

平衡发展的 AI 伦理观

——AI 可持续发展报告 2021-2022



坚持原创，让 AI 引领人类进步

面向全球，商汤科技坚持原创，致力于让 AI 推动经济、
社会和人类的发展，建设更美好的未来。

关于商汤

关于本报告

第一部分 概述08

- » 人类伦理的变迁：从“机器学习”到“机器伦理”09
- » 文明伦理的求同：“平衡发展”的 AI 伦理观10
 - » “可持续发展”原则：发展权是最大的社会责任与人权13
 - » “以人为本”原则：济困扶弱，包容尊重14
 - » “技术可控”原则：人类负责，敏捷治理15

第二部分 可持续发展实践18

- » 商汤实践：智能电网，“智能+”引领绿色能源革命19
- » 商汤实践：智慧医疗普惠全球20
- » 商汤实践：推广人工智能基础教育22

第三部分 以人为本实践26

- » 商汤实践：商汤智慧医疗 SenseCare 平台，赋能前线医生26
- » 商汤实践：赋能智慧文旅，迎接元宇宙世界到来28
- » 商汤实践：智慧乘客服务平台，赋能成都城市地铁交通系统30

第四部分 技术可控实践32

- » 商汤实践：数据隐私保护33
- » 商汤实践：算法 / 芯片评测34
- » 商汤实践：科研课题研究36

第五部分 AI 伦理与治理实践38

- » 商汤实践：组建 AI 伦理与治理委员会39
- » 商汤实践：人工智能伦理与治理风险控制机制41
- » 商汤实践：AI 伦理标准制定47
- » 商汤实践：AI 伦理外部联合研究47

结语50

【关于商汤】



作为人工智能软件公司，商汤科技以“坚持原创，让 AI 引领人类进步”为使命，“以人工智能实现物理世界和数字世界的连接，促进社会生产力可持续发展，并为人们带来更好的虚实结合生活体验”为愿景，旨在持续引领人工智能前沿研究，持续打造更具拓展性更普惠的人工智能软件平台，推动经济、社会 and 人类的发展，并持续吸引及培养顶尖人才，共同塑造未来。

商汤科技拥有深厚的学术积累，并长期投入于原创技术研究，不断增强行业领先的全栈式人工智能能力，涵盖感知智能、决策智能、智能内容生成和智能内容增强等关键技术领域，同时包含 AI 芯片、AI 传感器及 AI 算力基础设施在内的关键能力。此外，商汤前瞻性打造新型人工智能基础设施——SenseCore 商汤 AI 大装置，打通算力、算法和平台，大幅降低人工智能生产要素价格，实现高效率、低成本、规模化的 AI 创新和落地，进而打通商业价值闭环，解决长尾应用问题，推动人工智能进入工业化发展阶段。

商汤科技业务涵盖智慧商业、智慧城市、智慧生活、智能汽车四大板块，相关产品与解决方案深受客户与合作伙伴好评。

商汤科技现已在香港、上海、北京、深圳、成都、杭州、南平、青岛、三亚、西安、台北、澳门、京都、东京、新加坡、利雅得、阿布扎比、迪拜、吉隆坡、首尔等地设立办公室。另外，商汤科技在泰国、印度尼西亚、菲律宾等国家均有业务。

更多信息，请访问商汤科技网站、微信、微博和领英。

【关于本报告】

商汤科技有限公司（以下简称“商汤”、“公司”或“我们”）主动向社会公众报告公司的可持续发展状况，让全社会了解、监督商汤的可持续发展工作。

商汤向社会发布可持续发展报告以披露商汤的可持续发展理念和实践，促进商汤与利益相关方以及社会公众之间的了解、沟通与互动，实现企业的可持续发展。

本报告参照全球报告倡议组织（Global Reporting Initiative, GRI）《GRI Standards》核心“符合”方案进行编写，为了保证报告的可靠、公正和透明，公司聘请了外部审验机构 BV 对报告进行审验并出具独立的审验报告（拟定）。

作为独立的可持续发展报告，本报告于 2021 年 11 月以中文版发布，分为印刷版和电子版，如需在线浏览或下载本报告，敬请访问：<https://www.sensetime.com/cn/ethics#researchBor>

