



团 体 标 准

T/SBIAORG 0002—2025

基于大语言模型的药物警戒个例安全性  
报告人工智能辅助信息提取规范

Specifications of AI-assisted data extraction for individual case safety reports  
in pharmacovigilance based on large language models

2025-05-08 发布

2025-05-09 实施

上海市生物医药行业协会 发布  
中国标准出版社 出版

## 目 次

前言 .....	Ⅲ
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 总体考量 .....	2
5 大语言模型在个例安全性报告信息提取场景下的系统实施 .....	2
6 系统使用和质量控制 .....	4
7 数据安全和隐私保护 .....	6
8 利益相关方沟通 .....	7
参考文献 .....	8

## 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。

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专利的责任。

本文件由上海市药品和医疗器械不良反应监测中心提出。

本文件由上海市生物医药行业协会归口。

本文件起草单位：上海市药品和医疗器械不良反应监测中心、上海罗氏制药有限公司、上海商汤科技开发有限公司、上海复宏汉霖生物制药有限公司、上海复星医药(集团)股份有限公司、江苏恒瑞医药股份有限公司。

本文件主要起草人：许晋、胡骏、巢艾伦、杨依晗、栾国琴、黄天娇、袁博、步伐、浦义虎、韩澎、黎洋。

# 基于大语言模型的药物警戒个案安全性 报告人工智能辅助信息提取规范

## 1 范围

本文件规定了应用大语言模型技术处理药物警戒个案安全性报告数据的考量点与要求。

本文件适用于药品上市许可持有人、境外上市许可持有人的境内责任人、药品经营企业等组织和机构。此外,也能为药品监管部门、医疗机构及相关技术开发商和服务提供商,在开发、实施和使用基于大语言模型的药物警戒数据处理系统时提供参考。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**药物警戒 pharmacovigilance**

对药品不良反应及其他与用药有关的有害反应进行监测、识别、评估和控制。

[来源:《中华人民共和国药品管理法》,第十二条]

### 3.2

**大语言模型 large language model ;LLM**

生成式人工智能的一种,专注于文本生成的模型。

[来源:《欧洲药品管理局在监管科学及药品监管活动中使用大型语言模型的指导原则》,第二部分]

### 3.3

**生成式人工智能技术 generative artificial intelligence**

具有文本、图片、音频、视频等内容生成能力的模型及相关技术。

[来源:《生成式人工智能服务管理暂行办法》,第二十二条]

### 3.4

**微调 fine-tuning**

为提升人工智能模型的预测精确度,一种先以大型广泛领域数据集训练,再以小型专门领域数据集继续训练的附加训练技术。

[来源:GB/T 41867—2022,3.2.31]

### 3.5

**多模态数据 multi-modal data**

多种形态的数据。

注:包含结构化数据(例如业务系统数据等)、半结构化数据(例如XML文件、JSON文件等)和非结构化数据(例如文本、语音和图像视频等)。

[来源:GB/T 42135—2022,3.1]

3.6

**知识 knowledge**

通过学习、实践或探索所获得的认识、判断或技能。

[来源:GB/T 23703.2—2010,2.1]

3.7

**知识库 knowledge base**

用于知识管理的一种特殊的数据库,以便于有关领域知识的采集、整理及提取。

3.8

**词元 token**

文档中非空的连续字形或音素序列。

[来源:ISO 24611:2012,3.21]

3.9

**提示词 prompt**

提示语

使用大模型进行微调或下游任务处理时,插入到输入样本中的指令或信息对象。

[来源:GB/T 45288.1—2025,3.5]

