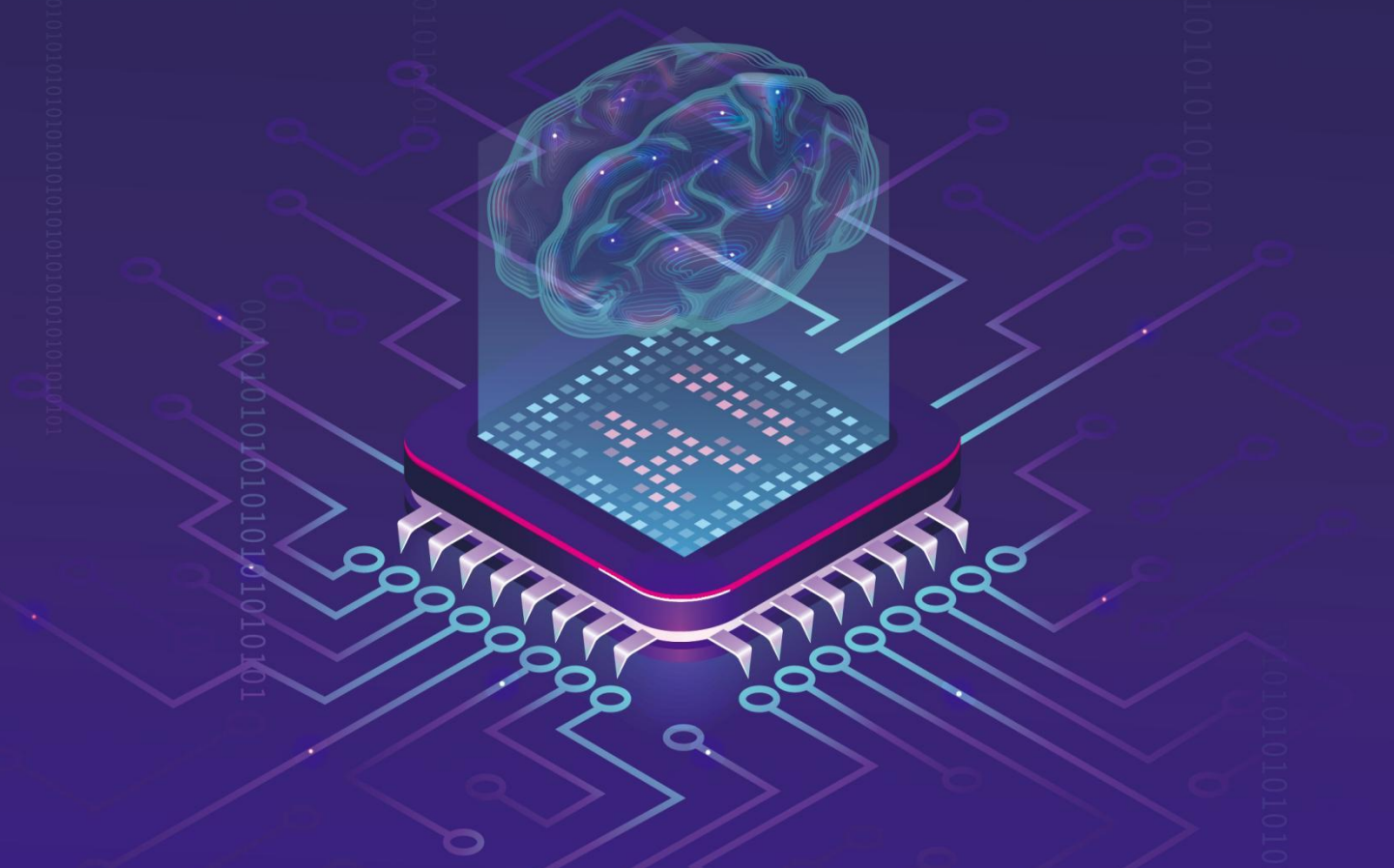
大数据技术标准推进委员会
Big Data Technology and Standard Committee

面向人工智能的数据治理（DG4AI）实践指南1.0



CCSA TC601 大数据技术标准推进委员会

2024年6月

版权声明

编制说明

本报告的撰写得到了数据治理、大数据和人工智能等领域多家企业与专家的支持和帮助，主要参与单位与人员如下（排名不分前后）。

参编单位：大数据技术标准推进委员会、中国联合网络通信集团有限公司、联通数字科技有限公司、中电信人工智能科技(北京）有限公司、中国联合网络通信有限公司软件研究院、中国人民大学、南京大学研究中心、广州信安数据有限公司、星环信息科技（上海）股份有限公司、交通银行股份有限公司、北京神州绿盟科技有限公司、央视频融媒体发展有限公司、亚信科技（中国）有限公司、广州小鹏汽车科技有限公司、北京枫清科技有限公司、华为云技术有限公司、腾讯云计算（北京）有限责任公司、普元信息技术股份有限公司、软通智慧科技有限公司、讯飞智元信息科技有限公司、中电科大数据研究院有限公司、电科云（北京）科技有限公司、上海浦东发展银行、创意信息技术股份有限公司、山东犀盐数据科技有限公司、芜湖明瞳数字健康科技有限公司、深圳市明源云科技有限公司、北京中软国际信息技术有限公司、中国移动紫金（江苏）创新研究院有限公司、杭州比智科技有限公司、云赛智联股份有限公司、湖北数据集团、北京卓信智恒数据科技股份有限公司、海南数造科技有限公司、一汽大众汽车有限公司。

参编人员：尹正，姜春宇，王妙琼，郭彦美，高倩倩，阚鑫禹，李雨霏，刘寒，周京晶，张娇婷，邱梦媛，周圣文，崔一妍，刘思达，

张一鸣，田明慧，马闻达，林木森，王宇龙，艾博焕，高海暘，安小米，蔡洛维，崔博亚，丁乙，何徐麒，胡斌，黄超，李建慧，李金夏，李凯东，屈晓龙，邝苗苗，史赟，谭晟中，王爱书，王瀚，王伟杰，王项男，王潼，闫龙，杨瑞，禹芳，徐松林，夏义堃，张艳红，赵丽丽，鲍立飞，陈韩霏，陈正伟，曹宗伟，崔壤丹，丁洪鑫，代威，方正，高雪峰，高华超，古伟，顾正嘉，龚禧，龚昱鸣，郭文鹏，花福军，黄启洲，胡文涛，姜丹丹，姜鹏，江龙兵，金依扬，刘頔，刘意凡，刘晨璐，刘庆会，刘燕，李光耀，李阳，李铁峰，李晓娟，卢科，梅珂夫，彭建辉，彭涛，钱龙，石荣达，万如意，王远，谢亚南，谢锋，肖美虹，徐超，徐聪颖，薛高飞，杨博，杨明皓，余震宇，袁雪梅，苑国跃，曾伟雄，曾云，张芬，张广庆，张玥玥，张可雨，张毓，张春雷，张文翔，周正斌，周小敏，周海涛，周维，周映，庄颂。

特别感谢以下专家对报告编制给予的专业性指导：安小米，蔡洛维，崔博亚，丁乙，何徐麒，胡斌，黄超，李建慧，李金夏，李凯东，屈晓龙，邝苗苗，史赟，谭晟中，王爱书，王瀚，王伟杰，王项男，王潼，闫龙，杨瑞，禹芳，徐松林，夏义堃，张艳红，赵丽丽。

引言

自 1988 年由麻省理工学院的学者启动了全面数据质量管理计划（TDQM）以来，随着大数据技术的迅猛发展，企业内数据量急剧上升，数据治理的内涵也在不断地变化和丰富。2021 年，随着以大模型为代表的生成式人工智能技术席卷全球，对人类的生产和生活都带来了革命性的变化，人工智能的发展从以模型为中心转变为了以数据为中心。以数据为中心的人工智能理论认为，好的人工智能需要高质量、大规模和多样性的数据。但在实践过程中，数据科学家们往往会遇到数据安全与隐私泄露、内容输出偏见与歧视以及数据“高量低质”的问题。如果放任这些问题不加管制，将会阻碍人工智能技术的进一步发展，甚至会危害个人、企业甚至国家的安全。

为了应对这些挑战，开发出更负责任、更可控的人工智能应用，面向人工智能的数据治理（DG4AI，Data Governance for Artificial Intelligence）概念应运而生。

当前，DG4AI 的需求极其迫切，其研究与实践还处于起步阶段，概念和实践方法论尚未形成。为凝聚共识、开宗明义，大数据技术标准推进委员会（CCSA TC601）组织大型银行、通信运营商、头部互联网公司共同编写《面向人工智能的数据治理（DG4AI）实践指南（1.0）》，旨在推动 DG4AI 理念的广泛应用。本指南第一章从数据治理的发展、面向人工智能的数据治理定义、治理主要阶段以及价值等明确人工智能数据治理的概念。第二章从治理的方法和技术对

DG4AI 的重点工作进行说明。第三章提出了一种 DG4AI 的数据治理步骤，为业界抛砖引玉，提供参考。最后在第四章提出了展望。在附录中我们以美国为主要研究对象，对比了中美在 DG4AI 在国家战略、法律类法规以及标准建设上的现状。

本指南在细节和深度上仍有较大提升空间，希望业界更多的专家能够不吝赐教，提出宝贵的修改意见。工作组将持续不断地完善这一指南，对我国 DG4AI 这一研究领域尽绵薄之力。

联 系 人：尹正

联系电话：15810811776

联系邮箱：yinzheng@caict.ac.cn

一、 人工智能数据治理概念界定

（一） 数据治理的发展

1. 数据治理的发展

数据治理的概念起源于企业管理领域，关于数据治理的定义研究众多，但由于业界权威研究机构、研究学者以及国内外标准组织研究视角不同，尚未形成较为统一的认知。

国际数据治理研究所（DGI）提出数据治理的定义为“一个根据既定模型针对信息相关过程的决策权和职责分配体系”。

梅宏院士在《数据治理之论》一书中提出数据治理的核心内容包括以释放数据价值为目标、以数据资产地位确立为基础、以数据管理体制为核心、以数据共享开放利用为重点、以数据安全与隐私保护为底线。

