上海市科学技术委员会

沪科提复〔2023〕56 号

对市政协十四届一次会议第 1027 号提案的答复

郑明委员：

您提出的“关于浅谈 ChatGPT 等 AI 引擎高速发展对相关行业的影响与应对的提案”收悉。经研究，现将办理情况答复如下：

您提出的建议，对本市加强 ChatGPT 等 AI 引擎治理相关工作有很重要的借鉴意义。随着 ChatGPT、GPT-4 等大型语言模型为标志的生成式人工智能的迅猛发展，人工智能迎来大模型时代。AI 大模型被认为是革命性的技术进展，将给经济社会发展带

来深远影响，AI 大模型的快速发展应用也持续引发各界对其伦理安全风险的担忧。

目前，我市已积极开展 ChatGPT 等 AI 引擎治理相关工作。**一是积极推进落实国家相关管理规定。**积极落实《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》相关规定，坚持发展和安全并重、促进创新和依法治理相结合的原则，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管。**二是成立上海市科技伦理和科研诚信建设协调机制办公室，其专家委员会中设立人工智能分委会。**2022 年 9 月 22 日，《上海市促进人工智能产业发展条例》经上海市十五届人大常委会第四十四次会议表决通过，明确设立人工智能伦理专家委员会。2022 年 12 月，上海市政府正式建立科技伦理和科研诚信建设协调机制。协调机制设立专家委员会，分设生命科学、医学、人工智能等三个分委员会，同时组建上海市科研诚信宣讲团。其中，人工智能分委会专门负责有关人工智能伦理治理。**三是依托上海人工智能实验室推进大模型的安全评测和价值对齐技术研发，打造夯实我国大模型伦理安全基座。**一方面，打造国家大模型评估评测体系，对照国家监管要求，围绕大模型安全评估等大模型上线前的重点环节，从价值观、合法合规、公平公正、安全无害、数据保护、文明进步等多个维度构建大模型评测顶层框架，建设大模型评测权威语料库与数据集，训练评测专用大模型，建设国家大模型公共评测平台。另一方面，构建我国大模型价值对齐体系，依托大模型训练的权威机构，形成多学科融合的价值对齐人才团队，系统研发人机对齐技

术，攻关数据去毒、中文常识偏差等瓶颈问题，构建我国特色的大模型价值对齐方法与技术路径，不断提升大模型伦理安全能力。**四是支持专家团队持续跟踪研判生成式 AI 的风险。**长期以来，我委鼓励专家学者开展人工智能法规体系、标准体系、监管体系研究，支持专家团队发起国家人工智能治理规则的研究和制定。自 2018 年起，已连续举办六届世界人工智能大会，2023 世界人工智能大会智能社会论坛、法治论坛等多个分论坛将生成式 AI 作为重要议题广泛讨论。在我委支持下，复旦大学成立中国科协—复旦大学科技伦理与人类未来研究院，以及上海人工智能实验室人工智能治理中心，围绕生成式 AI 进行持续跟踪研究。

下一步，我委将持续加强 ChatGPT 等 AI 引擎治理等工作，从评测监管和价值对齐两方面入手构建我国大模型伦理安全基座，及时更新完善 AI 研发领域相关法律法规的制定与落实，积极应对大模型的伦理安全挑战。

感谢您对本市科技创新工作的关系和支持！

上海市科学技术委员会

2023 年 11 月 3 日

抄送：市政府办公厅建议提案处，市政协提案办。

上海市科委办公室　　2023 年 11 月 3 日印发