中国联通发布人工智能共享数据集 AIData 并发起共建高质量数据集合作倡议

发布时间：2024-07-22 发布人：新闻宣传中心

2024 年 7 月 19 日下午，中国联通合作伙伴大会“人工智能创新发展论坛”在上海盛大召开。会上，中国联通数字化部副总经理娄瑜正式发布了中国联通人工智能共享数据集 AIData，并与新华社、阿里云、腾讯云、百度智能云、库帕思、人民数据、南方电网、中国数字文化集团、中国航信、中国信通院等多个合作伙伴联合发起了“共建高质量数据集合作倡议”。

这次发布会不仅仅是一次产品的亮相，更是中国联通在人工智能数据集建设方面迈出的重要一步。中国联通致力于通过汇集各行业的力量，共同推动构建大规模、多模态的高质量数据集，为推动人工智能技术的发展提供坚实的基础。

AIData 的核心优势

中国联通的 AIData 数据集在多个方面展现了其独特的优势和领先地位。首先，在数据获取速度上，中国联通展现了卓越的效率。通过高效的收集和处理机制，AIData 能够迅速整合来自移动通信、政务和新型工业化等多个重点行业的数据资源，沉淀了 13TB 的多模态高质量数据集。这种高速度的数据获取不仅提升了数据集的及时性，还确保了数据的鲜活度，使得 AIData 能够更好地适应快速变化的市场需求和技术发展。其次，AIData 的数据质量得到了充分的保障。中国联通采用了先进的数据筛选和清洗技术，确保每一份数据都经过严格的质量控制。无论是结构化数据还是非结构化数据，AIData 都能够提供高精度和高一

致性的数据集，这包括了 600PB 的结构化数据和 5PB 的非结构化数据。这种高质量的数据不仅提升了人工智能模型的训练效果，还大大降低了数据噪音对模型准确性的影响。在数据供给方面，AIData 展示了强大的稳定性和可靠性。中国联通通过建设高性能的数据中心和采用领先的数据传输技术，确保了数据供给的连续性和稳定性。无论在任何时候，用户都能够通过 AIData 平台获取到所需的高质量数据资源。这种稳定性和可靠性为各行业的人工智能应用提供了坚实的基础，确保了模型训练和应用的顺利进行。安全性是 AIData 的一大亮点。中国联通在数据安全方面投入了大量的资源，采用了多层次的安全防护措施，确保数据在存储、传输和使用过程中的安全性。AIData 平台配备了先进的加密技术和访问控制机制，有效防止数据泄露和未经授权的访问。这不仅保护了用户的数据隐私，还增强了用户对 AIData 平台的信任。最后，AIData 在数据的多样性和覆盖面上也具有显著的优势。中国联通通过与多个行业的深度合作，汇集了来自各行各业的丰富数据。这些数据不仅涵盖了传统行业，还包括了新兴领域的数据资源，通过 18 个行业军团的深度融入，确保了数据的多样性和广泛的覆盖面。这种多样性和广泛的覆盖面，使得 AIData 能够满足不同应用场景的需求，为各类人工智能应用提供全面的数据支持。

数据共享与合作倡议

共享数据集是中国联通在推动人工智能技术发展方面的重要举措之一。AIData 的推出不仅体现了中国联通在数据领域的领先地位，更展示了其在数据共享与合作方面的决心。中国联通深知，单靠一己之力难以应对日益复杂的人工智能数据需求，因此，借助合作伙伴的力量，共同构建一个开放、共享的数据生态系统显得尤为重要。中国联通在企业内部率先实现了电信行业内首家内部数据

100%集约化运营。这意味着，企业内部所有的数据都经过统一管理和优化，提高了数据的利用效率和质量。这种集约化运营不仅提升了企业内部数据的价值，更为外部合作提供了坚实的基础。在此基础上，中国联通还构建了完善的数据开放运营机制。通过这一机制，企业内部数据能够更高效地流通和共享，为人工智能模型的训练和应用提供了丰富的数据资源。对外，中国联通通过组建 18 个行业军团，深入政务、医疗、工业、交通、矿山、钢铁等 16 个行业的一线。这些行业军团不仅是数据采集的前沿阵地，更是数据应用的桥头堡。通过与各行业的深度合作，中国联通能够充分洞察行业 AI 应用需求，深入挖掘各领域的专业数据，形成具有行业特色的高质量数据集。这种深度融合不仅提升了数据的专业性和适用性，更为各行业的数字化转型和智能化升级提供了强有力的支持。为了推动数据共享与合作，中国联通联合多家行业伙伴共同发布了“共建高质量数据集合作倡议”。这一倡议以多方参与、鼓励数据共享共用、严守数据保护法律法规、坚持高水平的数据治理与关注数据的社会责任为主要原则，制定了数据集建设、数据平台开放、共性技术开发、商业模式创新、交流社区建设 5 个行动计划。通过这些行动计划，中国联通希望能够形成一个协同发展的数据生态系统，推动数据资源的高效利用和创新应用。中国联通在数据共享方面的努力还体现在技术支撑能力的建设上。通过创新联通链、可信数据空间技术与场景化数据质量管理体系，中国联通能够为数据共享提供全方位的技术保障。这些技术不仅提升了数据的安全性和可靠性，还增强了数据的可用性和可追溯性，为数据共享和合作提供了坚实的技术基础。