国际数据管理协会（DAMA）提出的数据治理概念为“在管理数据资产过程中行使权力和管控活动，包括计划、监控和实施。”

此外，在国际标准中，最早出现的数据治理（data governance）术语定义源自 ISO/TR 14872:2019 Health informatics — Identification of medicinal products — Core principles for maintenance of identifiers and terms，将其定义为“以管理信息的质量、一致性、可用性、安全性和可用性为重点的过程”，并强调该过程与数据所有权和管理的概念密切。从上述定义的内容来看主要对信息质量的管理，而后有国际标准从 IT 治理、数据资产管理等视角定义数据治理，逐渐凸显数据治

理应具有统筹协调、权责分配、资源调度等核心能力，涉及数据质量、数据安全、数据合规等关键治理内容。

在国家标准中，最早出现的数据治理术语定义源自 GB/T 35295-2017 《信息技术 大数据 术语》，将其定义为“对数据进行处置、格式化和规范化的过程。”，从定义内容来看主要是从数据管理视角来理解数据治理，而后有国家标准从数据管理权利、管控活动等视角定义数据治理，逐渐凸显数据治理具有过程性、集合性以及统筹与协调管控的特征。

结合通用场景下数据治理定义的特征来看，数据治理的核心治理内容主要围绕数据质量、数据安全、数据合规等内容展开，强调要围绕治理内容进行统筹协调、权责分配、资源调度等。

2. 数据治理的三个阶段

第一阶段，20 世纪 80 年代，随着数据库技术的发展，企业开始意识到数据的重要性。但当时数据管理主要依靠数据库管理系统（DBMS），直到 1988 年由麻省理工学院的两位教授启动了全面数据质量管理计划（TDQM），可以认为是数据治理最初的雏形。

第二阶段，伴随着数据仓库的建设，主数据管理与商务智能平台的实施，国内也逐步开始接受并利用数据治理的概念进行推广实践。

第三阶段，21 世纪 20 年代，以大模型为代表的生成式模型成为推动人工智能发展的重要驱动力。大模型的兴起对数据治理提出了新的挑战和需求。

第一阶段-DBMS **01**

数据治理主要依赖于数据库管理系统

- **1988 年**，麻省理工启动全面数据质量管理计划（TDQM），DAMA（国际数据管理组织协会）同年成立。
- **2002 年**，学术论文《数据仓库治理》中开始出现“数据治理”的实践内容。
- DGI、DAMABOK等数据治理框架开始发展。

第二阶段-EDW **02**

企业级数仓及大数据平台建设加速数据治理发展

- **2017年**，中国信通院《数据资产管理实践白皮书（1.0）》发布。
- **2018 年**，国家标准《数据管理能力成熟度评估模型》发布。
- **2023年，**《数据资产管理实践白皮书》更新至6.0，《DataOps实践白皮书（1.0）》发布

第三阶段-AGI **03**

通用人工智能的快速发展为数据治理带来新的挑战

- **2020年**，GPT3，生成式人工智能的成功席卷全球，对整个社会的生产模式产生影响。
- **2021年**，吴恩达提出人工智能将从“以模型为中心”转向“以数据为中心”，对高质量数据集提出要求。

3. 大模型时代数据治理的难题

随着人工智能技术的飞速发展，大模型已成为推动 AI 应用创新的重要驱动力。这些模型依赖于海量的数据、强大的算力以及复杂的算法参数来支撑其庞大的智能体系。在这一过程中，数据可谓是大模型的“灵魂”，塑造了其独特的“个性”。

大模型的智能程度与“个性”表现，促使人类社会生产力迈上新的台阶，同时也带来了更大的挑战与危险。

1） 数据“高量低质”

数据是人工智能技术的基石，是大模型训练和推断的原材料已成为共识。然而，数据的数量和质量并不总是成正比。在来源上，模型往往依赖于从互联网、社交媒体和公开数据库中采集的数据进行训练，这些数据的来源和质量无法得到有效控制。在管理上，我们面对多模态、非结构化数据缺乏理论与技术的支撑来客观评价数据质量的高低。

这些问题需要数据治理来解决，但传统的数据治理理论与实践更

多的适配于面向 BI 时代的结构化数据，在人工智能所需要的非结构化、半结构化、多模态数据上较为空白。为了应对这些挑战，开发出更负责任、更可控的人工智能应用，面向人工智能的数据治理（DG4AI，Data Governance for Artificial Intelligence）概念应运而生，它旨在通过创新的数据管理策略和技术，解决 AI 发展中的痛点问题。

2） 安全与隐私泄露频发

随着大模型对数据的依赖性日益增强，数据安全和隐私保护已成为核心问题。在大模型的全链路研发、管理和应用过程中，其各个阶段都存在着数据安全与隐私的问题和风险，包括但不限于数据的过度采集、样本的偏差、数据的投毒等情况，存在危害个人、企业甚至社会的安全与利益的巨大风险。

3） 偏见与歧视随处可见

在科技飞速发展的背景下，人工智能伦理和道德的关注程度及应对措施尚未完全跟上技术的步伐。自然语言处理技术的滥用案例日益增多。其中包括压制不同意见、侵犯隐私与匿名性等。随着人工智能技术驱动的应用逐步走向产业化，潜在的道德伦理问题成为备受关注的焦点。

这些伦理问题可能源于系统意外产生，也可能是恶意行为者蓄意开发。常见的负面后果包括因人口统计偏见而导致的不公平问题、面向不同用户群体的服务性能不平等、对话者需求的错误识别，以及有害内容与刻板印象的传播等。此外，许多应用只注重信息内容，对文本作者及其信息的社会意义缺乏足够的意识与关注。

DG4AI 的提出，是对现有数据治理体系的重要补充。它强调了数据治理在 AI 研发全生命周期中的重要性，并指出了实现高质量 AI 应用的关键路径。这一概念的实践，需要跨学科的合作、政策的支持以及技术的创新，以确保 AI 技术的健康发展，并最大化其对人类社会的积极影响。

（二） 面向人工智能的数据治理的定义

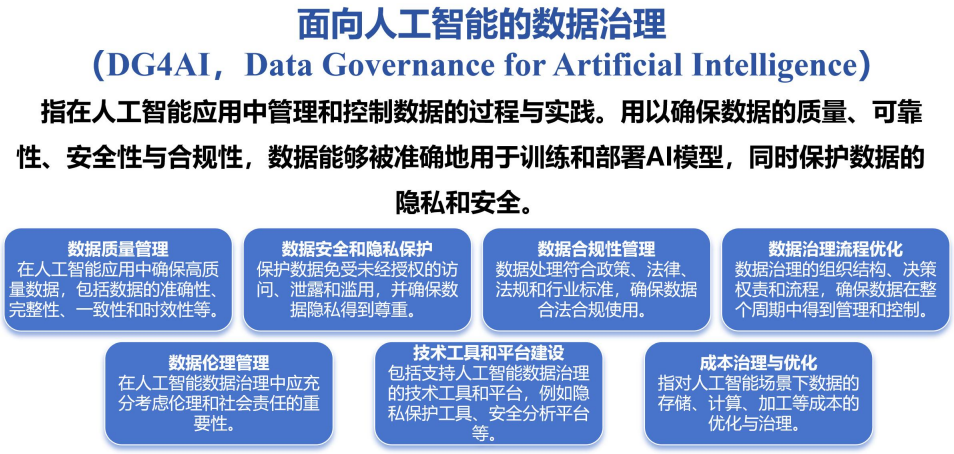
对于面向人工智能的数据治理定义建立于对人工智能和数据治理的共同理解之上，因此在这里我们先对几个关键概念进行明确：

人工智能：人工智能（AI，Artificial Intelligence）是一个与认知科学/心理学、哲学、语言学和数学等学科进行了知识融合的计算机科学，当前由于对于智能的定义存在困难，在学界并未有一个统一的定义，但是从商业的角度来看，AI 意味着使计算机能够执行各种高级功能（包括查看、理解和翻译口语和书面语言、分析数据、提出建议等能力），达到帮助替代或超越人类的工作的能力。

数据治理：根据 IBM、标准要求、数据治理协会等研究（见附录二）综合来看，数据治理主要是从组织层面对数据进行管理。其目的在于确保数据的质量安全性。可指代为企业数据价值化开展的一系列具体性工作，也可指代一系列数据管理活动的集合。

由此，我们可以认为面向人工智能的数据治理（DG4AI，Data Governance for Artificial Intelligence）是指在人工智能应用中管理和控制数据的过程与实践，用以确保数据的质量、可靠性、安全性与合规

性，数据能够被准确地用于训练和部署 AI 模型，同时保护数据的隐私和安全。



（三） 面向人工智能开展数据治理的主要阶段与对象

从组织层面开展数据治理工作方面来看，面向人工智能开展数据治理的工作，主要包含以下四个阶段：

1. 顶层设计阶段：
    - **治理目标：**确立数据治理的总体框架和战略目标，确保数据治理与组织的整体战略相匹配。
    - **工作重点：**根据组织的业务现状、信息化现状、数据现状和 AI 现状，设定组织中各机构和部门的职责、权力的利益，定义符合组织战略目标的整体数据治理目标和可行的行动路径。
2. 数据治理组织保障体系搭建阶段：
    - **治理目标：**确保面向 AI 的数据治理得到必要的支持和资源，包括人力、算力、算法、数据、技术和管理等支持。

