

人工智能科技伦理审查与服务办法（试行）

第一章 总则

第一条 为规范人工智能科技活动伦理治理，促进公平、公正、和谐、安全和负责任创新，推动人工智能产业健康发展，依据《中华人民共和国科学技术进步法》《关于加强科技伦理治理的意见》《科技伦理审查办法（试行）》（以下简称《伦理办法》）等法律法规和相关规定，制定本办法。

第二条 本办法所适用的人工智能科技活动是在中华人民共和国境内开展的，可能在人的尊严、公共秩序、生命健康、生态环境、可持续发展等方面带来科技伦理风险挑战的人工智能科学研究、技术开发等活动，以及依据法律、行政法规和国家有关规定需进行人工智能科技伦理审查的其他科技活动。

第三条 开展人工智能科技活动应将科技伦理要求贯穿全过程，遵循增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明、保护隐私安全、确保可控可信的人工智能科技伦理原则，遵守我国宪法、法律法规和有关规定。

第二章 服务与促进

第四条 建立和完善人工智能科技伦理标准体系，推动

制定相关国际标准、国家标准、行业标准、团体标准，支持搭建国际标准化交流合作平台。

鼓励高等学校、科研机构、医疗卫生机构、企业和科技类社会团体等参与人工智能科技伦理标准的制定、验证与推广。

第五条 推进人工智能科技伦理服务体系建设，强化人工智能科技伦理风险监测预警、检测评估、认证、咨询等服务供给，提升企业技术研发和人工智能科技伦理风险防范能力，加大对中小微企业人工智能科技伦理审查的支持和服务力度，推进人工智能科技伦理国际交流合作。

第六条 鼓励高等学校、科研机构、医疗卫生机构、企业和科技类社会团体等开展人工智能科技伦理审查研究，支持人工智能科技伦理审查技术创新，强化以技术手段防范人工智能科技伦理风险；促进人工智能科技伦理审查高质量数据集有序开源开放，加强通用性风险管理、评估审计工具研发，探索基于应用场景的科技伦理风险评估评测；推广符合科技伦理的人工智能产品和服务，保护科技伦理审查技术知识产权。

第七条 开展人工智能科技伦理宣传教育，发挥科技类社会团体在人工智能科技伦理宣传教育方面的作用，鼓励公众参与，促进实践示范，提升公众伦理意识和素养。引导大众传播媒介有针对性地进行人工智能科技伦理宣传教育。

第八条 支持高等学校、科研机构、医疗卫生机构、企业和科技类社会团体等开展人工智能科技伦理相关教育与培训，推动职业体系和课程体系建设，采取多种方式培养人工智能科技伦理人才，促进人才交流。

第三章 实施主体

第九条 从事人工智能科技活动的高等学校、科研机构、医疗卫生机构、企业等是本单位人工智能科技伦理审查管理的责任主体，应按照《伦理办法》第四条有关要求，设立人工智能科技伦理委员会（以下简称委员会）。委员会应配备必要的工作人员、办公场所和经费等条件，采取有效措施保障委员会独立开展工作。鼓励有资质的相关单位开展人工智能科技伦理管理体系相关认证。

第十条 委员会章程、组成和委员职责、义务应依照《伦理办法》第五条至第八条进行。委员会组成应包括人工智能技术、应用、伦理、法律等相应专业背景的专家。

第十一条 地方、相关主管部门可结合实际情况依托相关单位建立专业性人工智能科技伦理审查与服务中心（以下简称服务中心）。服务中心接受其他单位委托，提供人工智能科技活动伦理审查、复核、培训、咨询等服务。服务中心不得对同一人工智能科技活动同时提供审查和复核服务。服务中心应建立规范的管理制度和程序，配备具有人工智能科技伦理审查与服务能力的专职人员，接受地方或相关主管部

门监督。

第四章 工作程序

第一节 申请与受理

第十二条 对属于本办法第二条所列范围人工智能科技活动的，人工智能科技活动负责人应向本单位委员会申请，单位未设立委员会或委员会无法胜任科技伦理审查工作要求，向本单位委托的服务中心申请开展科技伦理审查，无单位人员应委托符合要求的服务中心开展科技伦理审查。人工智能科技活动负责人应依照规定向委员会或服务中心提交申请材料。申请材料主要包括：

（一）人工智能科技活动方案，包括研究背景、目的和计划，所涉及相关机构的合法资质材料、人员情况、经费来源，拟采用的算法机制机理，数据来源与获取方式、测试评估方法，拟形成的软硬件产品，预期应用领域及适用人群等；