3.10

**提示工程 prompt engineering**

人工智能中执行任务的描述文本被嵌入到输入中的一个概念。

4 总体考量

在大语言模型系统的开发、部署、验证和质量控制的全生命周期中,应遵循以下原则,以确保安全、负责和有效地使用LLM系统。

- a) 人工审查和监督(Human Oversight):确保LLM系统在整个开发和使用过程中始终有人工审查和监督。
- b) 可问责(Accountability):建立明确的责任分配和归属制度。
- c) 透明性(Transparency):保持系统的透明度,并能够向相关方提供准确和及时的信息,包括系统的功能、操作流程、使用情况和局限性等。
- d) 可解释性(Explainability):提供LLM决策过程的可理解解释,包括完整的日志和记录等,以便在必要时进行审查和分析,确保系统操作的可追溯性。
- e) 风险管理(Risk Management):采用基于风险的方法来管理LLM应用,将LLM系统纳入药物警戒质量管理体系,持续识别、评估并减轻潜在的风险。
- f) 数据保护和隐私保护(Data Protection and Privacy Protection):加强数据保护和隐私保护,确保个人数据和敏感信息的处理符合《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》《中华人民共和国网络安全法》及数据跨境等相关法律法规要求。
- g) 治理(Governance):建立有效的治理框架,以监督和规范LLM系统的使用。
- h) 持续学习(Continuously Learning):鼓励组织和用户持续学习和适应合理合规使用LLM系统,以使LLM的使用价值最大化。

5 大语言模型在个例安全性报告信息提取场景下的系统实施

5.1 模型部署和架构设计

5.1.1 大语言模型的选择

5.1.1.1 宜选择与应用场景匹配的大语言模型。例如,个例报告源文件可能较长,宜优先考虑在长文本

能力具有优势的大语言模型,大语言模型支持的上下文词元(Token)宜不低于128 K;个例报告信息提取的规则较多,应优先选取在逻辑推理方面能力较强的大模型;个例报告源文件包含文本、图片、语音等多种格式,宜考虑在多模态数据处理能力具有优势的大语言模型;宜综合应用场景复杂度、技术成熟度、成本预算等因素选择合适的商业版大语言模型或者开源版大语言模型。

- 5.1.1.2 宜选择支持持续预训练微调和微调定制化的大语言模型,以更好地匹配药物警戒应用场景。
- 5.1.1.3 宜考虑组织或用户对模型部署的具体需求,选择符合需求的大语言模型。
- 5.1.1.4 宜选择具备持续学习和更新能力的大语言模型,以满足未来业务的可持续发展。

## 5.1.2 模型调优

调优的候选技术手段可包括提示工程、知识库管理、模型微调等,旨在提高大模型输出结果的准确率。

- a) 提示工程:采用在提示词中加入少量示例,加入思维链引导模型推理,拆分提取任务(如先提取怀疑用药,再提取怀疑用药的属性信息,包括开始日期、适应症等),提示语结构优化等基于提示语的调优手段,提升大语言模型对任务的理解,降低大语言模型逻辑推理复杂度,从而提高输出结果的准确率。
- b) 知识库管理:对于需要补充参考语料(与任务相关的额外背景信息、领域知识或示例等)进行提取的信息,可将补充语料放入外部知识库中。宜参考补充制药企业遵循的政策法规要求,如《药物警戒质量管理规范》《个例药品不良反应收集和报告指导原则》《个例安全性报告 E2B(R3)区域实施指南》等,以及领域知识,如 MedDRA(Medical Dictionary for Regulatory Activities)、CTCAE(Common Terminology Criteria for Adverse Events)等。
- c) 模型微调:模型微调通常依赖于至少千条标注好的训练数据,并需通过自动化评估系统来检验微调成效。在决策是否进行微调或确定微调策略时,宜全面评估数据质量、标注成本和测试集的适用性。

## 5.1.3 模型部署和架构设计

可依据实际需求选择适宜的部署策略,在架构设计时还应考虑数据保护的合规性,确保符合数据安全、隐私保护、伦理、数据跨境等相关要求。

## 5.2 模型性能评估

大语言模型在药物警戒个例报告信息抽取任务的性能评估宜采用混淆矩阵(见图1)中的准确率(Accuracy)、精确率(Precision)、召回率(Recall)指标。

综合考虑药物警戒实际业务的评测数据情况,宜将混淆矩阵的 TP(True Positive)、FN(False Negative)、FP(False Positive)、TN(True Negative)转换为药物警戒场景下的业务术语:正提、漏提、错提、提错和正拒。术语定义和映射关系如下:

- a) 正提:源文件信息存在,且被大语言模型正确提取;
- b) 漏提:源文件信息存在,但未被大语言模型提取出来;
- c) 错提:源文件信息不存在,但大语言模型提取出了信息;
- d) 提错:源文件信息存在,但大语言模型提取出了错误信息;
- e) 正拒:源文件信息不存在,大语言模型没有提取出信息。

补充说明:

- a) 源文件:药物警戒个例安全性报告的原始信息文件,例如文献、电子邮件、来自市场研究项目或患者支持项目的报告表等;
- b) 源文件信息:药物警戒个例安全性报告所需填写的字段信息,例如患者信息(姓名、性别、出生日期等)、报告人信息(姓名、职业、所在单位、联系电话、电子邮箱等)、药品信息(批准文号、商品

名、通用名、剂型、用法用量等)、不良反应信息(不良反应术语、严重性、报告人评价等)。

	预测正类	预测负类
实际正类	TP (正提)	FN (漏提)
实际负类	FP (错提或提错)	TN (正拒)

图1 性能评估混淆矩阵

### 5.3 评估指标计算

#### 5.3.1 准确率计算见式(1)。

$$\text{准确率} = \frac{TP + TN}{TP + TN + FN + FP} = \frac{\text{正提} + \text{正拒}}{\text{正提} + \text{正拒} + \text{漏提} + \text{错提} + \text{提错}} \quad \dots\dots\dots(1)$$

注：准确率是最直观的性能指标，它衡量的是模型预测正确的样本占总样本的比例，旨在全面评估模型的整体性能。

#### 5.3.2 精确率计算见式(2)。

$$\text{精确率} = \frac{TP}{TP + FP} = \frac{\text{正提}}{\text{正提} + \text{错提} + \text{提错}} \quad \dots\dots\dots(2)$$

注：精确率衡量的是模型预测为正类的样本中，实际为正类的比例。精确率关注的是预测结果的准确性，即减少误报。

#### 5.3.3 召回率计算见式(3)。

$$\text{召回率} = \frac{TP}{TP + FN} = \frac{\text{正提}}{\text{正提} + \text{漏提}} \quad \dots\dots\dots(3)$$

注：召回率衡量的是所有实际为正类的样本中，被模型正确预测为正类的比例。召回率关注的是模型捕捉到的正类样本的完整性，即减少漏报。

#### 5.3.4 $F_\beta$ 分数( $F_\beta$ Score)计算见式(4)。

$$F_\beta \text{分数} = (1 + \beta^2) \frac{\text{精确率} \times \text{召回率}}{(\beta^2 \times \text{精确率}) + \text{召回率}} \quad \dots\dots\dots(4)$$

注： $F_\beta$ 分数是精确率和召回率的加权调和平均值。 $\beta$ 是一个参数，用于调整精确率和召回率的相对重要性。当 $\beta$ 系数取值等于1时， $F_\beta$ 分数简化为 $F_1$ 分数，即精确率和召回率的调和平均值。 $\beta$ 系数取值大于1时，召回率的权重更高。 $\beta$ 系数取值小于1时，精确率的权重更高。例如，在比较关注不良反应漏报率时，召回率(灵敏度)是一个非常非常重要的指标， $\beta$ 系数可以大于1，而在个例安全性报告信息提取场景中，更加关注精确率，本文件建议 $\beta$ 系数取0.5。

## 6 系统使用和质量控制

### 6.1 制定大语言模型系统使用策略

根据系统输出准确率、法规合规性要求、数据准确度要求等方面进行综合评价，确定大语言模型使用

策略。具体可分为基本依赖大语言模型进行个例安全性报告信息的自动提取、大语言模型与人工相结合但以大语言模型为主进行个例安全性报告信息的提取和大语言模型与人工相结合但以人工为主进行个例安全性报告信息的提取。