- **工作重点：**分析领导层、管理层、执行层等利益相关方的需求，建立健全面向 AI 数据治理的相关管理制度和标准，并基于数据治理所需的专项能力和业务价值目标构建支持面向 AI 的数据治理体系。

3. 数据治理工程建设阶段：
   - **治理目标：**基于数据战略目标，结合 AI 数据治理的特点，制定并执行数据治理实施计划，确保数据治理能够按照既定目标和流程进行。
   - **工作重点：**包括数据收集、数据预处理/清洗、特征工程、数据标注、数据划分、数据增强、模型训练、模型验证与测试、模型推理等实施步骤。
4. 数据治理运营优化与 AI 应用融合阶段
   - **治理目标：**提升 AI 应用的规模化落地效果，实现数据治理与 AI 应用的良性互动。进一步，形成数据治理与 AI 应用相互促进的闭环，实现数据价值的最大化。
   - **工作重点：**通过数据治理提升 AI 模型的拟合效果，同时利用 AI 技术优化数据治理流程，形成良性闭环系统。

从面向人工智能场景下所需的数据来看，大致可分为原始的多模态数据集、训练数据集、验证数据集、测试数据集和推理数据集。

从面向人工智能场景的工程建设阶段来看，可分为数据收集、数据预处理/清洗、特征工程、数据标注、数据划分、数据增强、模型

训练、模型验证与测试、模型推理等九个阶段。

对于人工智能训练和推理的阶段与数据治理对应关系如下：

1) **数据收集阶段：**此阶段治理对象包括：结构化数据、非结构化数据、半结构化数据、空间地理数据、时间序列数据等多种模态数据集。数据来源的选择和收集策略直接影响后续的数据质量。数据的有效性和代表性在此阶段就已开始形成。此阶段我们需要保障相关来源的数据量和多样性。
2) **数据预处理/清洗阶段：**此阶段数据治理对象是数据收集阶段所采集的多模态数据。此阶段对收集到的数据进行初步处理，去除无关信息，修正错误数据，处理缺失值、异常值、重复值等问题，确保数据质量。数据必须具备高度的质量和准确性，保证训练模型时使用的样本数据能够反映真实世界的情况。
3) **特征工程阶段：**此阶段治理对象包括：原始数据集，中间数据和特征变量、标签数据集等。此阶段将原始数据转化为适合机器学习算法使用的特征表示，包括特征提取、特征选择、特征构造等。对于非结构化数据，可能需要进行特征提取，如文本分词、图像特征提取等。特征的选择、构造与转换过程决定了模型能否有效

捕捉到数据中的有用信息，特征的质量直接影响模型的表现力和泛化能力。

4) **数据标注阶段**：此阶段治理对象主要是标注数据集。对于监督学习任务，需要人工或半自动方式对数据进行标注。高质量的标注数据对于模型的学习至关重要。准确、一致且全面的标注能显著提升模型训练效果。

5) **数据划分阶段**：此阶段治理对象主要是训练集、验证集和测试集三类。本阶段将数据集划分为训练集、验证集和测试集，训练集用于训练模型。数据划分阶段的质量治理重点在于保障数据分布和数据平衡，合理地将数据划分为训练集、验证集和测试集，确保每个集合都能代表总体数据分布，有助于避免过拟合或欠拟合。

6) **数据增强阶段**：此阶段治理对象主要是合成数据。为了提高模型的泛化能力和应对不平衡数据问题，合成数据是通过模拟或生成技术生成的人工数据，用于模型训练、隐私保护等目的。虽然对合成数据的质量治理不是直接改善原始数据质量，但能间接提高模型对各种情况的适应性和泛化能力。合成数据质量治理包括对合成数据的生成过程、使用限制等方面进行规范和管理。

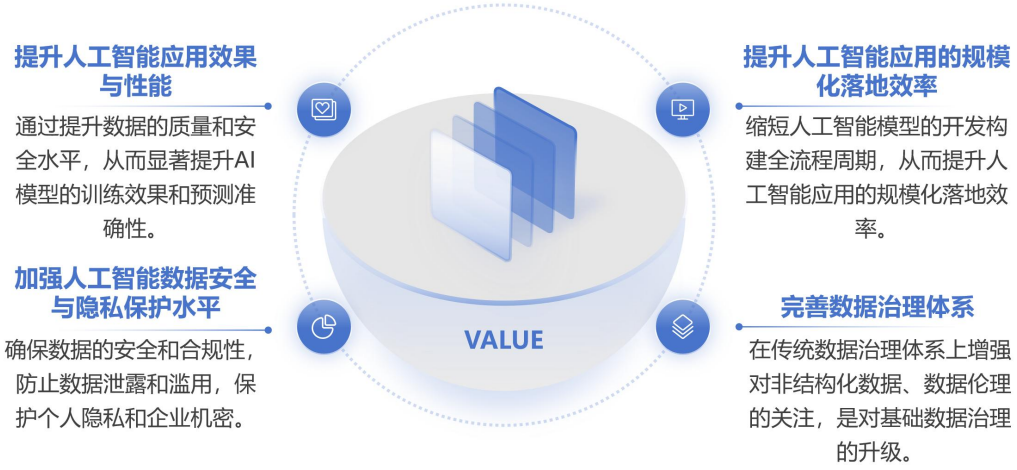
7) **模型训练阶段**：此阶段治理对象主要是训练数据。使用高质量的数据训练模型，会得到更准确、稳定的结果。训练过程中，如果数据质量不佳，模型容易学得有偏差或者过拟合。训练数据的数据质量治理重点保障数据的完整性、准确性、一致性、多样性和代表性。

8）**模型验证与测试阶段：**此阶段治理对象主要是验证数据和测试数据，包括对抗性样本、稀有事件或者小样本数据等。模型的性能验证和测试依赖于独立的高质量测试集，只有当测试数据具有良好的代表性时，才能准确评估模型在新样本上的真实性能。在模型验证与测试阶段，对数据的要求和活动更加聚焦于检验模型在未知数据上的表现和鲁棒性，确保模型不仅在训练集上表现出色，而且在新的、未见过的数据上也能维持良好的性能。

9) **模型推理阶段：**此阶段治理对象主要是推理数据集。在模型推理阶段，除了确保模型自身的性能以外，还要关注用于推理的实时数据的质量，通过一系列的数据处理活动来保证模型在实际应用中的效果和稳定性。推理数据集质量治理的关注点主要包括推理数据集的数据格式兼容性、数据质量监控、数据有效性验证、实时数据更新与维护、在线特征提取与转换等。

（四） 面向人工智能的数据治理价值

通过在面向人工智能场景下对数据集与数据工程流程进行系统化、标准化的治理，一是能够提高人工智能模型的准确性和可靠性。二是能够缩短人工智能模型的开发周期，降低开发与维护成本。三是能够提升整个 AI 系统的安全水平。此外，这项工作还能够完善对未来全域数据治理理论版图的构建。



（五） 面向人工智能的数据治理原则

在开展面向人工智能的数据治理工作时，建议遵循以下原则，从而更好的提升人工智能应用效果、保护个人隐私与社会安全、降低成本、消除歧视与偏见。

- **标准化原则：**人工智能数据治理应具有灵活性、可操作性和实用性，能够被实际应用到数据治理中，通过制定和使用统一的标准、规范和流程并不断迭代完善。这有助于降低数据管理成本，提高数据质量和效率。
- **透明性原则：**人工智能数据治理应该是透明的，运作方式和数据处理过程应该是可解释和可理解的，有助于建立信任和可靠性，并确保合规性。
- **合规性原则：**人工智能数据治理应该符合相关法律法规和行业标准的要求，如隐私法、知识产权法等。
- **安全性原则：**人工智能数据治理应注重数据的安全性，采取适当的安全措施，如加密、访问控制等，以保护敏感数据的机密性和

完整性。

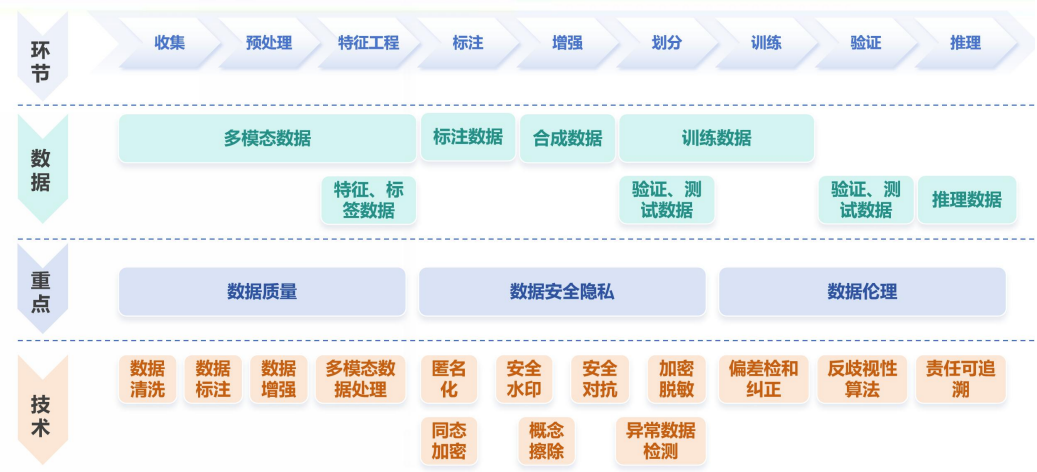
- **负责任原则**：人工智能数据治理应该遵循道德和伦理标准，保证对数据和个人隐私的尊重，避免歧视和不公平的结果。
- **公正性原则**：人工智能数据治理应该确保公正和平等的对待所有用户和利益相关方，不偏袒特定群体或个人。
- **可审计原则**：人工智能数据治理应建立审计机制，对数据的收集、存储、处理和使用等过程进行监控和记录，以便及时发现和解决潜在问题。