如对本报告有任何建议和意见，请通过以下方式与商汤联系：

电子邮箱：airesearch@sensetime.com

第一部分 概述

- » 人类伦理的变迁：从“机器学习”到“机器伦理”
- » 文明伦理的求同：“平衡发展”的 AI 伦理观
 - » “可持续发展”原则：发展权是最大的社会责任与人权
 - » “以人为本”原则：济困扶弱，包容尊重
 - » “技术可控”原则：人类负责，敏捷治理

人类文明的 AI 律己与利他

——平衡发展的 AI 伦理观

科学求真，技术向善。中国著名科学家钱学森坦言：“我作为一名中国的科技工作者，活着的目的就是为人民服务。”而英国生物化学家、科技史专家李约瑟异曲同工分享道：“对科学家来说，不可逾越的原则是为人类文明而工作。”人工智能伦理就是研究 AI 为全人类服务。所以人工智能从业者既有责任向全社会科普、推广人工智能原创应用，又有责任联手社会各界共同监督、治理、改进人工智能伦理风险。

人类伦理的变迁：从“机器学习”到“机器伦理”

计算机引发“数据爆炸”，而人工智能把数据转化为知识。摩尔定律带来了一系列“链式反应”，包括通信革命（互联网）、计算革命（云计算）、媒介革命（数字视频），最终的产物是“失控增长”的大数据量——一种全球化的新型生产资料，日夜不停，取之不尽，用之不竭。

生产资料的爆发式增长往往带来生产力的新科技变革。工业革命因为全球各地源源不断的原棉、矿石运输到英国，人工劳动效率的瓶颈倒逼蒸汽机的诞生。互联网、工业互联网、车联网、环境联网的大数据资源指数级增长，人工知识萃取与转化效率瓶颈，倒逼人工智能的诞生。

表 数据“知识化”的自动化进程

数据知识化	工业时代	信息时代	互联网时代	人工智能时代
数据采集	人工	人工	半自动化	自动化
数据存储	人工	半自动化	自动化	自动化
知识提取	人工	人工	半自动化	自动化
优化业务	人工	人工	人工	自动化

新一代人工智能逐步赋能人类感知、认知、决策范式的升级，成为人类科技创新的源动力，另一方面，人类自身就是伦理价值观的载体，所以在经验归纳、推理演绎、灵感猜想中都自然带入价值观。当我们用 AI 辅助创新与决策时，是否也会自然而然地复制自身价值观呢？那是否意味着机器学习全过程的伦理倾向应有人类监督、机器监督呢？例如在贷款审批、资讯推荐、社交网络审核等方面都存在着因文明而异的默认伦理准则。又如根据数十万名人类司机驾车操作学习的自动驾驶模型，就是用大数据替代人类的经验归纳，驾驶行为中的道路安全、礼让行人等不同价值观都对 AI 训练有影响；而让自动驾驶汽车模型跑在虚拟城市、虚拟高速路上进行模拟测试，就是用仿真取代人工推理演绎，好处是大幅提升创新速度、降低试错成本、测试长尾小概率场景，其间也涉及到车辆安全、行人安全、动物安全、环境安全等伦理准则。诚然，人脑是由 860 亿个相互连接的神经元编织的网络组成，每个神经元与平均约 2000 个其他神经元通过突触结构连接，共有 1.5 百兆个连接（ 1.5×10^{14} ），每个突触每秒可执行上百次“计算”，共 1.5 艾次计算 / 秒（ 1.5×10^{18} ），相当于 10 万个 GPU 的计算量，

这个“复杂巨系统”中伦理观的运行机制我们尚无法掌握，但在人类应时刻监督 AI 产品从设计到退役全过程中的价值观倾向与隐藏风险，在人类通过“AI 大装置”探索未知世界的基础科研、产业研发之旅中，坚守伦理初心，持续升级“机器伦理”风险控制机制。

表 人类认知与伦理范式

	归纳	推理	猜想	伦理
人类认知	1. 经验归纳	2. 推理演绎	3. 直觉灵感	4. 人类伦理
机器认知	5. 大数据	6. 仿真模拟	7. 机器猜想	8. 机器伦理