面向未来的展望

尽管中国联通的 AIData 在数据集建设方面已经取得了显著的成就，但未来

依然面临着诸多挑战。例如，数据流通共享不足是一个亟待解决的问题。当前，各行业的数据壁垒依然存在，数据孤岛现象严重，限制了数据的跨行业流动和共享。中国联通认识到，仅靠单个企业的努力难以打破这种局面，因此需要在更大范围内推动数据共享机制的建立，打破数据壁垒，实现数据资源的最大化利用。政策的不完善也是一大障碍。尽管国家层面已经出台了一系列支持数据开放和共享的政策，但在具体实施过程中，依然存在政策落地难、执行不到位等问题。这不仅影响了数据的合法合规使用，也限制了数据价值的充分发挥。中国联通在这一方面积极与政府部门沟通，推动相关政策的细化和落实，确保数据共享的合法合规性，为行业树立了良好的示范作用。质量难以评价是另一个重要挑战。在数据量急剧增加的背景下，如何确保数据的高质量成为一大难题。数据质量直接影响到人工智能模型的训练效果和应用性能。因此，中国联通在数据质量管理方面投入了大量资源，通过建立严格的数据筛选和清洗机制，确保数据的准确性、一致性和完整性。此外，中国联通还引入了第三方数据质量评估机构，对数据集进行独立评估，确保数据质量符合行业最高标准。开放过程中的可信性与安全性问题也不容忽视。随着数据开放共享的推进，数据安全和隐私保护问题日益凸显。如何在保证数据开放的同时，确保数据不被滥用和泄露，成为一大挑战。中国联通在这一方面采用了多层次的安全防护措施，包括数据加密、访问控制和审计追踪等技术手段，确保数据在开放共享过程中的安全性和可信性。在应对这些挑战的过程中，中国联通积极布局产业生态，通过率先发布 AIData 人工智能共享数据集，旨在协同打造人工智能数据集建设共享的机制体系。中国联通通过完善技术支撑能力，进一步推动人工智能共享数据集的发展态势，助力国家战略打造新质生产力。高质量数据集的建设需要汇集全行业的力量共同推动构建。为此，中

国联通联合各行业伙伴共同发布了“共建高质量数据集合作倡议”。这一倡议以倡导多方参与、鼓励数据共享共用、严守数据保护法律法规、坚持高水平的数据治理与关注数据的社会责任为主要原则，制定了数据集建设、数据平台开放、共性技术开发、商业模式创新、交流社区建设 5 个行动计划。通过这些行动计划，中国联通希望能够形成一个协同发展的数据生态系统，推动数据资源的高效利用和创新应用。

在数据集建设方面，中国联通强调多方参与的重要性。各行业的参与不仅可以丰富数据集的内容，还可以带来不同的视角和专业知识，提升数据集的多样性和适用性。数据共享共用是提升数据价值的重要途径。通过建立数据共享平台，打破数据孤岛，实现数据资源的互联互通，为各行业的人工智能应用提供丰富的数据支持。严守数据保护法律法规是数据共享的前提和保障。中国联通在数据保护方面严格遵守国家相关法律法规，通过建立完善的数据治理体系，确保数据共享过程中的合法合规性。高水平的数据治理是数据集建设的关键。中国联通通过引入先进的数据治理技术和管理方法，提升数据治理水平，确保数据的高质量和高可信度。关注数据的社会责任是中国联通一直以来坚持的原则。数据不仅是一种资源，更是一种责任。中国联通在数据共享和开放过程中，始终关注数据的社会影响，确保数据的使用符合社会道德和伦理标准，为构建和谐社会贡献力量。通过这些努力，中国联通将联合各行业伙伴，共同推动人工智能数据集构建，强化以数据和知识供给赋能数智化转型升级，为经济社会创造更大价值。这不仅是对当前人工智能数据集建设工作的总结，更是对未来发展方向的展望。中国联通将继续发挥其在数据和知识供给方面的优势，推动数智化转型升级，为经济社会创造更大价值。