（二）人工智能科技活动科技伦理风险评估情况、防控及应急处理预案，包括人工智能技术预期应用带来的潜在科技伦理风险评估，对科技伦理风险的监测预警措施，对可能产生科技伦理风险的防控计划等；

（三）遵守人工智能科技伦理和科研诚信等要求的承诺书。

第十三条 委员会或服务中心应根据申请材料决定是否受理申请并通知申请人，决定受理的，应根据科技伦理风险

发生的可能性和程度、紧急情况等明确适用的一般、简易或应急程序，根据不同程序要求采取线下、线上等形式开展科技伦理审查，材料不齐全的应一次性完整告知需补充的材料。

第二节 一般程序与简易程序

第十四条 人工智能科技伦理审查会议由委员会主任委员或其指定的副主任委员主持，到会委员应不少于5人，且应包括本办法第十条所列的不同类别的委员。服务中心可参考委员会的规定组织实施工作。

根据审查需要，可邀请相关领域不存在直接利益关系的顾问专家等提供咨询意见。顾问专家不参与会议表决。

第十五条 委员会或服务中心开展人工智能科技伦理审查，重点关注以下方面：

（一）在人类福祉方面，人工智能科技活动是否具有科学价值、社会价值；研究目标对增进人类福祉、实现社会可持续发展等是否具有积极作用；人工智能科技活动的风险是否受益合理。

（二）在公平公正方面，训练数据的选择标准，算法、模型、系统的设计是否合理；是否采取措施防止偏见歧视、算法压榨，保障资源分配、机会获取、决策过程的客观性与包容性。

（三）在可控可信方面，是否能够保障模型、系统的鲁

棒性，以应对开放环境、极端情况和干扰性因素；是否能够保障使用者控制、指导和干预模型、系统的基本操作；是否制定持续监测方案、突发状况处理预案。

（四）在透明可解释方面，是否合理披露算法、模型、系统的用途、运行逻辑、交互方式说明、潜在风险等信息；是否采用有效技术手段提升算法、模型和系统的可解释性。

（五）在责任可追溯方面，是否具有日志管理等措施清楚记录数据、算法、模型、系统各个环节的充分信息，保障全链路可追踪和管理；科技人员资质是否符合相关要求。

（六）在隐私保护方面，数据的收集、存储、加工、使用等处理活动以及研究开发数据新技术等，是否采取充分措施确保隐私数据得到有效保护。

第十六条 委员会或服务中心应在申请受理后的 30 日内作出批准、修改后再审或不予批准等决定。情况复杂或需要补充、更正材料等特殊情况可适当延长并明确延长时限。

对于需要修改或不予批准的，委员会或服务中心应提出修改建议或说明理由。申请人有异议的，应在自决定送达之日起 3 个工作日内向委员会或服务中心提出申诉。申诉理由充分的，委员会或服务中心应在 7 个工作日内重新作出决定。

第十七条 人工智能科技活动负责人应及时识别科技伦理风险变化，并将相关变化情况向委员会或服务中心报告。委员会或服务中心依照《伦理办法》第十九条规定对审查批

准的人工智能科技活动开展跟踪审查，及时掌握科技伦理风险变化情况，必要时可作出暂停或终止相关科技活动等决定。跟踪审查间隔一般不超过 12 个月。

第十八条 多个单位合作开展人工智能科技活动，可根据实际情况开展单位间人工智能科技伦理审查结果互认。

第十九条 有下列情形之一的可以适用简易程序：

（一）人工智能科技活动伦理风险发生的可能性和程度不高于日常生活中遇到的常规风险；

（二）对已批准的人工智能科技活动方案作较小修改且不增大风险受益比；

（三）前期无重大调整的人工智能科技活动的跟踪审查。

第二十条 委员会或服务中心应制定适用简易程序审查的工作规程和跟踪频度。简易程序审查由委员会主任委员指定两名或两名以上的委员承担。服务中心可参考委员会的规定组织实施工作。

简易程序审查过程中出现审查结果为否定性意见、对审查内容有疑义、委员间意见不一致等情形的，应调整适用一般程序。

第三节 专家复核程序

第二十一条 工业和信息化部、科技部会同有关部门制定发布“需要开展科技伦理专家复核的人工智能科技活动清

单”（以下简称复核清单），根据工作需要进行动态调整。

第二十二条 开展纳入复核清单的人工智能科技活动，通过委员会或服务中心的初步审查后，由本单位申请开展专家复核。多个单位参与的，由牵头单位负责申请。中央企业，中央和国家机关直属高等学校、科研机构、医疗卫生机构等直接报请相关主管部门组织开展专家复核。其他单位报请地方开展专家复核。