文明伦理的求同：“平衡发展”的 AI 伦理观

在以人工智能为代表的科技创新日趋复杂之时，我们从《三体》中读到“给岁月以文明，而不是给文明以岁月。”即人文原则第一的基础理念。科学家应是人文、科学、艺术三位一体的存在，人类至高的智慧不是发明，而是良知。科学研究不是万能钥匙，人文精神是打开未来世界大门的钥匙。所以我们搜集分析了全球所有的 AI 伦理政策、倡议、标准，将 AI 伦理观分解为三大类，“可持续发展”类覆盖保护环境、保护和平、包容共享、开放协作、社会认知、敏捷治理等发

展性原则，“以人为本”类覆盖保护人权、保护隐私、人类可控、公平无歧视、造福人类等人文关怀原则，“技术可控”类覆盖可验证、可审查、合法性、可信任、可解释、安全可靠、公开透明、负责任等技术责任原则，特定国家、地区、国际组织成为对标方，发布过越多的相关主题的政策、指南、标准，则雷达图中的圆点就越多，总体呈现趋势如下。

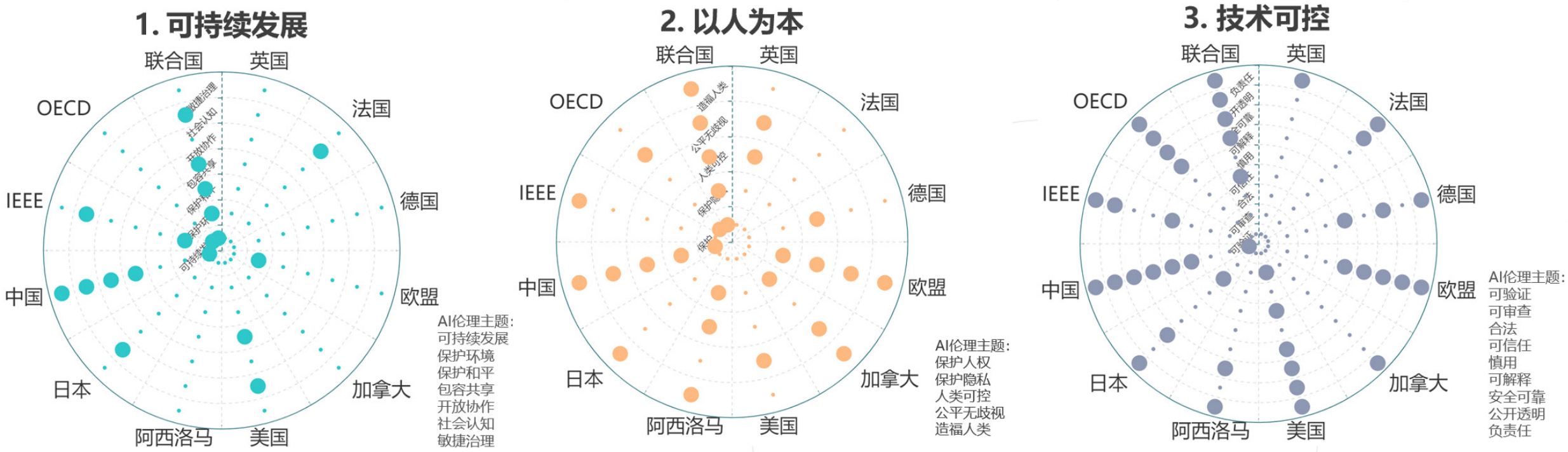


图 全球 AI 伦理观一览

首先，AI 伦理应与人类价值观统一。图灵奖得主杨立昆在《科学之路》一书中所说：“（AI 伦理）是一个机器的价值与人类普遍价值一致性的问题。”过去数千年，人类将道德价值体系编纂入法律，通过教育孩子明辨是非规范社会伦理，现在人类将道德价值观从日常生活抽象到人工智能算法、数据、算力中，并不可能短时间内穷尽所有长尾场景，所以采用案例法为基础的敏捷治理原则持续补充完善 AI 伦理原则体系，以使机器伦理更符合人类普世伦理。所以纵览全球各地区、各个机构的 AI 伦理政策、倡议、标准、法规，能够发现“可持续发展”、“以人为本”、“技术可控”形成了与人类价值观相符的 AI 伦理观。

