行业标准

YY/T 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

征求意见稿

2022年6月30日

××××－××－××发布　　××××－××－××实施

国家药品监督管理局　发布

目　　次

目　　次 I
前　　言 II
1 范围 1
2 规范性引用文件 1
3 术语和定义 1
3.1 1
3.2 1
4 可追溯性要求 1
4.1 概述 1
4.2 设计开发与验证过程的可追溯要求 2
4.3 软件功能与界面要求 3
4.4 临床部署与软件迭代可追溯要求 4
5 可追溯性评价方法 4
5.1 设计开发与验证过程的可追溯性 4
5.2 软件功能与界面的可追溯性 4
5.3 临床部署与软件迭代的可追溯性 4
附　录　A （资料性附录）软件可追溯性分析模板 5
A.1 软件可追溯性分析报告 5
A.2 数据集可追溯记录 7
参　考　文　献 8

前　言

YY/T 1833的总标题是《人工智能医疗器械 质量要求和评价》，已发布的部分如下：

——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。

本部分为YY/T 1833的第4部分。

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准的结构和编写》给出的规则起草。

请注意本文件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。

本文件由全国人工智能医疗器械标准化技术归口单位归口。

本文件起草单位：

本文件主要起草人：

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

本标准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。本标准适用于人工智能医疗器械独立软件、软件组件。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是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。

YY/T 1833.1-2022 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

YY/T 1833.3-202X 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

YY/T 0316-2016 医疗器械 风险管理对医疗器械的应用

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标准。

3.1

可追溯性 traceability

人工智能医疗器械的设计开发过程、数据、决策过程、应用情况、部署环境能被记录的程度

[来源：GB/T 36061-2018, 定义3.1，有修改]

注：决策过程适用于辅助决策类人工智能医疗器械。

3.2

可追溯性矩阵 traceability matrix

记录两个或多个开发过程的任务之间的关系的矩阵。例如，记录给定的算法的需求和设计之间的关系的矩阵。

[来源：GB/T 11457-2006, 定义2.1753，有修改]

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，需要对设计开发与验证过程、产品功能与界面、临床部署与产品迭代三个层面提出要求。

设计开发与验证过程考虑数据集、算法设计、算法需求规范、源代码、风险管理、算法测试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。

软件功能与界面考虑人工智能医疗器械的数据输入、决策过程、输出结论的可追溯，例如通过软件运行日志、可视化界面等形式呈现。

临床部署与软件迭代考虑临床使用环境、使用反馈、软件迭代过程的可追溯。触发软件迭代的因素宜纳入可追溯的范围。

4.2 设计开发与验证过程的可追溯要求

4.2.1 总则

制造商应在研发生产阶段对人工智能医疗器械开展可追溯性分析，考虑需求分析、软件设计、风险管理、源代码、验证确认、数据集，形成可追溯性分析报告。

注：附录 A 给出软件可追溯性分析的模板。

4.2.2 需求分析追溯

制造商应根据人工智能医疗器械软件的预期用途和使用场景，建立软件需求分析，重点考虑数据收集、算法性能、使用限制等要求。

制造商应建立需求阶段的可追溯，追溯分析软件需求与产品需求，软件需求与风险分析，人工智能算法需求（若无单独追溯，可与软件需求分析过程合并）与风险分析。

制造商应对软件需求进行评审，确保人工智能医疗器械声明周期内需求、设计、验证和确认的一致性。

制造商应确保需求的垂直可追溯性和水平追溯性，制造商可根据产品情况确定需求层级，下级需求应继承上级需求的要求。

制造商应建立需求数据库，制定需求命名规则，并对需求的垂直可追溯性和水平可追溯性进行跟踪。可通过追溯性分析工具予以实施。

注1：垂直可追溯性指的是贯穿软件开发文档到软件编码的需求可追溯，指软件开发过程中前后阶段的可追溯。以典型的V模型为例，系统需求需要覆盖用户需求、架构设计覆盖系统需求、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。以系统测试为例，测试条件覆盖对应的测试依据（例如：系统需求）、测试用例覆盖测试条件、测试规程覆盖测试用例，再到发现的缺陷对应相关的测试规程（更多的会体现在测试用例的对应上）。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 风险管理追溯