上述原则为人工智能数据治理提供了指导方针，确保 AI 系统的可靠性和公正性，保护个人隐私和数据安全，促进组织的数字化转型和发展。

二、 面向人工智能数据治理的重点工作

数据治理理论已有多年的发展，相关治理逐步趋于完善和成熟。然而，针对人工智能领域的数据治理，由于治理对象与评价方式的特殊性，各方实践仍处于初步探索阶段。本章节我们将从数据质量治理、数据安全与隐私治理、数据伦理治理三方面展开探讨，为人工智能场景下的数据治理工作提供指导。这三大重点工作与传统的数据治理理论过程类似（如 PDCA 循环、安全分级分类等），但鉴于人工智能应用场景的特殊性，其在技术和实施细节上有所差异。

随着技术与产业实践的不断演进，数据治理的重点工作将持续补充和完善，以更好地适应人工智能领域日新月异的需求和挑战。



（一） 数据质量治理

推进面向人工智能的数据质量治理是保障人工智能应用高质量、准确、持续可用的重要基础，这要求企业找准人工智能应用建设过程中的数据质量治理需求，把握人工智能数据质量治理与传统数据质量治理的差异，精准识别人工智能数据质量治理的范围和对象，从而帮

助企业科学设计人工智能数据质量治理的机制、方法和步骤，建设面向人工智能的全面的数据质量治理技术能力，以良好的数据质量治理实践预防和杜绝数据质量问题，提高人工智能应用的性能和效果。

1. 治理方法

面向人工智能的数据质量治理是确保数据质量对人工智能模型效果的重要保障。数据质量管理贯穿于人工智能研发、管理和应用的整个生命周期，需要建立一套完善的数据质量管理体系，制定相应的数据质量管理制度和流程，并明确各环节的责任和要求。

同时，还需要建立有效的数据质量评估和监控机制，以确保数据质量符合要求，为人工智能模型的训练、调优提供高质量、高可信度的数据资源，从而提升模型表现效果。

1） 需求分析与质量目标设定

在面向人工智能的数据治理过程中，数据质量治理始于需求分析与质量目标设定阶段。明确人工智能应用对于数据质量和数量的具体需求，设立针对性的数据质量基准和目标。通过这一系列的数据需求分析与质量目标设定工作，可以为后续的数据清洗、预处理、特征工程、标注和增强等数据质量管理活动提供清晰的方向和依据，确保整个数据治理体系围绕既定目标有序展开，最终提升人工智能模型的性能和稳定性。

2） 制定数据质量管理体系

在面向人工智能的数据治理过程中，制定数据质量标准和规范是极为关键的一环，其主要任务是建立一个全面、严谨的数据质量评价指标体系。通过这样的数据质量标准和规范体系，能够对数据治理过程中的数据进行全方位、多层次的质量评估，指导数据清洗、预处理、标注等活动的开展，进而保障人工智能应用所用数据的质量，推动模型训练和应用效果的提升。

3） 数据源评估及采集

从数据的源头控制好数据质量，让数据“规范化输入、标准化输出”是解决人工智能数据质量问题的关键所在。同时不仅关注数据来源的可靠性，而且考虑数据是否涵盖足够的维度和场景，确保数据来源、质量和多样性。

4） 数据预处理

数据标注与数据增强是数据质量管理活动中不可或缺的部分，它们能够在有限的数据资源基础上，通过创新技术和策略，生成大量高质量的训练数据，有力支撑人工智能模型的高效训练和准确预测。

a. **数据清洗：**此阶段聚焦于消除数据中的错误、不完整、不一致和重复等问题。具体措施包括：

- **缺失值处理：**通过填充（如使用平均值、中位数、众数等）、插值或其他方法处理缺失值。
- **异常值检测与处理：**识别并移除或者替换那些明显偏离正

常范围的数据点，防止其对模型训练产生不良影响。

- **数据一致性校验**：对同一实体在不同数据源中的记录进行比对和整合，保证数据的一致性。
- **去重处理**：识别并移除非唯一标识的重复数据记录，避免因重复样本导致的模型训练偏差。

b. 数据标注：在某些 AI 任务中，尤其是监督学习场景下，模型需要依赖带有标签的高质量数据进行训练。高质量的数据标注能有效提高模型理解和学习数据的能力，为后续模型训练奠定基础。

c. 数据增强：即使经过精心标注，实际可用的数据量也可能受限，这可能会导致模型过拟合等问题。数据增强技术旨在通过一系列规则或算法人为地扩增训练数据，如在图像识别任务中采用翻转、旋转、裁剪、色彩变换等方式生成新的训练样本；在文本数据上，可通过同义词替换、句式变换等方式生成不同的表达形式。数据增强不仅可以有效扩大训练样本空间，还可以提高模型的泛化能力和鲁棒性，降低过拟合的风险。

5） 特征工程

对已预处理过后的数据，可能仍然存在不满足人工智能需求，其目的是通过对已清洗和预处理过的数据进行深层次的分析和转换，提取、构造出最具价值的特征，以满足后续人工智能模型构建的需求。通过特征工程的实施，可以极大提升数据对于人工智能模型的解释能力和预测能力，进而增强模型的泛化能力和实用性。

6） 数据偏见检测与矫正

在数据治理的过程中，深入挖掘并矫正数据偏见是构建负责任且公平的人工智能系统的基石，有助于避免 AI 应用在社会生活中可能造成的歧视和不公。

a. 数据偏见检测： 数据偏见检测主要通过统计分析、可视化手段以及特定的偏见评估框架来发现潜在的不公平现象。例如，在分类任务中检查不同群体的误分类率是否存在显著差异，在回归任务中审视因变量预测值是否受无关属性（如性别、种族）的影响过大等。

b. 偏见矫正： 在发现数据存在偏见后，数据质量管理活动会采取一系列措施来矫正这些偏见，从而提升模型决策的公平性。

7） 常态化数据质量监控

在面向人工智能的数据治理过程中，数据质量管理活动的一个核心环节是常态化数据质量监控，该环节贯穿于整个 AI 项目生命周期，既包括模型训练阶段，也涵盖模型推理阶段。常态化的数据质量监控不仅是数据治理的重要组成部分，也是确保人工智能应用成功运行、产出高质量成果的必要条件。无论是训练还是推理阶段，都需要对数据质量进行严格的把关，以应对不断变化的数据环境，持续优化数据质量和模型效果。

a. 模型训练阶段的数据质量监控： 在模型训练前，系统应具备实时数据质量监测功能，对输入的训练数据持续进行完整性、一致性、精确性、及时性等方面的监控。一旦发现数据异常或质量问题，如数

据分布突然变化、出现大量缺失值、新增数据格式不合规等情况，应及时触发告警机制，以便快速定位问题源头并采取相应措施进行修复或清洗。只有确保数据质量达到预期标准，才能将其用于模型训练和微调，以期得到精准、可靠且具有泛化能力的 AI 模型。

b. 模型推理阶段的数据质量监控： 即使模型进入推理阶段，数据质量监控工作依然不能松懈。对于模型接收的实时或批量推理请求所携带的数据，也需要执行严格的数据质量检查，确保数据格式正确、内容合法、业务逻辑合理，避免无效数据、恶意攻击数据或不符合模型输入规范的数据对模型运行造成干扰或损害。通过实时数据质量监控，能够迅速发现并拦截有问题的数据输入，保障模型推理的稳定性和准确性。

8） 持续改进

在面向人工智能的数据治理过程中，数据质量管理活动的持续改进是一个动态迭代、持续优化的过程。不断地总结经验教训，优化数据质量管理的方法和流程，实现数据质量的螺旋式上升，从而为人工智能系统提供更为精准、可靠的数据支撑，推动 AI 技术的健康发展。该过程可能包括：定期复盘过去的数据质量管理活动，总结成效、发现问题，为改进提供方向；提炼经验教训，制度化有效策略并规避同类问题；紧跟 AI 技术发展，优化数据采集、预处理、清洗、标注等环节，提升管理方法和流程的科学性与高效性；密切关注并适时引入新的数据处理与质量管理工具，以应对复杂挑战，提高数据质量，促

进 AI 模型训练与应用效果。

2. 治理技术

在面向人工智能的数据治理过程中，数据质量治理涵盖了从数据准备到模型训练、再到模型部署应用的全过程，每个阶段都有相应的数据治理技术来确保数据质量。借助一系列先进的治理技术，实现了从数据获取到模型应用全程的数据质量问题管理和控制，为 AI 模型的成功构建和稳定运行提供了强有力的支持。

数据清洗：数据清洗是数据预处理中的一项基本任务，旨在去除数据中的噪声、异常值和重复信息。随着机器学习和人工智能的发展，数据清洗技术也在不断进步。现在，通过使用自动化工具和机器学习算法，可以更高效地识别和纠正数据中的错误和不一致性，提高数据的质量和可靠性。

自动化与智能标注：传统的数据标注方法通常需要大量的人工劳动，成本高昂且耗时。为了解决这个问题，自动化标注技术应运而生。自动化标注使用机器学习算法和计算机视觉技术自动识别和标注数据，大大提高了标注效率并降低了成本。对于某些复杂的数据类型，完全自动化的标注可能无法达到高精度。在这种情况下，交互式标注和可编辑的标注成为一种有效的解决方案。这些方法允许用户对自动标注的结果进行手动编辑和调整，以提高标注的准确性和精度。

特征工程：特征工程是将原始数据转换为特征向量，供机器学习算法使用的技术。随着机器学习和人工智能技术的发展，特征工程技

术正朝着自动化和智能化方向发展。通过使用机器学习算法和自然语言处理技术，可以自动识别和转换数据中的特征，提高数据处理的效率和准确性。

数据增强：数据增强是通过生成新的训练样本扩展数据集的技术。在深度学习和计算机视觉领域中，数据增强扮演着重要的角色。通过使用旋转、翻转、裁剪等技术，可以增加数据的多样性和丰富性，从而提高模型的泛化能力。

