生成式人工智能（大语言模型）上线备案表（**1.0版**）

备案单位 （盖章）

年 月 日

单位法人 （签字）

年 月 日

安全评估负责人 （签字）

年 月 日

**填 表 须 知**

1.封面和自愿承诺须由相关人员亲笔签名。填报日期具体到日。

2.本表中有“□”栏的为选择栏，请根据实际情况在对应的“□”内划“√”。如勾选了“其他”项，需注明具体内容。

3.若需填写的内容较多，现有的表格中填写不下时，可扩展相应表格填写。有关情况需要进一步说明的，可另附页说明。

4.本表中标“★”号内容须重点关注。

**一**、基本情况

|  |  |  |  |
| --- | --- | --- | --- |
| 模型名称 |  | 单位名称 |  |
| 负责人 |  | 联系电话 |  |
| 单位属地 |  | | |
| 主要功能 | □人机对话 □文字生成 □图片生成 □声音合成 □视频合成 □代码生成或优化 | | |
| 适用人群 |  | | |
| 适用场合 | □自动控制 □医疗信息服务 □心理咨询 □关键信息基础设施 □其他 | | |
| 具体说明： | | |
| 服务范围 | □未限定特定领域  □限定特定领域  具体说明： | | |
| 推出时间 |  | | |
| 是否具有舆论属性或社会动员能力 | | ☑是 □否 | |

说明：（1）单位属地填写单位注册地或总部实际所在地。（2）主要功能一栏，可多选。（3）适用人群须重点说明是否适用未成年人、学生等。（4）推出时间填写开始规模测试或适用的时间。

二、模型研制

**1.模型备案情况**

|  |  |  |  |
| --- | --- | --- | --- |
| 是否采用已备案模型 | □未采用 | | |
| □采用 | 模型名称 |  |
| 备案单位 |  |
| 备案时间 |  |
| 备案号 |  |
| 采用方式 | □商业合作 □开源免费 |

说明：对于使用已备案的模型，填写此表即可；对于在已备案模型基础上优化训练的情况，在表2-5中填写增量情况。

**2.训练算力资源（自研模型）**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 训练服务器数量及GPU厂商 | | |  | |
| 采用第三方云平台或服务器情况 | □未采用 | | | |
| □采用 | 第三方情况 | | |
| 单位名称 | |  |
| 单位性质 | | □外资 □民企 □国企 □事业单位 |
| 平台/服务器地点 | |  |

说明：（1）采用自研模型的，填写表2-5。（2）平台/服务器地点一栏，境外地点说明国家/地区，境内地点详细到省，可填写多个地址。

**3.训练语料和标注语料来源与规模**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 文本训练语料规模 | | 存储规模 | | | | |  | | | | |
| 语料数量 | | | | |  | | | | |
| 各类型语料规模 | | | | | | | | | | | |
| 中文  文本 | 英文  文本 | | | 代码 | 图片 | | | 音频 | | 视频 | 其他 |
|  |  | | |  |  | | |  | |  |  |
| 语料来源 | | | | □开源语料 □自采语料 □商业语料 | | | | | | | |
| 境外开源网站  语料来源★ | | | |  | | | | | | | |
| 境外开源网站语料规模★ | | | | | | | | | | | |
| 中文  文本 | 英文  文本 | | | 代码 | 图片 | | | 音频 | | 视频 | 其他 |
|  |  | | |  |  | | |  | |  |  |
| 自采语料规模 | | | | | | | | | | | |
| 中文  文本 | 英文  文本 | | | 代码 | 图片 | | | 音频 | | 视频 | 其他 |
|  |  | | |  |  | | |  | |  |  |
| 商业语料规模 | | | | | | | | | | | |
| 中文  文本 | 英文  文本 | | | 代码 | 图片 | | | 音频 | | 视频 | 其他 |
|  |  | | |  |  | | |  | |  |  |
| 文本和图片标注数量 | | |  | | | | | | | | |
| 标注人员数量 | | | |  | 标注人员类型 | | | | □内部 □外包 | | |
| 标注人员培训 | | | | 培训 人，每人 小时 | | | | | | | |
| 标注规则★ | | | □有 □无 | | | | | | | | |
| 标注内容准确性人工核验比例 | | | | | |  | | | | | |

说明：（1）文本训练语料规模中，存储规模填写文本格式存储时语料大小，语料数量填写Token数量。（2）图片类型语料规模按张计算。（3）音频、视频语料规模按小时计算。（4）境外开源网站语料来源（如CommonCrawl、Pile、Wikipedia等），须全部列举。（5）文本和图片标注数量按标注单元填（通常按条数、张数）。（6）标注人员培训需说明培训数量及时间。（7）如有标注规则，需另附。

**4.语料合法性★**

|  |  |
| --- | --- |
| 语料来源是否合法 | □是 □否 |
| 语料是否包含侵害他人知识产权内容 | □是 □否 |
| 语料是否包含违法违规的个人信息内容 | □是 □否 |

**5.算法模型**

|  |  |
| --- | --- |
| 模型架构 | □Transformer □其他 |
| 训练框架 | □Pytorch □TensorFlow □PaddlePaddle □MindSpore □其他 |

三、服务与安全防范

**1.推理算力资源**

|  |  |  |  |
| --- | --- | --- | --- |
| 推理服务器数量及GPU厂商 | |  | |
| 并发服务能力 | | □十万级 □百万级 □千万级 □其他 | |
| 采用第三方云平台或服务器情况 | □未采用 | | |
| □采用 | 第三方情况 | |
| 单位名称 |  |
| 单位性质 | □外资 □民企 □国企 □事业单位 |
| 平台/服务器地点 |  |
| 租用期限 |  |