使用策略建议见表1。

表1 模型使用策略建议

模型输出准确率/ 全部数据点准确率	系统可用性等级	关键数据点准确率	策略建议
>98%	非常高	不适用	以大语言模型为主,需要人工控制质量
>90%	高	>98%	以大语言模型为主,部分依赖人工,需要人工控制质量
		<98%	以人工为主,部分依赖大语言模型,需要人工控制质量
≥60%;≤90%	中	不适用	以人工为主,部分使用大语言模型与人工相结合方式,需要人工控制质量
<60%	低	不适用	大语言模型与人工相结合但以人工为主或不考虑大语言模型
<p>注1: 关键数据点包括并不限于不良事件术语、药品信息、患者信息、报告人严重性判断和相关性判断等。</p> <p>注2: 上述建议主要适用于有法规递交要求的安全性信息提取。如果所处理数据只为内部使用,不涉及任何法规递交,如既往数据库的数据迁移,则需要全面评估,决定使用策略。</p>			

## 6.2 标准操作流程(SOP)

基于现有流程体系和大语言模型使用策略,建立相关标准操作流程并及时更新,嵌入现有流程体系,确保流程体系的完善性和准确性,从而确保合规性。

## 6.3 结果复核

应对大语言模型系统输出的结果进行人工复核。可基于实际应用场景和合规性要求,设计复核方式和流程。

## 6.4 定期回顾和更新

6.4.1 应定期对大语言模型系统的输出结果进行总结和分析,识别精确率较低的数据元素,并采取有针对性的纠正行动,提高大语言模型的提取准确率。

6.4.2 应及时针对大语言模型系统的知识库进行更新和补充。

6.4.3 根据具体情况决定是否对大语言模型进行升级,如决定升级,应在开展充分的性能评估后进行升级。

6.4.4 在针对大语言模型系统的重要更新和升级后,必要时再次进行综合评估,更新大语言模型使用策略,并相应更新相关SOP和具体操作流程。

## 6.5 文档记录

6.5.1 大语言模型系统的提取结果文档不应视作源文件,应存档最原始的源文件文档。

6.5.2 大语言模型系统不宜作为个例安全性报告数据存档系统,相关数据应及时清理,但需要大语言模型使用记录等相关事宜的记录。

## 6.6 培训

按照药物警戒流程体系,针对使用大语言模型系统的人员进行定期培训,确保相关人员按照药物警戒流程体系和相关法规要求开展相关工作。

## 6.7 合规性

实时监测大语言模型使用过程中的合规性。对使用过程中产生的偏差需要及时记录和纠正,必要时采取相应的纠正和预防措施,确保合规性。

## 6.8 审计

按照药物警戒相关审计要求,定期对大语言模型使用的相关流程进行审计,确保合规使用。

## 6.9 业务持续性

按照公司或药物警戒部门业务持续性要求,进行业务持续性评估和必要的演练,确保在大语言模型系统使用出现重大问题时,能够正常开展个例安全性报告数据处理工作。

# 7 数据安全和隐私保护

## 7.1 缓解措施

### 7.1.1 通则

在规划、设计和运行系统时应充分考虑数据安全和隐私保护,宜实施有效的缓解措施,以防止:

- a) 提示词注入;
- b) 不安全地输出处理;
- c) 数据中毒;
- d) 模型拒绝服务;
- e) 供应链漏洞;
- f) 不安全的插件设计;
- g) 过度代理;
- h) 过度依赖;
- i) 模型盗窃;
- j) 漂移。

缓解措施宜包括但不限于7.1.2~7.1.4中的描述。

7.1.2 遵循中国关于数据安全与隐私保护方面的法律法规,特别是可以参考人工智能方面的法律法规要求,如《生成式人工智能服务管理暂行办法》。

7.1.3 可参考国际组织对于数据信息安全资质的要求:

- ISO/IEC 27001:2022 信息安全管理体系;
- ISO/IEC 29151:2017 个人身份信息保护管理体系;
- ISO/IEC 27701:2019 隐私信息安全管理;
- ISO/IEC 27017:2015 云安全管理;
- ISO/IEC 27018:2019 公有云个人信息保护;
- ISO/IEC 23894:2023 信息技术-人工智能-风险管理指南;