其次，人类共同发展成为“普世价值观”。哈佛大学教授塞缪尔·亨廷顿在《文明的冲突与世界秩序的重建》中，将当今世界划分为八大文明板块：中华文明、日本文明、印度文明、伊斯兰文明、东正教文明、西方文明、拉丁美洲文明、非洲文明。2020 年以来，八大文明为代表的全人类都面对异常严峻的共同挑战，如极端高温、野火、洪水、飓风、新冠疫情等天灾，唯有“大家一起发展才是真发展，可持续发展才是好发展”，遵循“创新、协调、绿色、开放、共享”的共同发展理念，世界大同，和合共生，用“天下观”超越“文明观”，用“人类福祉”超越“国别利益”，各国携手承担起“人类命运共同体”的发展使命与创新责任，平衡发展的 AI 伦理观才能最大限度地体现全人类的普世价值诉求。

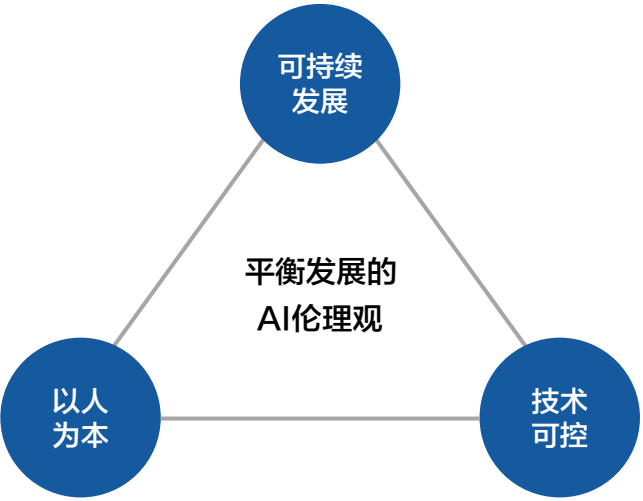


图 平衡发展的 AI 伦理观

最后，平衡、包容的 AI 伦理观，是对人类文明存续发展的关键驱动力。经济实力的壮大确实能够极大增强区域文明的自信心和“世界以我为中心”的错觉，但人类历史上从不是只有一条文明之河，而是诸多差异化很大的文明百川入海。英国著名历史学家汤因比在《历史研究》中陈述：“道德是与意识同时出现在生物圈中，两者共同构成一种存在形式，即精神形式。”和“即使是发达的文明人，仍然像原始人一样是他自己时代和地域的囚犯。”所以，塞缪尔·亨廷顿将全球划分为中华文明、日本文明、印度文明、伊斯兰文明、东正教文明、西方文明、拉丁美洲文明，每个文明社会的伦理观来自与众不同的人文历史，以宗教、文化、人口、经济、语言融为一体的文明多样性，带来了多元化的社会伦理观。从社会伦理映射到 AI 伦理观，则顺势形成了文明板块中各国因地制宜、自主平衡的差异化科技人文标准，尤以“可持续发展”、“以人为本”、“技术可控”为最受各文明板块关注的高优先级因素，驱动人类文明走上一条创新、协调、绿色、开放、共享的可持续发展道路。