多模态数据处理和分析：随着多模态数据的普及，数据质量特征工程技术正朝着多模态数据处理和分析方向发展。多模态数据包括文本、图像、音频和视频等多种形式，每种形式都有自己的特征和属性。通过多模态数据处理和分析技术，可以综合利用不同模态的数据，提高数据分析和预测的准确性和可靠性。

（二） 数据安全与隐私治理

推进面向人工智能的数据安全与隐私治理是保障人工智能被安全、可靠使用的基础。在训练算法模型的过程中，会利用到企业和个人的身份、隐私和交易等数据，通过对人工智能应用全生命周期的数据安全与隐私数据进行治理能够有效的保护个人隐私、防止数据泄露并且避免算法被数据投毒所侵害，研发可以被放心使用的人工智能应用。

1. 治理方法

1） 建立数据全生命周期安全监督机制

为确保数据全生命周期的安全性，需建立包含数据采集、处理、存储及输出的全面安全保障机制。这涉及确保数据集多样性与公平性、实施数据处理与加密措施、制定严格的数据使用规定等。在数据治理中，透明度与知情权的保障至关重要，需明确告知用户数据收集目的，并允许用户自主决定是否共享个人信息，增强隐私控制。遵循数据最小化原则，仅收集必要的个人数据，并定期审查清理，以降低隐私风险，确保数据安全。

2） 制定数据集安全风险分类管理体系

为应对人工智能领域的安全挑战，需构建一个基于应用场景、影响范围和风险的分类分级管理体系。该体系应对高风险领域进行定期的数据安全能力评估，并根据风险级别采取差异化的管理策略，实现精细化管理。

首先，明确隐私数据的定义和范围，对数据进行等级分类，指导数据使用模块对不同安全等级的数据进行模糊化处理，以降低泄露风险。数据安全管理应贯穿整个数据治理生命周期，通过分类分级为信息安全管理提供指导，帮助制定安全策略和保护措施，确保数据治理全面合规。数据分级还涉及对数据破坏后果的预估和公众危害程度的分析，确保各级数据得到适当保护。

其次，建立数据安全管理系统，支持数据分类的增删、搜索和敏感词管理，优化资源分配与共享，增强数据安全意识，引导用户主动

保护数据安全。

通过这些措施，可以有效地评估和管理数据安全风险，确保数据的安全性和合规性。

3） 数据加密

应用先进的加密算法对数据进行加密，确保数据在存储和传输过程中的安全。其次，构建全面的安全管理体系，涵盖安全审计和漏洞扫描等环节，以监测和防御潜在风险。技术层面上，实施坚固的加密技术和身份验证机制，防止未授权访问，增强系统的整体安全性。这些措施共同构成了强化数据安全保障的核心，有效抵御数据泄露和网络攻击。

4） 风险评估

为了确保人工智能模型的安全性和可靠性，需要加强模型评估，以判断其对潜在威胁的反应能力和逃避监管的可能性。这包括评估模型是否具有危险行为的倾向，并验证其行为是否与设计预期相符，同时对模型的内部机制进行审查。此外，风险评估是数据治理的持续任务，需要定期执行以识别和防范数据安全和隐私方面的风险。通过这些措施，组织可以制定有效的应对策略，确保数据和隐私得到有效保护，及时应对安全挑战。

5） 教育与培训

安全与隐私是人工智能应用的红线，为组织内部人员提供关于数据安全和隐私的培训，增强人员安全隐私的意识，有助于构建一个全员参与的安全文化。同时，向用户提供关于安全实践和隐私保护的教育，使其更加自觉地保护个人信息，进一步提升整个生态系统的安全性。

6） 监管与合规审计

建立有效的监管机制，监督人工智能系统的运行，以及进行定期的合规审计，确保系统的运行符合相关法规和政策，是保障治理有效性和可持续性的关键步骤。通过这一系列综合的治理方法，可以全面而有效地应对人工智能数据安全与隐私治理的复杂挑战。

2. 治理技术

1） 安全治理技术

异常数据检测：利用异常样本和良性样本的分布差异或在隐藏空间上的特征差异，检测数据中的异常值。

数据增强：数据增强对于对抗攻击、后门攻击、投毒攻击来说都是有效防御机制，在丰富数据集多样性的同时，可降低异常数据的有效性。

鲁棒训练：通过改进训练过程来降低恶意数据的影响，提高大模型面对对抗样本的预测准确率。

数字水印：水印技术是一种在数据中嵌入隐蔽标记的方法，用于

追踪数据来源，增强数据安全性和可溯源性。技术的最新进展在于对抗性水印，它通过复杂算法和模型抵抗包括篡改和定向攻击在内的多种攻击，保护数据完整性和隐私。

安全对抗：安全对抗技术对人工智能数据安全与隐私治理至关重要，安全对抗技术是一种主动的安全策略，通过模拟攻击行为来预先发现和修复系统潜在漏洞。通过提高安全对抗技术的有效性，人工智能系统能够更主动、全面地保护数据安全，确保用户信息不被恶意获取或滥用。

加密与脱敏：加密技术和安全协议是确保数据在传输和存储过程中不被未授权访问、窃听或泄露的关键手段，尤其是在云环境或网络传输中。数据脱敏技术通过变形敏感信息，保护个人隐私，同时保持数据的可用性，降低隐私风险，对 AI 数据安全治理起着核心作用。

2） 隐私治理技术

差分隐私：通过对数据加噪，确保训练集中某一数据不论存在与否，对模型预测结果的影响都有限，从而阻止攻击者根据模型输出推断数据集中的具体数据信息。

同态加密：同态加密在明文和密文上进行计算得到的结果相同，因此可以直接在加密后的隐私数据上进行运算，保障数据隐私。但同态加密时间复杂度高，面对海量数据效率较低。

安全多方计算：安全多方计算允许各参与方输入对其他方保密的情况下，根据输入共同计算一个函数，确保了整个系统中个体敏感数据的

隐私性。

匿名化： 匿名化技术的最新进展集中在不可逆加密方法和差分隐私技术上。不可逆加密方法通过使用无法逆向解密的算法，确保个体身份信息在处理后无法还原，有效保护数据主体隐私。差分隐私技术则通过在数据发布或处理时加入噪声，防止重新识别攻击，保护个体身份信息不被泄露。

概念擦除： 概念擦除技术通过修改数据中的敏感信息来降低隐私泄露的风险。最新的概念擦除方法不仅关注隐私保护，还考虑在信息擦除的同时保持数据的分析可用性。

（三） 数据伦理治理

在科技高速发展的背景下，对人工智能中伦理道德的关注程度及应对方法仍未完全跟上技术的发展步伐。生成式语言模型作为人工智能的最新范式，其通过深度学习技术和大规模数据集生成文本内容，其算法“平等”的保留了训练数据中的偏见、黄色、恐怖和暴力等不当内容，进而在应用中造成不可逆的严重后果。针对这些问题，设计及加强对人工智能所需数据集的管控，成为至关重要的发展方向。

1. 治理方法

1）制定数据伦理政策

在人工智能数据处理过程中，数据伦理政策的制定是首要且不可

或缺的一步。这些政策为整个数据处理流程提供了道德和法律的框架，确保了人工智能技术的公平、透明和负责任的使用。

为了确保数据伦理政策的有效实施，需要建立相应的监督机制和违规处理措施。例如，可以设立独立的数据伦理审查委员会，负责监督数据处理活动是否符合伦理准则和政策要求，并对违规行为进行调查和处理。

2）提升透明度和可解释性

透明度和可解释性是人工智能数据伦理的重要方面，它们有助于增强公众对人工智能技术的信任和理解。透明度要求数据处理过程公开、透明，让数据持有者能够了解数据被如何使用、与谁共享以及用于何种目的。

为了实现透明度，我们可以采取向数据持有者提供详细的数据处理说明、建立数据主体访问和更正其数据的机制、公开算法原理和模型结构等多种方式。

可解释性则强调人工智能模型应能够为其决策或预测提供合理的解释。这对于决策性人工智能模型尤为重要，用户需要理解算法是如何得出特定结论的，以便对结果进行评估，提升对模型的信任程度。

为了提高模型的可解释性，我们可以采用简洁明了的模型结构、提供易于理解的模型输出解释以及使用可视化工具展示模型决策过程等。

3）规范数据收集和标注

在人工智能数据处理过程中，数据收集和标注是两个关键环节，它们直接影响到模型的质量和性能。

在数据收集阶段，我们需要确保采集的数据具有代表性、多样性和均衡性，以避免模型在特定群体上的偏见和歧视。此外，我们还需要关注数据来源的合法性和道德性。

在数据标注阶段，我们需要注意避免歧视性的标签和评价。标注人员应接受数据标准相关培训，确保标注人员能够客观、公正地进行标注工作。同时还需要建立标注质量评估和审核机制，对标注结果进行定期检查和纠正，确保数据的准确性和一致性。

4) 开展风险评估和缓解措施

在人工智能数据处理过程中，我们需要对可能导致不公正或歧视性结果的风险进行评估，并采取相应的缓解措施。风险评估可以通过对模型性能进行定量分析、对数据处理流程进行审查以及对相关法规和社会期望进行解读等方式来实现。

针对评估结果中暴露出的风险点，我们可以采取多种缓解措施。例如，对于模型偏见问题，我们可以通过增加多样性样本、调整模型参数或使用公平性增强算法来降低偏见程度；对于数据泄露风险，我们可以加强数据加密、访问控制和安全审计等措施来保护数据安全；对于算法决策不透明问题，我们可以采用可解释性强的模型结构或提供模型输出解释来提高透明度。