制造商应按照YY/T 0316的要求编写风险管理文档，根据软件预期用途和自治能力宜考虑算法风险、操作风险、数据集风险、场景风险。

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、过拟合等。

注2：常见的操作风险包括算法失控（用户无法控制产品）、人为操作失误等。

注3：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4：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。

4.2.4 数据集追溯

制造商应按照YY/T 1833.2条款5.2.11建立数据集可追溯的记录。

4.2.5 软件设计追溯

制造商应根据软件需求规范开展垂直追溯，形成软件设计规范。制造商应确保所有软件需求均可以追溯到恰当的设计元素。

4.2.6 软件实现追溯

4.2.6.1 源代码/软件单元

制造商通过编写源代码将软件设计规范转换为软件系统。

制造商应通过配置管理系统实现软件编码的配置管理，并追溯分析源代码/软件单元与软件设计、源代码与测试用例的关系。

制造商应确保软件设计规范中的每个要素均已通过源代码/软件单元予以实现。

4.2.6.2 人工智能算法设计

制造商应根据人工智能算法需求规范，建立算法设计规范，主要考虑算法选择、算法训练、算法调优等要求。

制造商应建立人工智能算法版本的命名规则以及根据4.2.4建立数据集命名规则。

制造商应追溯每个算法版本以及对应的训练集、调优集版本信息以及迭代历史。

4.2.7 软件验证与确认追溯

制造商应在软件验证与确认过程中建立可追溯记录，如单元测试、集成测试、系统测试、用户测试各级测试用例，用于开展可追溯性分析。

软件单元测试可追溯至软件详细设计规范中的要素和风险分析，并针对每个人工智能算法进行独立算法测试。

软件集成测试应追溯至软件概要设计规范，确保设计规范中要素予以验证。

软件系统测试追溯至软件需求和风险管理。

用户测试阶段追溯分析用户测试与软件需求、用户测试与风险管理的关系。

制造商应根据软件实际情况，开展与之相适宜的可追溯性分析活动。

4.2.8 设计开发与验证过程的可追溯性矩阵要求

人工智能医疗器械的需求可追溯性矩阵宜呈现需求分析、算法设计、源代码、验证与确认、风险管理、数据集等要素之间的关系，以算法为例，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）数据集说明文档和可追溯记录。

软件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，避免文件的错误使用；

——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 软件功能与界面要求

4.3.1 软件功能要求

软件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）算法辅助决策的原始结果、预测概率、阈值；

6）操作人员的标识信息；

7）操作人员对AI决策的处理，如接受、修改、驳回；

8）操作人员通过产品给出的最终决策。

软件宜保护人工智能算法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 软件界面要求

软件界面宜提供可视化的方式区分AI辅助决策、人工决策、人机协同产生的结果。

4.4 临床部署与软件迭代可追溯要求

4.4.1 临床部署

制造商宜提供技术手段，帮助授权用户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、人机协同效果等方面的反馈。

4.4.2 软件迭代

如适用，制造商宜记录软件上市后的迭代过程，包括触发产品迭代的原因、面向新场景、新用途的风险评估、迭代需求分析、产品迭代使用的训练集、调优集、再评价使用的测试集、测试结果。制造商可参考4.2的要求，建立可追溯性分析矩阵。

5 可追溯性评价方法

5.1 设计开发与验证过程的可追溯性

检查可追溯性分析报告和过程文件，可识别的关系应体现正确性、准确性、一致性、完备性，可填写评价量表，结果应符合4.2的要求；注：附录A给出量表的示例。

5.2 软件功能与界面的可追溯性

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

检查软件界面，开展实际的操作验证，应符合4.3.2的要求。

5.3 临床部署与软件迭代的可追溯性

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4.4.1的要求。

检查软件迭代的过程记录，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分析矩阵，应符合4.4.2的要求。