第二十三条 人工智能科技活动承担单位应依照《伦理办法》第二十七条规定提交专家复核材料。

地方或相关主管部门应依照《伦理办法》第二十八条至第三十条规定组织成立复核专家组，对初步审查意见的合规性、合理性等内容开展复核，在收到复核申请后 30 日内向申请单位反馈复核意见。

地方或相关主管部门可委托服务中心具体实施复核工作。

第二十四条 委员会或服务中心应根据专家复核意见作出科技伦理审查决定。

第二十五条 委员会或服务中心对纳入复核清单的人工智能科技活动加强跟踪审查，跟踪审查间隔一般不超过 6 个月。

科技伦理风险发生重大变化的，应依照《伦理办法》第二十条规定重新开展相关科技伦理审查并申请专家复核。

第二十六条 人工智能科技活动在深度合成、算法推荐、生成式人工智能服务管理等方面实行登记、备案、行政审批等监管措施且将符合科技伦理要求作为审批条件、监管内容的，可不再开展专家复核。

第四节 应急程序

第二十七条 委员会或服务中心应制定人工智能科技伦理应急审查制度，明确突发公共事件等紧急状态下的应急审查流程和标准操作规程。应急审查一般在 72 小时内完成。对于适用专家复核程序的人工智能科技活动，专家复核前的审查一般在 36 小时内完成。

第二十八条 委员会或服务中心应保证科技伦理应急审查的质量与时效，加强跟踪工作和过程监督。必要时，可邀请相关领域顾问专家参会，提供咨询意见。

第五章 监督管理

第二十九条 科技部负责统筹指导全国科技伦理监管工作，工业和信息化部会同有关部门负责人工智能科技伦理治理工作，加强应急伦理审查的协调指导。各部门依照职责权限负责本行业、本系统人工智能科技伦理审查的监督管理工作，各地方依照职责权限负责本地区人工智能科技伦理审查的监督管理工作。

第三十条 单位应依照《伦理办法》第四十三条至第四十五条规定，将委员会相关情况、纳入复核清单的人工智能

科技活动通过国家科技伦理管理信息登记平台进行登记，并提交上一年度委员会工作报告、纳入复核清单的人工智能科技活动实施情况报告等相关材料。服务中心应依照上述规定登记并提交上一年度工作报告。

科技部与相关主管部门对人工智能科技伦理相关登记信息同步共享。

第三十一条 地方、相关主管部门、从事人工智能科技活动的单位应结合本行业、本系统、本单位的实际情况，畅通能够反映人工智能科技伦理违法违规问题的渠道，并按照相关规定处理。

第三十二条 在人工智能科技活动或开展人工智能科技伦理相关工作的过程中，违反本办法规定的，依照《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《中华人民共和国科学技术进步法》等法律法规和相关规定，进行调查处理，给予相应处罚。

第六章 附则

第三十三条 本办法涉及期限的规定，未标注为工作日的，均为自然日。

本办法所称“地方”是指省级人民政府确定的负责人工智能领域科技伦理审查和管理工作的省级管理部门，“相关主管部门”是指国务院相关主管部门。

第三十四条 地方、相关主管部门可依照本办法规定，结合实际情况制定或修订本地方、本行业、本系统的人工智能科技伦理的审查与服务办法、细则等制度规范。科技类社会团体可制定本领域的人工智能科技伦理审查与服务具体规范和指南。

第三十五条 相关主管部门对本行业、本系统人工智能科技伦理审查与服务有特殊规定且符合本办法精神的，从其规定。本办法未作规定的，依照《伦理办法》及相关法律法规的规定执行。

第三十六条 本办法由工业和信息化部会同有关部门负责解释。

第三十七条 本办法自印发之日起施行。

附件：需要开展科技伦理专家复核的人工智能科技活动清单

附件

需要开展科技伦理专家复核的人工智能 科技活动清单

一、对人类主观行为、心理情绪和生命健康等具有较强影响的人机融合系统的研发。

二、具有舆论社会动员能力和社会意识引导能力的算法模型、应用程序及系统的研发。

三、面向存在安全、人身健康风险等场景的具有高度自主能力的自动化决策系统的研发。

本清单将根据工作需要动态调整。