01 “可持续发展”原则：发展权是最大的社会责任

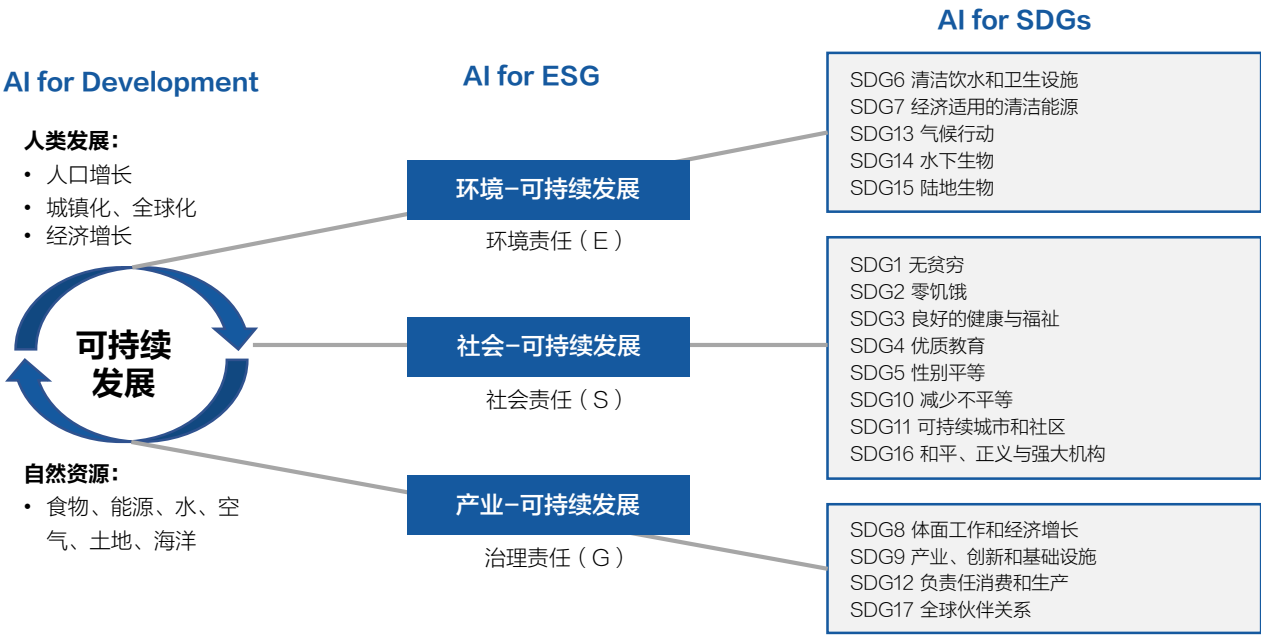


图 AI 可持续发展体系

首先，产业、社会、环境的可持续发展是最大的企业社会责任。19 世纪，美国地理与生态学家马什在《人与自然》书中提出：“地球不仅影响人，人也在影响地球。人们应在为自身以及子孙后代利益考虑的基础上，停止对自然的过度破坏，保护自然，改造自然，让生活更美好。”拉开了“可持续发展”理念和实践的大幕。2019 年，联合国人权理事会通过决议并发布《发展权利宣言》，确认发展权利是一项不可剥夺的人权。例如中国解决了 14 亿人的温饱，使 8.5 亿人脱贫，为 7.7 亿人提供就业，将广大人民的幸福生活“生存权与发展权”放在首位，并向全球各国开放市场、共享科技等生产力，带动众多发展中国家通过新一代基础设施和数字经济生态，实现人类命运共同体的共同发展、共同富裕，取得了世界人民举世瞩目的成就。当经各国已经取得阶段性共识，商业文明功在当代，生态文明利在千秋、惠及子孙，每一个企业、政府、个人都是人类命运共同体的组成部分，都肩负着产业可持续发展责任、社会可持续发展责任、环境可持续发展责任，密不可分三类责任组成为国家治理、科技伦理、产业发展的基石，也称为科技创新企业和科技用户企业的社会责任(ESG)。

其次，科技创新是人类可持续发展的“金钥匙”，人类生存发展面对的共同挑战加速人工智能解决方案惠及全球。地球面临生存危机，人类可持续发展目标将全世界国家凝聚在一起，加速科技创新努力平衡、改进人类无限的发展需求和有限的地球生态资源之间的根本矛盾。“熊皮特奖”得主布莱恩·阿瑟在《技术的本质》中说：“技术是让我们这个世界发生重大变革的第一驱动力。”当今世界，人类利用风光氢储等新能源取代石油应对全球气候灾难，通过人工智能检测新冠病毒探索特效药，研发机器人、脑机接口帮助弱势群体，用 5G 物联网和 AI 遥感技术保护自然资源和野生动物……

