《人工智能系统生命周期治理指南》地方标准

编制说明

目前人工智能企业、社会、政府对人工智能系统治理的重要性和必要性已达成基本共识。然而，当前人工智能治理的标准侧重于建立人工智能治理的基本原则，而非形成技术指导文件。总体上，目前国际与国内缺乏标准化、可应用落地的人工智能系统生命周期治理的技术性指南。因此，制定人工智能系统生命周期治理指南，形成人工智能系统治理、评估和管控方法，加强前瞻预防与约束引导，有利于实现安全可信、负责任的人工智能应用，最大程度地释放人工智能在社会、经济、环境等各方面的正向价值。

一、任务来源

经上海市市场监管局评审、公示，根据2022年8月26日发布的《上海市市场监督管理局关于下达2022年度第三批上海市地方标准制修订项目计划的通知》（沪市监标技〔2022〕371号），《人工智能系统生命周期治理指南》正式立项。本标准由上海市经济和信息化委员会提出，上海市人工智能标准化技术委员会归口，上海计算机软件技术开发中心、上海市人工智能行业协会、上海依图科技有限公司、上海商汤智能科技有限公司、公安部第三研究所、上海人工智能实验室、上海交通大学、上海人工智能研究院有限公司、华东师范大学等等单位共同起草。

二、编制背景与意义

当前，人工智能已成为新一轮科技革命和产业变革的重要驱动力量，上海已将人工智能定位为重点培育的三大先导产业之一。目前，全市正在加快建设具有全球影响力的人工智能高地，全行业呈现技术越做越精、场景日益做深、价值红利不断释放的趋势。与此同时，人们在享受人工智能技术进步带来的红利和价值的同时，也面临着其引发的伦理和技术风险，因此迫切需要技术性指南来指导人工智能系统生命周期的治理，确保人工智能产业健康发展和对社会的正面影响。2017年7月，国务院印发的《新一代人工智能发展规划》，明确提出在大力发展人工智能的同时，必须高度重视安全风险挑战，确保人工智能安全、可靠、可控发展。2019年6月，国家新一代人工智能治理专业委员会发布《新一代人工智能治理原则——发展负责任的人工智能》，提出公平公正、安全可控、开放协作、敏捷治理等治理原则。

目前人工智能企业、社会、政府对人工智能系统治理的重要性和必要性已达成基本共识。比如ISO/IEC SC42发布了ISO/IEC TR 24368:2022 《Information technology — Artificial intelligence — Overview of ethical and societal concerns》（信息技术 人工智能 伦理和社会关注概述）与 ISO/IEC 38507:2022 《Information technology - Governance of IT - Governance implications of the use of artificial intelligence by organizations》（信息技术 IT治理 组织使用人工智能的治理影响）。然而，上述人工智能治理的标准侧重于建立人工智能治理的基本原则，而非形成技术指导文件。总体上，目前国际与国内缺乏标准化、可应用落地的人工智能系统生命周期治理的技术性指南。因此，制定人工智能系统生命周期治理指南，形成人工智能系统治理、评估和管控方法，加强前瞻预防与约束引导，有利于实现安全可信、负责任的人工智能应用，最大程度地释放人工智能在社会、经济、环境等各方面的正向价值。

本标准参考了ISO/IEC 22989、ISO/IEC TR 24368、ISO/IEC 38507等人工智能国际标准中人工智能生命周期、人工智能治理相关条文，提出了人工智能系统生命周期治理的总则、治理框架和治理流程，提供了人工智能系统生命周期治理的建议，适用于上海市企业内部在研发人工智能系统时治理的实施和自我评价，以及政府监管机构、第三方机构对企业人工智能系统研发、应用过程中治理能力的评价。

三、编制原则

本标准的编写符合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》中的有关规定和要求。

标准中的相关内容广泛参考了ISO/IEC 22989、ISO/IEC TR 24368、ISO/IEC 38507等人工智能国际标准中人工智能生命周期、人工智能治理相关条文，同时参考了全国信标委人工智能分委会发布的《可信赖人工智能标准化白皮书（2022）》、《人工智能伦理治理标准化指南（2023版）》等关于人工智能可信、伦理的技术白皮书，以及《新一代人工智能发展规划》（国发〔2017〕35号）、《新一代人工智能治理原则——发展负责任的人工智能》（国家新一代人工智能治理专业委员会）、《生成式人工智能服务管理暂行办法》等国家政策文件，确保标准内容符合国家相关政策、符合人工智能技术发展的趋势。