附 录 A
（资料性附录）软件可追溯性分析模板

A.1 软件可追溯性分析报告

A1.1 软件可追溯性分析范围

本次追溯性分析主要分析某人工智能医疗器械产品从需求规范、设计规范、源代码、软件测试和风险管理、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。

A1.2 软件需求可追溯性分析过程举例

制造商参考表A.1的形式，编写软件需求可追溯性矩阵。

表A.1 软件可追溯性分析矩阵

需求 ID	需求简述	软件设计 ID()	源代码(软件单元)	单元测试 ID	集成测试 ID	系统测试 ID	用户测试 ID
SRSXXX	用户管理	DSXXX DSXXX DSXXX DSXXX ……	XXX XXX XXX XXX ……	UTXXX UTXXX UTXXX UTXXX ……	SITXXX SITXXX SITXXX SITXXX ……	STXXX SITXXX SITXXX SITXXX ……	USTXX ……

表A.1在实现过程中，制造商可根据其质量管理体系要求调整并编写软件可追溯性矩阵，也可用分表的形式呈现，参考表A.2~表A.7。

表 A.2 用户需求到软件需求的追溯

用户需求ID	需求简述	软件需求ID
URSXXX	********	VPRSXXX

表A.3 产品风险到软件需求的追溯

产品风险 ID	风险简述	软件需求 ID
RISKXXX	********	PRSXXX

表A.4 网络安全风险到产品需求的追溯

网络安全风险 ID	风险简述	软件需求 ID
SRISKXXX	********	PRSXXX

表A.5 软件需求到设计的追溯

产品需求 ID	需求简述	架构设计	单元设计
PRSXXX	********	ARCHXXX	UNITXXX

表A.6 软件需求到软件验证的追溯

产品需求 ID	需求简述	验证测试(单元+集成+系统)用例 ID	验证测试用例结果
PRSXXX	********	VERTXXX，UTXXX SITXXX STXXX	通过/不通过

表A.7 用户需求到用户测试的追溯

用户需求 ID	需求简述	用户测试用例 ID	确认测试用例结果

URSXXX	********	VALTXXX	通过/不通过

A1.3 算法可追溯性分析过程举例

表A.8 算法可追溯性分析矩阵举例

算法需求ID	算法需求简述	算法设计ID	源代码（软件单元）	算法测试ID	算法版本	数据集标识与版本	数据集说明文档
SRSXXX	算法检出性能	DSXXX	XXX	UTXXX	V1.0	XXX	YYY
SRSXXX	算法分割性能	DSXXX	XXX	UTXXX	V1.0	XXX	YYY

以需求为导向建立算法可追溯性映射关系的过程举例如下：对某个预期用于病灶检出的人工智能算法，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；制造商为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，在算法设计文件中进行记录。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.8第二行的各个要素。

对于产品迭代的情形，可追溯性分析过程在参考表A.8的基础上宜增加新旧模型的对比，具体形式由制造商决定。

A1.3 可追溯性分析量表

表A.9 软件可追溯性分析量表举例

软件需求均由软件设计继承	□符合 □不符合
软件设计能追溯到源代码	□符合 □不符合
每个详细设计均由单元测试验证	□符合 □不符合
每个架构设计均由系统集成测试验证	□符合 □不符合
每个软件需求均由测试用例验证	□符合 □不符合
每个用户需求均由测试用例确认	□符合 □不符合

表A.10 算法可追溯性分析量表举例

算法需求分析中的性能指标要求具有依从性文件，如标准规范、临床指南、医学文献、用户需求描述	□符合 □不符合
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
数据集的说明文档能否响应算法需求分析	□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
每个算法性能指标均通过测试集验证	□符合 □不符合
测试集具有标识、版本信息	□符合 □不符合
算法完成验证和确认后，制造商应确认其发布版本	□符合 □不符合

注1：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与软件需求可追溯性分析文档合并。

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。

A.2 数据集可追溯记录

按照YY/T 1833.2条款5.2.11的要求建立数据集全生命周期的可追溯记录，具体格式由制造商确定。

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》[Z]. 国家药品监督管理局医疗器械技术审评中心，2022.

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

[8]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.