联合国 17 个可持续发展目标指引各国经济发展模式、科技创新方向、企业社会责任。2021 年 4 月 22 日，第 52 个世界地球日上，联合国秘书长安东尼奥·古特雷斯呼吁国际社会以从新冠肺炎疫情的恢复为契机，推动世界走上一条更清洁、更绿色、更可持续的发展道路。2021 年 6 月，为了激活人工智能的巨大潜力、加速可持续发展目标的实现、治理科技风险，联合国经济社会局 (DESA) 与联合国秘书处联合发布《人工智能战略资源指南》英文版，系统性提出国际 AI 伦理呼吁、技术治理标准、各国 AI 战略资源与创新案例。

最后，可持续发展依赖国际性包容性数字经济共同体。2019 年 6 月，联合国《数字相互依存的时代》报告中提议国际社会共同努力优化数字技术的使用和降低风险，提出 5 项建议：(1) 建立一个包容的数字经济和社会；(2) 发展人力和机构能力；(3) 保护人权和人类自主性；(4) 促进数字信任、安全和稳定；(5) 促进全球数字合作。一年后（2020 年 5 月），联合国发布了《数字合作路线图》：“预计人工智能的使用到 2022 年将为全球市场带来近 4 万亿美元的附加值。”但我们仍需显著增强“全球人工智能讨论中发展中国家的代表性和包容性”、汇编全球 160 套原则的人工智能伦理和治理共同平台、公共部门为实现可持续发展目标部署与监督人工智能。同时，在国际范围内的数字能力建设（尤其是发展中国家）、数字信任和安全、人工智能安全性与自主性、互联网空间保护弱势群体、数据保护和隐私权、网络骚扰与内容治理等领域，皆需要全球数字共享架构、分布式共同治理架构、互联网治理论坛组成的数字合作方式，以解决成员国、企业、学术机构、民间社团在人工智能包容性、协调和能力建设方面的全球数字合作目标。

首先，人是可持续发展的主体，AI 伦理以人为本，跨越“发展鸿沟”。据《重构地球》数据显示，在 2030 到 2050 年，地球上将有 90-100 亿人口，需要有营养的食物、干净的饮水、体面平等的就业、摆脱贫困的基本收入、可靠可负担的能源、能抵御风险的基础设施。这意味着，城镇化率达到 68%，人类对食物、能源和水的需求将分别增加 60%、80%、55%。回到现在的世界，全球 8 亿人忍受饥饿，20 亿人营养不良，15 亿人缺少现代化电力，8 亿人饮用不安全水源——用先进科技解决人类生存问题刻不容缓！纵然世界上的多元文明存在着显著差异的不同价值观，但扶贫济困、共同富裕成为超越个别文明的“共同性原则”。

其次，尊重人类基本权利是以人为本的核心，跨越“人权鸿沟”。根据《联合国宪章》和《世界人权宣言》，人工智能应保护和促进妇女、儿童、残疾人、少数人种、灾民、老年人、精神疾病患者、失业者和其他弱势群体的基本权利，并实现隐私保护、避免算法歧视。例如商汤通过 K12 AI 基础教育课程 (SenseStudy AI 教学实验在线平台) 帮助青少年获得科技创新能力与正确的伦理价值观，增强了疫情期的“教育权”；我们研发智能避障眼镜帮助视力障碍者户外行走时获得交通信号、环境障碍物等语音提示，保护残疾人的“自由行动权”。我们用 SenseCare® 胸部 CT 智能临床系统赋能超过 6 个省市的新冠肺炎重点筛查医院实现“秒级自动筛查”，提升大众的“社会健康权”。

最后，充满人文关怀、互相尊重的包容性 AI 伦理观是以人为本原则的关键，跨越“文化鸿沟”。加拿大诺贝尔和平奖得主莱斯特·皮尔逊认为：“一个不同文明必须学会在和平交往中共同生活的时代，相互学习，研究彼此的历史、理想、艺术和文化，丰富彼此的生活。否则，在这个拥挤不堪的窄小世界里，便会出现误解、紧张、冲突和灾难。”AI 伦理虽然从属于社会伦理的子领域，但包容性伦理观却能从影响规模、社会效能上充分体现跨文明伦理共识，所以商汤在国际产品上市前都会进行文化伦理、法律法规的适配优化，并根据全球 AI 伦理动态定期进行风险评估与改进措施。