四、编制过程

本标准的编制过程主要包括以下几个阶段：

1.标准立项阶段（2022年3月-2022年8月）

2022年3月，上海计算机软件技术开发中心、上海市人工智能行业协会召集相关单位组成标准编制项目组，讨论并确定本项目的重点任务、工作计划和任务分工。项目组系统查阅和梳理了关于人工智能治理的国内外文献以及相关标准，并进行预调研，深入分析了人工智能生命周期治理的技术路线，初步形成《人工智能系统生命周期治理指南》的总体编写思路。2022年4月份起，工作组内部多次召开研讨会，讨论完善框架内容并初步确立了标准要素，并据此起草了标准草案。

2022年4月-6月，项目组完成《上海市制修订地方标准项目建议书》和《人工智能系统生命周期治理指南》标准草案，正式立项并完成立项答辩。经上海市市场监督管理局评审、公示，2022年8月项目获批。

2.草案编写阶段（2022年8月-2023年4月）

标准编写组根据人工智能治理相关文献，逐步完善标准草案。

3.标准启动会（2023年4月）

2023年4月，由标准牵头单位上海计算机软件技术开发中心组织召开了标准项目启动会，会议邀请了上海市经济和信息化委员会、上海市市场监督管理局、上海市人工智能行业协会的领导以及各参编单位相关人员参与。会议回顾了标准立项情况，讨论了标准草案，确定了后续标准编写分工。会上共收集标准修改意见15条，全部采纳。

4.集中编写、完善和讨论阶段（2023年4月-2024年6月）

项目组广泛征集参与单位，邀请上海市人工智能行业协会、上海依图科技有限公司、上海商汤智能科技有限公司、公安部第三研究所、上海人工智能实验室、上海交通大学、上海人工智能研究院有限公司、华东师范大学等单位的专家参与标准的编制和讨论。2023年9月，为充分听取社会各界的意见和建议，项目组面向社会进一步征集参编单位，开展了更广泛的理论研究和实际调研，形成了较为科学的标准框架和主要内容。

项目组按照GB/T 1.1-2020的最新要求，进一步完善了标准草案内容并撰写了编制说明，在2024年4月和5月组织了标准线下讨论修订会，进一步提升了标准内容的科学性和可操作性，形成标准征求意见稿。

5.意见征集阶段（2024年6月至今）

2024年6月~7月，项目组在人工智能标委会内部进行意见征求，收集标准修订意见41条，已全部采纳。

2024年7月，项目组、项目提出单位、人工智能标委会面向相关主管部门、技术机构、高校、社会公众等相关主体，开展线上、线下相结合的公开意见收集。将根据相关意见，进一步完善标准内容。

五、标准主要结构与关键内容说明

本标准提出了人工智能系统生命周期治理总则、治理框架和治理流程，提供了人工智能系统管理者对人工智能系统各生命周期治理的建议，包括初始阶段、设计开发、测试验证、部署、运行监测、重新评估、退出使用等全生命周期的治理措施。本标准适用于指导本市企业内部人工智能系统生命周期治理的实施和自我评价，以及指导本市政府监管机构、第三方评估机构对企业的人工智能治理能力的评价。

**1.** **人工智能系统生命周期治理总则**

明确了人工智能系统生命周期治理的目标和任务。人工智能系统生命周期治理的目标是保障人工智能系统的安全、合规，建立人工智能系统的治理体系，满足市场及利益相关者的期望。人工智能系统生命周期治理的任务包括评估人工智能系统生命周期治理的现状及需求、系统内数据管理和系统运行情况；指导人工智能系统生命周期治理体系的构建和人工智能系统生命周期治理的实施；制定合理的评价体系与审计规范，监督人工智能系统生命周期治理内控、合规和绩效。

**2.** **人工智能系统生命周期治理框架**

提出人工智能系统生命周期治理框架。人工智能系统生命周期治理框架包含人工智能系统在初始阶段、设计开发、测试验证、部署、运行监测、重新评估、退出使用等全生命周期中的治理，治理措施包括决策治理、数据治理、伦理治理、合规治理、风险治理。

**3.人工智能系统生命周期治理流程**

提出人工智能系统生命周期治理流程，包含统筹和规划、构建和运行、监控和评价以及改进和优化等步骤，同时考虑多主体协同敏捷治理。

**4.人工智能系统生命周期治理措施**

本章节详细列举了人工智能系统在各个生命周期的治理措施，包括通用的治理措施、初始阶段的治理措施、设计开发阶段的治理措施、测试验证阶段的治理措施、部署阶段的治理措施、运行监测阶段的治理措施、重新评估阶段的治理措施、退出使用阶段的治理措施。每个阶段的治理措施分为五个方面，分别为决策治理、数据治理、伦理治理、合规治理、风险治理。同时在附录A中，提供了人工智能系统生命周期治理监督检查表，表中列出了人工智能系统各生命周期阶段中治理实施的输出产物。