——BS 10012 个人信息安全管理体系。

7.1.4 应符合组织内部的关于数据安全和隐私保护方面的合规要求。例如,某些组织内部会依据数据的敏感度进行分类(如公开、内部、机密、高度机密)。对高敏感数据实施严格的控制措施,以防止未经授权的访问和泄露。

## 7.2 模型云服务提供商的报告或证书要求

Cloud 云服务用户数据保护能力检验报告。

Cloud 云服务用户数据保护能力检验证书。

AIDC 智算中心基础设施支撑平台-网络信息安全等级保护三级。

## 7.3 条件限制

使用知识产权、个人数据或个人敏感信息等时,应预先设定使用限制或强制前提条件。

## 8 利益相关方沟通

大语言模型系统使用方应与监管机构、医疗机构等利益相关方保持沟通,确保其了解 LLM 的使用情况和潜在影响,确保 LLM 的使用满足监管要求,并能够支持监管决策。

参 考 文 献

- [1] 中华人民共和国药品管理法
  - [2] 中华人民共和国个人信息保护法
  - [3] 中华人民共和国数据安全法
  - [4] 中华人民共和国网络安全法
  - [5] 药物警戒质量管理规范(国家药品监督管理局公告2021年第65号)
  - [6] 个例药品不良反应收集和报告指导原则(国家药品监督管理局2018第131号通告)
  - [7] 个例安全性报告E2B(R3)区域实施指南(国家药品监督管理局药品评价中心2019年11月22日发布)
  - [8] 生成式人工智能服务管理暂行办法(国家互联网信息办公室 中华人民共和国国家发展和改革委员会 中华人民共和国教育部 中华人民共和国科学技术部 中华人民共和国工业和信息化部 中华人民共和国公安部 国家广播电视总局令第十五号)
  - [9] GB/T 23703.2—2010 知识管理 第2部分:术语
  - [10] GB/T 41867—2022 信息技术 人工智能 术语
  - [11] GB/T 42135—2022 智能制造 多模态数据融合技术要求
  - [12] GB/T 45288.1—2025 人工智能 大模型 第1部分:通用要求
  - [13] 欧洲药品管理局在监管科学及药品监管活动中使用大型语言模型的指导原则(EMA Guiding principles on the use of large language models in regulatory science and for medicines regulatory activities)
  - [14] ISO 24611:2012 Language resource management—Morpho-syntactic annotation framework (MAF)
  - [15] ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection—Information security management systems—Requirements
  - [16] ISO/IEC 29151:2017 Information technology—Security techniques—Code of practice for personally identifiable information protection
  - [17] ISO/IEC 27701:2019 Security techniques—Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management—Requirements and guidelines
  - [18] ISO/IEC 27017:2015 Information technology—Security techniques—Code of practice for information security controls based on ISO/IEC 27002 for cloud services
  - [19] ISO/IEC 27018:2019 Information technology—Security techniques—Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors
  - [20] ISO/IEC 23894:2023 Information technology—Artificial intelligence—Guidance on risk management—First Edition
-

上海市生物医药行业协会  
团体标准  
基于大语言模型的药物警戒个案安全性  
报告人工智能辅助信息提取规范  
T/SBIAORG 0002—2025

\*

中国标准出版社出版发行  
北京市朝阳区和平里西街甲2号(100029)

网址 www.spc.net.cn

总编室:(010)68533533 发行中心:(010)51780238

读者服务部:(010)68523946

中国标准出版社秦皇岛印刷厂印刷  
各地新华书店经销

\*

开本 880×1230 1/16 印张 1 字数 18 千字  
2025年7月第1版 2025年7月第1次印刷

\*

书号:155066·5-15562 定价 31.00 元

如有印装差错 由本社发行中心调换  
版权专有 侵权必究  
举报电话:(010)68510107



T/SBIAORG 0002-2025