02 “以人为本”原则：济困扶弱，包容尊重

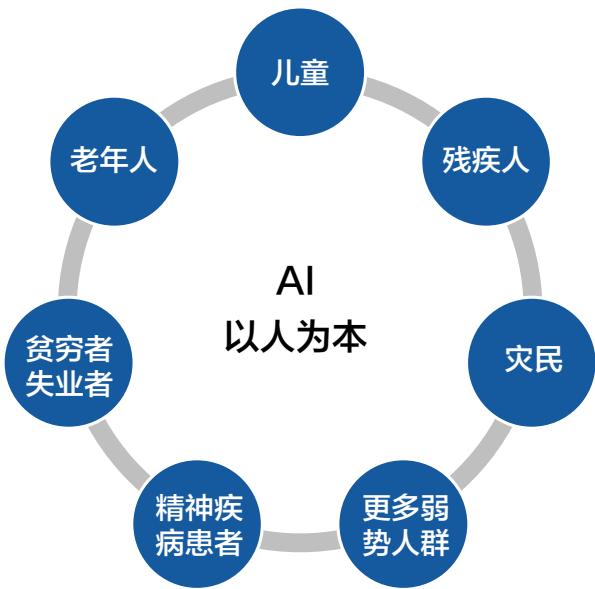


图 AI 以人为本

03 “技术可控”原则：人类负责，敏捷治理

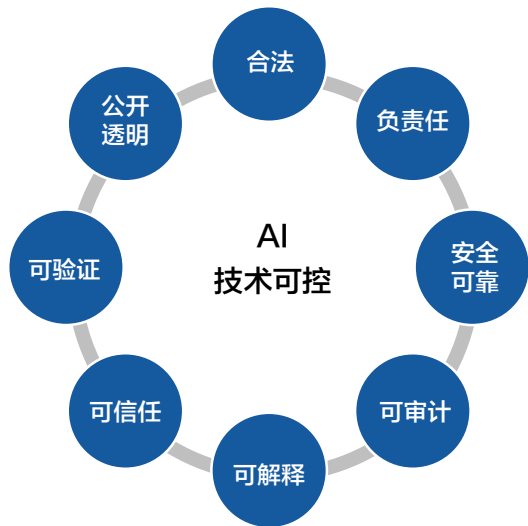


图 AI 技术可控

首先，人工智能是由人类开发并为人类服务的超级工具，所以 AI 应完全受人类控制（技术可控），科技伦理责任由其控制者（人类）承担。自古以来，由于工具是无意识、无自主性的，所以工具的伦理责任必然落在人类身上，包括人工智能的研发者、使用者、监管者等。例如研发方给予不完整数据训练集，就会因数据偏差导致算法歧视，或者因为应用场景滥用和安全防护不足导致敏感数据收集和泄露。亚当·斯密在《道德情操论》中坚信：“人性之尽善尽美，就在于多为他人着想而少为自己着想，就在于克制我们的自私心，同时放纵我们的仁慈心。”

