

TC260

TC260-003

成式人工智 服务安全基本 求

2024-02-29 发布

全国 安全标准化技术委员会发布

次

前



生成式人工智能服务安全标准

1 范围

2 规范性引用文件

3 术语和定义

3.1
生成式人工智能服务 generative artificial intelligence service

3.2
服务提供者 service provider

5 料安全 求

5.1 料来源安全 求

注1:

注2:

注3:

5.2 料内容安全 求

5.3 料标注安全 求

6 模型安全 求

注：

7 安全措施 求

注：

8 其他 求

8.1 关 库

8.2 成内容测 库

8.3 拒 测 库

8.4 分 模 型

9 安全 估 求

9.1 估方法

注1:

注2:

注3:

9.2 料安全 估

9.3 成内容安全 估

9.4 拒 估

录 A

(性)

料及 成内容 主 安全

A. 1 包含 反 会主义核心价值 内容

A. 2 包含歧 性内容

A. 3 商业 法

A. 4 侵犯他人合法权

A. 5 无法满 特定服务 型 安全 求

参 文 献