5）定期审查和更新

随着法规和社会期望的不断变化以及人工智能技术的快速发展，我们需要定期审查和更新数据伦理准则和政策以确保其与时代保持一致。审查过程应包括对现有准则和政策的全面评估对相关法规和社会期望的解读以及对新技术进展的跟踪和分析等。

在审查过程中，如果发现现有准则和政策存在不足或过时之处，应及时进行更新和完善。更新后的准则和政策应重新发布并告知所有相关方以确保其得到有效执行。同时还需要建立持续监督机制来确保更新后的准则和政策得到长期有效的执行和维护。

2. 治理技术

偏差检测和纠正：识别算法中存在的偏见，并通过调整算法参数或重新训练来消除偏见。

反歧视性算法设计：确保算法在处理数据时不会因种族、性别、宗教等因素产生歧视。

责任与可追溯性技术：确保 AI 系统的决策过程有清晰的记录，出现问题时可以追溯到责任方。

三、 面向人工智能的数据治理步骤

当前产业对于人工智能的工程建设工作包括数据收集、数据预处理/清洗、特征工程、数据标注、数据划分、数据增强、模型训练、模型验证与测试、模型推理等九个阶段。

过程上，总体由业务方提出需求后，数据工程师、算法工程师、数据科学家等角色进行分头开发。其总体上缺乏一套科学的方法论将各个团队、人员、角色进行串联，形成“流水线”式的作业。从而往往导致团队间的协作困难、工作效率不高、工作成本不低、责任分工不清、应用效果不及预期等问题。

通过利用 DataOps[1]所强调的加强团队间协作沟通、要求数据流程具有可重复性和可追溯性、重视自动化和持续集成、关注数据集的监控和持续改进、强调数据流程的监控与反馈等特性。能够有效改善这些问题，实现数据流程的自动化、质量管理的持续改进以及合规性、伦理性和隐私保护的强化，从而提高人工智能项目的效率和效果。

以下我们将创新性的提出利用 DataOps 理念来赋能人工智能模型研发、治理和运营的一体化流程方法，为产业提供一种实践思路，并在今后的报告中不断打磨。

[1] DataOps：数据研发运营一体化（DataOps）是数据开发的新范式，将敏捷、精益等理念融入数据开发过程，通过对数据相关人员、工具和流程的重新组织，打破协作壁垒，构建集开发、治理、运营于一体的自动化数据流水线，不断提高数据产品交付效率与质量，实现高质量数字化发展。——《DataOps 实践指南》

（一）明确应用目标与需求

确定应用目标与效果：首先，与相关方共同明确人工智能应用的目标和预期效果。这可能涉及解决的业务问题、改善的业务流程或实现的业务目标。

明确应用场景：确定人工智能应用的具体场景和应用范围。这包括确定应用的环境、用户和操作方式。

澄清数据需求：与算法方、需求方和数据方共同澄清所需的数据类型、数据数量和数据质量标准。这可能涉及确定需要的数据来源、数据格式、数据标签等。

评估数据获取难度：评估当前已有数据与所需数据的可获取性和使用难度。这包括考虑数据获取的法律、道德和技术限制，以及数据采集和标注的成本和时间。

（二）设计数据集治理要求与规范

制定数据标注流程：制定清晰的数据标注流程，包括数据标注的

步骤、标准和质量控制措施。这可能涉及确定标注人员的资质要求、标注工具的选择和标注结果的审核流程。

明确数据质量标准：明确数据质量的衡量标准和评估方法，包括数据准确性、完整性、一致性和可靠性等方面。这可以通过制定数据质量指标和监控机制来实现。

考虑合规和隐私要求：确保数据采集、标注和使用过程符合法律、道德和行业规范的要求，特别是涉及个人隐私和敏感信息的情况下。这可能涉及制定数据保护策略、访问控制机制和数据使用协议。

管理规范持续更新：建立一个持续更新的数据管理规范，确保规范与技术发展和业务需求的变化保持同步。这可能包括定期审查和更新数据管理政策、流程和工具。

（三）开展数据集处理与研发

数据清洗和标注：进行数据清洗、标注和预处理，以确保数据质量和一致性。这可能涉及识别和处理数据中的噪音、缺失值和异常值，以及为数据添加标签和元数据。

特征工程和数据增强：进行特征工程和数据增强，以提取数据的有效特征并增加数据的多样性。这可以通过使用统计方法、机器学习算法和数据增强技术来实现。

沟通与验收：加强算法方与数据方的沟通，确保数据处理过程中的偏差得到及时纠正。这可能包括定期的数据处理进展报告和算法人员的阶段性验收动作。

（四）进行数据集洞察与交付

数据集持续维护：对交付的数据集进行持续维护和更新，确保数据集的及时性和适用性。这可能包括定期的数据质量评估、数据集版本管理和数据集权限管控。

洞察数据集构成与分布：对当前企业内数据集的构成、分布、质量和成本进行洞察，以优化数据集的组织和利用。这可以通过数据集分析和数据集使用情况监控来实现。

版本管理与权限管控：建立数据集的版本管理系统和权限管控机制，确保数据集的一致性和可追溯性。这可能包括对数据集的版本记录、变更审批和访问权限控制等。

（五）持续数据集运营与优化

构建监控指标体系：构建全局的监控指标体系，综合考虑数据集的使用频率、更新频率、质量评价和成本效益等方面。这可以通过建立数据集运营指标和监控仪表板来实现。

持续优化数据集流程：对整个数据集构建周期的流程进行持续优化，提高数据集的效率和效果。这可能包括对数据处理流程的自动化、工作流程的优化和团队协作的改进等方面。

四、 展望

（一）人工智能数据产业分工更加明确

总体来看，产业中存在大量对数据的重复标注、重复采集、重复加工的现象。这无疑是对时间、资金、资源和人才的铺张浪费。

未来，供给人工智能数据集的产业在采集、加工、交易、消费等环节将更加清晰和成熟，通过市场化的调节机制可以合理分配产业的人才、资金与资源，更高效的推进人工智能应用发展。

（二）数据治理或成为大模型的胜负手

当前，产业界普遍通过大力发展大模型产品应用来抢占市场份额。然而，随着市场集中度的提高（马太效应），只有少数企业可能在这场竞争中胜出。通过有效的数据治理，企业将有可能获得竞争优势。

（三）服务化

随着 DG4AI 技术、实践和理论的成熟，DG4AI 将更加标准化与流程化，进而发展为服务化，能够高效、高质量、安全可控的提供标准的数据产品（数据集、语料库）。

附录一 国内外政策法规与标准建设发展情况

（一）美欧人工智能数据治理政策实践概述

随着人工智能的快速发展，从ChatGPT迭代更新到Sora的重磅推出，人工智能自动化的程度不断提高，以远超人类的生产力和效率水平完成日益复杂的任务，并在此过程中产生巨大的社会和经济效益。但同时产生的数据安全和隐私保护问题引发了各界的担忧和关切，成为人工智能在开发和应用过程中面临的严峻安全挑战。因此，探索人工智能数据治理成为当下急需解决的问题。美国和欧盟作为较早开展人工智能数据治理工作的典型代表，分析其战略部署到立法的政策实践经验，能够为我国开展人工智能数据治理提供有益借鉴。

1. 美国人工智能数据治理

1） 战略层面

通过对美国人工智能政策梳理发现，在人工智能数据治理战略层面，美国重视高质量数据集和模型开发与评估以及个人隐私保护，整体上体现出充分的政策的供给，确保美国领先地位。

特征一：重视数据集和模型开发与评估

早在2016年10月，美国白宫科技政策办公室（OSTP）发布《国家人工智能研究与发展战略计划》，并在2019年和2023年进行更新。该计划在战略1“长期投资基础和负责任的人工智能研究”中提出以数据为中心的知识发现方法，并将“建立并共享人工智能培训和测试专用公共数据集和环境”作为人工智能发展战略之一，具体包括：开发可无障碍访问数据集，满足多样化的人工智能应用需求；开发大规模、专业化的共享先进计算和硬件资源；加快测试资源对商业和公共利益的响应；开发开源软件库和工具箱。2019年2月，美国总统特朗普签署行政令《维护美国人工智能领导地位的行政命令》（Executive Order on Maintaining American Leadership in Artificial Intelligence）中明确提出：所有相关机构的负责人都应该重新评估其拥有的联邦数据与模型，以确定是否有机会增加非联邦AI研究团体的访问与运用，同时要确保数据安全、隐私和机密性。具体而言，各机构应改进数据和模型库存文档，以实现发现和可用性，并应根据AI研究团体的用户反馈，优先改进AI数据和模型的访问和质量。

特征二：重视个人隐私保护

2022年10月4日，美国白宫科技政策办公室（OSTP）发布《人工智能权利法案蓝图》（Blueprint for an AI Bill of Right）提出将数据保护隐私作为人工智能技术应用的五项原则之一，默认将隐私保护嵌入到人工智能系统中，并鼓励人工智能系统在可能的情况下尊重个人关于收集，使用，访问，传输和删除个人数据的决定，从而确保公众不受滥用数据行为的影响，拥有关于个人数据使用方式的自主权。蓝图指出，数据隐私是实现该框架中所有其他原则所需的基础性和交叉性原则。认为目前美国缺乏一个全面的法律或监管框架来管理公众在个人数据方面的权利。基于此，蓝图提出了几点对自动化系统的期望：首先，自动化系统应通过设计和默认来保护隐私。自动化系统的设计和构建应以默认方式保护隐私，且数据收集和使用的范围应该是有限的，有具体的、狭窄的确定目标。同时，系统能主动识别和缓解风险，并保护隐私的安全。其次，保护公众免受不受约束的监视，这需要有限和相称的监视系统，并加强对这些系统的监督，对监控进行范围限制，从而保护权利和民主价值。再次，应为公众提供适当和有意义的同意、访问和控制其数据的机制。系统在特定用途的需求上需要征得用户同意，且同意请求应是简短、通俗、直接的。