说明：平台/服务器地点一栏，境外地点说明国家/地区，境内地点详细到省，可填写多个地址。

**2.服务**

|  |  |  |
| --- | --- | --- |
| 服务方式 | □APP □网页 □API接口 □其他 | |
| 注册认证方式 |  | |
| ①声明模型适用人群、场合、用途的方式★ |  | |
| ②提供服务过程中收集保存个人信息情况★ | 类 型 |  |
| 数 量 |  |
| 用 途 |  |
| 保存期限 |  |
| ③收集个人信息征得个人同意的  方式★ |  | |
| ④受理处理使用者关于查阅、复制、更正、补充、删除个人信息请求的方式★ | 条 件 |  |
| 途径方法 |  |
| ⑤图片、视频  标识★ | 标识样式 |  |
| 标识位置 |  |
| 频 度 |  |
| ⑥接受公众或使用者投诉举报★ | 途 径 |  |
| 反馈方式 |  |
| ①至⑥是否已经写入模型服务协议 | 已写入：□① □② □③ □④ □⑤ □⑥ | |

说明：（1）服务方式可多选。（2）注册认证方式包括通过手机验证码等方式。（3）个人信息类型包括姓名、身份证号、联系方式、家庭住址等。（4）标识样式一栏，按1:1比例将标识粘贴在此栏。标识位置一栏，说明标识在图片、视频中的位置。频度包括每帧、跳帧等。

**3.非法内容拦截措施**

|  |  |
| --- | --- |
| 监看人员数量 |  |
| 预置关键词拦截 |  |
| 分类模型检测 | □有 □无 |

说明：（1）预置关键词拦截一栏，提供拦截关键词列表。（2）分类模型检测一栏，如有，须参照本《备案表》“模型研制”表格内容，另附分类模型情况。

**4.模型更新、升级**

|  |  |
| --- | --- |
| 在何种情况下（如较频繁发现生成非法内容）重新进行预训练 |  |
| 重新预训练所需时间 |  |
| 计划优化训练（fine-tuning）频度及所需时间 |  |

四、安全评估

**1.基本情况**

|  |  |  |  |
| --- | --- | --- | --- |
| 评估方式 | □自评估 □第三方评估 | | |
| 评估负责人姓名 |  | 联系电话 |  |
| 所在部门或单位 |  | 职 务 |  |
| 国 籍 |  | 常住地点 |  |
| 身份证/护照号码 |  | | |
| 从事安全评估或相近工作时间 |  | | |
| 评估团队专业人员数量 |  | | |
| 评估起止时间 | 年 月 至 年 月 | | |

说明：常住地点应说明所在省/自治区/直辖市。

**2.评估情况★**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **①语料内容评估** | | | | |
| 人工抽检 | 数量 |  | 合格率 |  |
| 关键词抽检 | 数量 |  | 合格率 |  |
| 分类模型抽检 | 数量 |  | 合格率 |  |
| **②生成内容评估** | | | | |
| 人工抽检 | 数量 |  | 合格率 |  |
| 关键词抽检 | 数量 |  | 合格率 |  |
| 分类模型抽检 | 数量 |  | 合格率 |  |
| **③涉知识产权、商业秘密评估** | | | | |
| 方法 | 1. □开展涉知识产权、商业秘密等评估，说明评估方法、判断标准、评估结果等： | | | |
| 结论 |  | | | |
| **④涉民族、信仰、性别等评估** | | | | |
| 方法 | □开展涉民族、信仰、国别、地域、性别、年龄、职业、健康等评估，说明评估方法、判断标准、评估结果等： | | | |
| 结论 |  | | | |
| **⑤涉透明性、准确性、可靠性等评估** | | | | |
| 方法 | □开展涉透明性、准确性、可靠性等评估，说明评估方法、判断标准、评估结果等： | | | |
| 结论 |  | | | |
| **⑥模型性能（拒答率）评估** | | | | |
| 方法 | □开展问题拒答率评估，说明测试题集规模、来源，回答数量、拒答数量，并附测试题集： | | | |
| 结论 |  | | | |

说明：“语料内容评估”和“生成内容评估”重点针对涉意识形态、国家主权，个人隐私、个人肖像权和名誉权等方面的评估。“语料内容评估”须附关键词列表；“生成内容评估”须附测试题集、关键词列表；涉及分类模型的，参照本《备案表》“模型研制”表格内容，另附分类模型情况。

1. 自愿承诺

|  |
| --- |
| 本人代表 （单位名称） 自愿承诺：  1.本备案表中的信息是真实的。如有弄虚作假或不符合实际情况的内容，自愿接受监管部门处罚。  2.本备案表中提到的安全评估是客观公正的，本企业/单位对评估结论负责。  3.本企业/单位依法承担网络信息内容生产者责任，履行网络信息安全和个人信息保护义务。  4.本企业/单位依法配合监管部门的监管。  承诺人（签字）：  年 月 日  注：承诺人应为单位法人 |

**六、附件及备注**

|  |
| --- |
| 1.附件   1. 安全评估报告 2. 模型服务协议 3. 语料标注规则 4. 拦截关键词列表（如较长，附电子版即可） 5. 评估测试题（如较长，附电子版即可） |
| 2.备注  对于训练语料、所用算法、安全防范措施相同，只是参数大小不同的大语言模型，可视为同一模型，无须重复备案。 |