同时，确保用户数据访问的便利、撤销与删除数据的及时性以及系统本身的支持。最后，需进行独立评估和报告证明数据隐私和用户控制受到保护。除此之外，蓝图还指出要对敏感领域相关数据提供有额外保护措施，收集、使用、共享或存储与这些敏感领域相关的数据的自动化系统应为数据提供强化保护。并且蓝图认为，通过法律、政策和实际的技术和社会技术方法来保护权利和机会，能够使得这些原则落地从而服务民众。

2023 年 10 月 30 日，美国总统拜登于签署颁布《关于安全、可靠、值得信赖地开发和使用人工智能的行政命令》（Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence），该行政令明确了美国政府对待人工智能的政策法制框架，将隐私保护列为八个重点行动领域之一，提出联邦机构应按照指示要求，加强隐私保护研究和技术，包括让人工智能系统在训练的同时保护训练数据隐私的技术，支持和评估针对人工智能的隐私保护技术，并制定进一步的隐私指南，以应对人工智能风险。此外，拜登总统还呼吁国会通过两党数据隐私立法。

2） 立法层面

随着人工智能的不断发展，特别是生成式人工智能技术的迅速迭代以及应用的快速普及，使得美国在人工智能立法层面“广开听证、高频提案”。本节针对美国近期人工智能提案的数据治理部分进行梳理，探索美国人工智能数据治理的特征和趋势。主要包括五个部分：

第一：建立人工智能机构

2023 年 2 月 11 日，美国提出《确保人工智能安全、可靠、道德和稳定的系统法案》或《评估人工智能法案》（Assuring Safe, Secure, Ethical, and Stable Systems for AI Act or ASSESS AI Act），该法案指示总统任命一个工作组，评估人工智能（AI）对隐私、公民权利和公民自由的影响。工作组应评估人工智能应用和相关数据的现有政策、监管和法律空白；并向国会和总统提出立法和监管改革建议，以确保在联邦政府运作中使用人工智能和相关数据符合言论自由、平等保护、隐私、公民自由、公民权利和正当程序。2023 年 7 月 13 日，美国国会提出《人工智能领导法案》(AI LEAD Act）旨在设立首席人工智能官委员会、首席人工智能官和人工智能治理委员会。指出各机构的职责。其职责包括：每个机构的负责人应确保该机构负责地研究、开发、获取、应用、治理和使用符合民主价值观的人工智能，包括隐私、公民权利和自由、信息安全、非歧视原则、透明度和可靠性。此外，各机构负责人应制定人工智能战略，使机构可信赖地采用人工智能，以更好地实现机构为美国人民服务的使命。战略内容应包括该机构将如何与私营部门合作，确保采购的人工智能系统或能力包括保护个人权利和安全以及保护联邦政府数据和其他信息的保护措施等。2023 年 9 月 28 日，美国国会提出《防止深度虚假诈骗法案》(Preventing Deep Fake Scams Act)，社交媒体的普及使得不法分子更容易获得潜在目标的视频和音频。这些材料可以用来复制其他人的声音和外表，以进行数据盗窃、身份盗窃或欺诈。该法案旨在建立金融服务部门人工智能工作组，就金融服务部门人工智能相关问题向国会报告，内容包括：对不良行为者使用人工智能窃取消费者数据、窃取消费者身份和实施欺诈可能导致的潜在风险的描述。

第二：提升政府机构人工智能素养

2023 年 4 月 27 日，美国国会提出《人工智能领导力培训法案》（AI Leadership Training Act），要求人事管理办公室主任为联邦管理官员和监管人员及其他目的制定或以其他方式确保提供人工智能培训计划。该计划应包括的信息有：数据在人工智能系统中的作用，以及在这些系统中没有使用具有足够代表性的培训数据的风险，包括与偏见相关的风险。

第三：重视人工智能应用场景化立法

2023 年 5 月 16 日，美国国会提出《儿童人工智能保护法案》(AI Shield for Kids Act)，该法案限制未成年人使用包含人工智能（AI）功能（如聊天功能）的程序、服务、应用程序或其他产品。具体来说，联邦通信委员会（FCC）必须发布规则，禁止任何实体（如社交媒体公司）在未经未成年人父母同意的情况下向未成年人提供具有人工智能功能的产品。2023 年 9 月 12 日，美国国会《保护选举免受欺骗性人工智能法案》(Protect Elections from Deceptive AI Act)，禁止分发与联邦公职候选人及其他目的有关的具有欺骗性的人工智能生成的音频或视频媒体。2024 年 1 月 24 日，美国国会发现，人工智能（AI）技术的最新进步和深度伪造软件的发展对个人保护自己的声音和形象不被盗用的能力产生了不利影响，提出《禁止人工智能欺诈法案》(No AI FRAUD Act)，该法案旨在规定个人的肖像权和声音权。

第四：加强人工智能生成内容与机理的披露

2023 年 5 月 2 日，美国国会提出《真实政治广告法》(REAL Political Advertisements Act)，该法案扩大了对政治竞选的某些披露和免责要求，包括要求对含有人工智能（AI）生成内容的广告进行免责。具体而言，该法案要求通信（例如，政治广告）如果通信包含全部或部分使用人工智能生成的图像或视频片段，则必须以明确和明显的方式包括声明。2023 年 6 月 5 日，美国国会提出《人工智能披露法案》(AI Disclosure Act）要求生成式人工智能披露其输出是由人工智能生成的，并用于其他目的。具体要求披露生成式人工智能的使用情况，生成式人工智能应在该人工智能产生的任何输出中包括以下内容：（1）免责声明“免责声明：此输出是由人工智能产生的。”（2）联邦贸易委员会的执法。2023 年 9 月 12 日，美国国会《人工智能生成内容咨询法案》(Advisory for AI-Generated Content Act)，指出为人工智能生成的材料和其他目的要求水印。“人工智能生成的材料”是指可以生成各种类型内容的人工智能技术，包括文本、图像、音频或合成数据。2023 年 11 月 21 日，美国国会《人工智能标签法案》(AI Labeling Act）要求披露人工智能生成的内容，其中输出的元数据信息应包括内容为人工智能生成内容的标识、用于创建内容的工具的标识以及创建内容的日期和时间。2023 年 12 月 22 日，美国提出《人工智能基础模型透明度法案》（AI Foundation Model Transparency Act），指导联邦贸易委员会建立标准，公开人工智能基础模型中使用的训练数据和算法。该提案中要求基础模型的创建人披露训练数据的来源，以便于原始的版权。持有者知道自己的作品被用于模型训练。

第五：构建人工智能风险问责框架

2023 年 10 月 25 日，美国国会提出《人工智能问责法案》(Artificial Intelligence Accountability Act)，该法案要求美国国家电信和信息管理局（NTIA）研究并报告人工智能（AI）系统的问责措施，例如人工智能系统的问责措施如何有助于弥合数字鸿沟，并协助促进美国的数字包容。具体来说，NTIA 必须研究、征求利益相关者的反馈，并向国会报告有关机制（例如审计、认证和评估），以确保人工智能系统是值得信赖的。2023 年 11 月 2 日，美国国会提出《联邦人工智能风险管理法》(Federal Artificial Intelligence Risk Management Act)，该法案指示联邦机构使用由美国国家标准与技术研究院（NIST）制定的关于人工智能（AI）使用的人工智能风险管理框架。联邦采购政策管理员应与 NIST 院长协商，提供合同语言草案，供各机构在需要人工智能供应商的采购中使用，包括提供对 NIST 院长定义的数据、模型和参数的适当访问，以便进行充分的测试和评估、验证和确认。2023 年 11 月 15 日，美国国会提出《人工智能研究、创新和责任法案》(Artificial Intelligence Research, Innovation, and Accountability Act)，为人工智能创新和问责制提供框架。该法案通过设置人工智能系统分类认证及信息披露等制度，为人工

智能系统的开发、识别和部署的基础立法工作奠定基础。

2. 欧盟人工智能数据治理

1) 战略层面

2018 年 4 月 25 日，欧盟委员会发布《欧洲人工智能战略》(The Age of Artificial Intelligence: Towards a European Strategy for Human-Centric Machines）政策文件，首次系统地提出了欧盟版的人工智能发展战略规划，并提出了首个欧洲人工智能倡议和确定了欧洲人工智能发展的三大目标：（1）提升欧盟技术、产业能力，推进人工智能的广泛应用。（2）为应对人工智能技术可能带来的社会经济变革做好准备。（3）在欧盟现有价值观与《欧盟基本权利宪章》的基础上，确立合适的人工智能伦理与法律框架。作为该政策文件的核心内容，三大目标明确了欧盟人工智能发展的总体原则，为后续政策文件的出台提供了指导方针。

2019 年 4 月 8 日发布了由人工智能专家委员会撰写的《可信人工智能伦理指南》(Ethics Guidenelines for Trustworthy AI)，旨在为人工智能技术的开发和应用提供指导、约束，推动人工智能技术的可持续、负责任和可信发展。并提出将隐私与数据保护列为 7 个实践原则的关键要求之一。

2020 年 2 月 19 日欧盟发布了《人工智能白皮书》(White Paper on Artificial Intelligence)，主要围绕“卓越生态系统”（ecosystemof excellence）和“信任生态系统”（ecosystem of trust）两个方面的展开，着重构建了可信赖与安全的人工智能监管框架。

2021 年 4 月 21 日欧盟发布更新版本《人工智能协调计划 2021 年审查》(Coordinated Plan on Artificial Intelligence 2021 Review)是对 2018 年首次发布的《人工智能协调计划》的更新，报告主要集中在三个方面：建立一个能够有效获取、积累并分享人工智能政策见解的治理框架；挖掘数据潜能以充分释放其潜力以及建设关键基础设施，以支持能力建设并促进人工智能的发展。其中，数据层面的原因是：人工智能技术的发展通常需要大规模、高质量且安全可靠的数据集，因此，确保数据能够在符合欧盟法规（包括保护个人数据的《通用数据保护条例》以及欧盟的国际承诺）前提下，实现欧盟内部、与贸易伙伴之间以及跨部门的“流动”，是非常重要的。

2022 年 7 月 25 日，欧洲议会未来与科学和技术小组（STOA)，发布《审计算法决策系统中使用的数据集的质量》(Auditing the quality of datasets used in algorithmic decision-making systems）报告，主要聚焦在算法决策所使用的数据集的审计。研究发现有：减轻偏见的一个重要步骤是创建或使用高质量的特定领域训练数据集，以保证对基于人工智能的系统公平地表示“现实世界”的知识。应实施监督和问责机制，以不断评估数据的质量和完整性等。并提出相关的政策建议，包括： 不需要针对歧视问题的专门立法；在数据收集阶段减少歧视和偏差；推进数据集的认证；为受人工智能决策影响的主体提供数据访问权；促进人工智能法案的实施。

2) 立法层面

欧盟人工智能数据治理特征：首先从以数据为中心转变为以平台为中心，最后形成以人工智能为中心的严格监管模式。

第一：以数据为中心。2018 年 5 月 25 日，欧盟正式实施《通用数据保护条例》（GDPR），在数据安全及隐私安全上制定了全球最严格的合规政策，其中涉及人工智能的主要有：①GDPR 要求人工智能的算法具有一定的可解释性，例如：第 5 条规定个人数据处理必须遵循：合法地、公平地并且以公开透明的方式对数据主体的个人数据进行处理等。②GDPR 第 22 条指出数据主体有权不受仅基于自动化处理行为得出的决定的制约。为了确保

人工智能的发展尊重人权并获得信任。

第二：以平台为中心。2022 年 7 月 5 日，欧洲议会通过《数字服务法》（Digital Services Act，DSA）和《数字市场法》（Digital Markets Act，DMA），《数字服务法》旨在更好地保护用户及其在线基本权利，加强对在线平台的公共监督，建立在线平台的透明机制和明确的问责框架，杜绝算法偏见威胁平台用户权益，确保平台对其算法负责，防止系统滥用及研究人员随意访问平台的关键数据，而对于超大型平台，欧盟委员会加强监督和执法，以实现对用户的有效保护。《数字市场法》旨在规制互联网巨头（通常是指满足拥有强大的经济地位，对内部市场影响显著，活跃于多个欧盟国家等条件的大型在线平台），防止其滥用市场支配地位，解决数据垄断问题，并建立一套狭义的客观标准，将大型在线平台充当数字市场的“守门人”（gatekeeper），为守门人规定义务，规定其在日常运营中必须遵守的要求，承担起数据共享、算法公开等义务，如允许它们的业务用户访问在使用平台时生成的数据，禁止平台阻止用户有意愿地卸载任何预装的软件或应用程序等。

第三，以人工智能为中心。2023 年 6 月 14 日，欧盟议会通过了《人工智能法案》（AI Act），这是世界上首部针对人工智能的重大立法。法案指出高质量数据和获取高质量数据在提供结构和确保许多人工智能系统的性能方面发挥着至关重要的作用，特别是在使用涉及模型训练的技术时，目的是确保高风险人工智能系统按预期安全运行，并且不会成为欧盟法律禁止的歧视来源。用于训练、验证和测试的高质量数据集需要实施适当的数据治理和管理实践。法案第 10 条数据和数据治理相关规定包括：

1. 使用涉及使用数据训练模型的技术的高风险人工智能系统应基于符合第 2 至 5 段所述质量标准的培训、验证和测试数据集进行开发。

2. 培训、验证和测试数据集应遵守适当的数据治理和管理实践。这些实践应特别关注：

(a) 相关的设计选择；

(b) 数据收集过程；

(c) 相关的数据预处理操作，如注释、标签、清洗、丰富和聚合；

(d) 制定相关的假设，特别是关于数据应测量和表示的信息；

(e) 对所需数据集的可用性、数量和适用性的预先评估；

(f) 检查可能影响自然人的健康和安全或导致联盟法律禁止的歧视的潜在偏见；

(g) 确定任何可能存在的数据差距或不足之处，以及如何解决这些问题。

3. 培训、验证和测试数据集应具有相关性、代表性，并在最大程度上无误且完整。它们应具有适当的统计特性，包括在适用的情况下，关于高风险人工智能系统预期使用的人群或人群组的特性。这些数据集的特性可以在单个数据集或其组合的水平上满足。

4. 培训、验证和测试数据集应根据预期目的，在必要的程度上考虑特定地理、行为或功能环境的特征或要素，即高风险人工智能系统预期使用的环境。

5. 在确保高风险人工智能系统进行偏见监测、检测和纠正的目的严格必要的情况下，这些系统的提供商可以根据《欧洲议会条例（EU）2016/679》第 9(1）条、《欧盟指令（EU）2016/680》第 10 条和《欧洲议会条例（EU）2018/1725》第 10(1）条的规定处理特殊类别的个人数据，前提是采取适当的保护措施，保障自然人的基本权利和自由，包括采用最先进的安全和隐私保护措施，如匿名化或在匿名化可能显著影响所追求目的的情况下采用加密技术。

6. 对于不使用模型训练技术的高风险人工智能系统，第 2 至 5 条仅适用于测试数据集。

（二） 标准现状

1. 国际标准（ISO/IEC、ITU、IEEE）

表 1 人工智能数据治理国际标准一览

序号	标准号	标准名称
1	ISO/IEC 8183	Information technology — Artificial intelligence — Data life cycle framework
2	ISO/IEC DIS 5259-1	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples
3	ISO/IEC DIS 5259-2	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 2: Data quality measures
4	ISO/IEC DIS 5259-3	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 3: Data quality management requirements and guidelines
5	ISO/IEC DIS 5259-4	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 4: Data quality process framework
6	ISO/IEC DIS 5259-5	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 5: Data quality governance framework
7	ISO/IEC CD TR 5259-6	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 6: Visualization framework for data quality
8	ISO/IEC AWI TR 42103	Information technology — Artificial intelligence — Overview of synthetic data in the context of AI system
9	ITU-TF.748.13	Technical Framework for Shared Machine Learning System
10	IEEE 2830-2021	Standard for Technical Framework and Requirements of

2. 国内标准

表 2 人工智能数据治理国内标准一览

序号	标准号	标准名称
1	GB/T 42755-2023	人工智能 面向机器学习的数据标注规程
2	DB14/T 2463—2022	人工智能 数据标注总体框架
3	DB14/T 2464—2022	人工智能 数据标注一般技术要求
4	DB14/T 2465—2022	人工智能 数据标注通用工作规程
5	20231740-T-469	人工智能 风险管理能力评估
6	20231736-T-469	人工智能 预训练模型 第 1 部分：通用要求
7	20231746-T-469	人工智能 预训练模型 第 2 部分：评测指标与方法
8	20231741-T-469	人工智能 预训练模型 第 3 部分：服务能力成熟度评估
9	20231169-T-469	人工智能 联邦学习技术规范
10	20232020-T-469	人工智能 多算法管理技术要求
11	YY/T 1833.4-2023	人工智能医疗器械 质量要求和评价 第 4 部分：可追溯性
12	YY/T 1907-2023	人工智能医疗器械 冠状动脉 CT 影像处理软件 算法性能测试方法
13	—	网络安全技术 人工智能生成合成内容标识方法

附录二 相关名词解释

提出者	应用名词	名词解释
Franz J（2003）	Artificial Intelligence	人工智能（AI）是计算机科学领域的一门学科，受到认知科学/心理学、哲学、语言学和数学的影响。主要目标是创建包含或表现出一定智能的系统。以及研究人类执行需要智能的任务的方式。
Nemade et al（2022）	Artificial Intelligence	人工智能（AI) 与计算机科学相协调，致力于开发能够完成通常需要人类智能才能完成的任务的计算机操作机器。
Kok et al（2009）	Artificial Intelligence	对于 AI 的定义很混乱，并随着时间发生变化，但是主要的定义方式可以分为像人类一样思考的系统 、像人类一样行动的系统、理性思考的系统、 理性行动的系统。

提出者	应用名词	名词解释
IBM	data governance	数据治理是通过不同的策略和标准提高组织数据的可用性、质量和安全性。这些过程决定了数据所有者、数据安全措施以及数据的预期用途。总体而言，数据治理的目标是维护高质量的数据。
GBT34960.5-2018	数据治理	数据资源及其应用过程中相关管控活动、绩效和风险管理的集合。
ISO8000-2:2022	data governance	制定和执行与数据管理相关的政策
吴信东 等人 （2019）	数据治理	数据治理从本质上看就是对一个机构（企业或政府部门）的数据从收集融合到分析管理和利用进行评估、指导和监督（EDM）的过程， 通过提供不断创新的数据服务， 为企业创造价值
艾瑞咨询	数据治理	数据治理以数据源汇入为伊始，对数据进行

（2022）	清洗加工，并在数据存储、数据计算、数据服务应用等环节予以持续的治理服务，是企业实现数据服务与应用的重要环节。从数据层面来看，数据本身存在着从生产到消亡的生命周期，而数据治理会 在数据生命周期的各阶段通过相应工具与方法论进行规范与定义，在企业内部构建出切实有效的数据闭环，使数据发挥出 更大的价值。

大数据技术标准推进委员会
Big Data Technology and Standard Committee

联系人:尹正
联系电话:15810811776
联系邮箱:yinzheng@caict.ac.cn