其次，能解决问题的人工智能，应采用多元治理、发展式治理，以释放科技造福人类的创新发展潜力。诚然没有 AI 就没有 AI 伦理风险，但治理的目的不是消灭领先的科技第一生产力，而是采用“促进发展、保障底线”的发展式、多元、敏捷治理模式，以新思路解决新风险与新问题。虽然人类对机器学习的充分理解仍需时间，但判定 AI 应用结果的善恶却十分容易，即符合普世的社会伦理观，即便非技术领域的人文学者、普罗大众用户都能进行多元监督、多元自治自律。“中国航天之父”钱学森常说：“Knowledge was boundless, nothing is final！”（人类的知识认知永无止境）所以人类对科技的认知是持续进步与自我颠覆的，但这并不意味着放弃科技创新的成果，只要新科技能够拯救人类于危难之时、水火之中，人类就应该及时应用，并采用广泛观察、敏捷治理的方式予以持续改进。例如 1897 年首次合成的阿斯匹林每年挽救数百万人的生命，但直到 1971 年人类才明确其作用机制；又如人类到现在也无法完全理解人脑与人体的本质运行逻辑，但并不妨碍 2000 多年来用中医治病救人；伯努利定理不能完全解释飞机机翼的升力成因，但每年超过 40 亿人乘坐航班旅游和出差。敏捷治理能够加速科研创新，视觉神经科学启发了卷积网络，而卷积网络又阐释了视觉皮层的功能，人工智能的“学习机”能力已经落地各种前沿科研领域。商汤以原创 AI 技术（基于图神经网络的蛋白相互作用分析框架、混合图卷积网络模型 DeepCDR、基于病理细胞图像的基因突变预测、药物重定向分析框架）加速新药研发、上市流程，缩短新药研发周期、降低药品研发成本，准确预测癌症药物反应。

最后，实现人工智能的法律合规、主动保护数据安全，是可信任 AI 的关键成功因素。第一，通过自主合规驱动产品升级，通过匿名化、去标识化保证符合《个人信息保护法》，在 AI 所有产品中保障个人敏感信息处理获得授权、审慎性使用、最小化存储、隐私数据的可删除性。第二，用风控优化产品设计，通过收集全球 AI 产品负面案例库，补充风控检查原则，定期对全部产品进行风险评估与改进，比如在数据采集时应用“隐私默认设计”机制，采用不可逆的数据加密技术，实现数据安全。第三，实现多元 AI 治理模式，邀请产学研用所有利益相关方参与研讨会、评审会，针对社会应用场景中的向善性、无偏性跨领域评估，提供全视角的敏捷治理思路与方法。AI 日渐成为各学科科研与各行各业的创新基础设施，审慎负责任的人工智能才能获得社会广泛信任与接纳。

第二部分

可持续发展实践

- » 商汤实践：智能电网，“智能 +” 引领绿色能源革命
- » 商汤实践：智慧医疗普惠全球
- » 商汤实践：推广人工智能基础教育



联合国可持续发展目标旨在到 2030 年解决社会、经济和环境三个维度的发展问题，走上可持续发展道路。可持续发展目标呼吁“所有国家（不论该国是贫穷、富裕还是中等收入）行动起来，在促进经济繁荣的同时保护地球”。该目标指出，应长期致力于消除贫困、实现良好健康与福祉、实现优质教育、性别平等、清洁饮水和卫生设施、经济适用的清洁能源、创新和基础设施、减少不平等、可持续城市、气候变化和保护环境等。随着人工智能技术的发展，人们越来越意识到，人工智能在这些领域全都大有可为。

2020 年 6 月 12 日发布的《联合国秘书长数字合作路线图》报告中指出，应以最佳方式部署人工智能支持实现可持续发展目标，并造福公众。AI 可持续发展的理念来源于联合国 17 个变革世界的目标，换言之，实现 AI 的可持续发展能够实现经济目标之外的社会价值，有助于实现共筑人类命运共同体的愿景。在这一愿景下，商汤积极推进可持续发展框架进程，以实际行动推动 AI 可持续发展的各项原则在不同细分产业的落地，履行了其作为 AI 行业领导者的社会责任和使命。

以下，我们聚焦能源电网、医疗和教育三大行业，来探讨以人工智能为代表的信息技术如何赋能产业数字化转型，从而推动经济增长的良性循环。

01 商汤实践：智能电网，“智能 +” 引领绿色能源革命

7

经济适用的清洁能源



SDG7

可负担的清洁能源

电力工业是全球最主要的 CO2 排放源（45% 的 CO2 来自电力生产）。因此，实现电力减排和清洁生产，降低电力输送损耗，全面优化电力生产、输送、消费全过程，将有助于推动低碳电力、低碳能源乃至低碳经济的发展。在此过程中，智能电网在推动电力清洁生产，促进电力高效利用以及保障可靠电力供应等方面将发挥重要作用，已经成为世界电网发展的必然趋势。

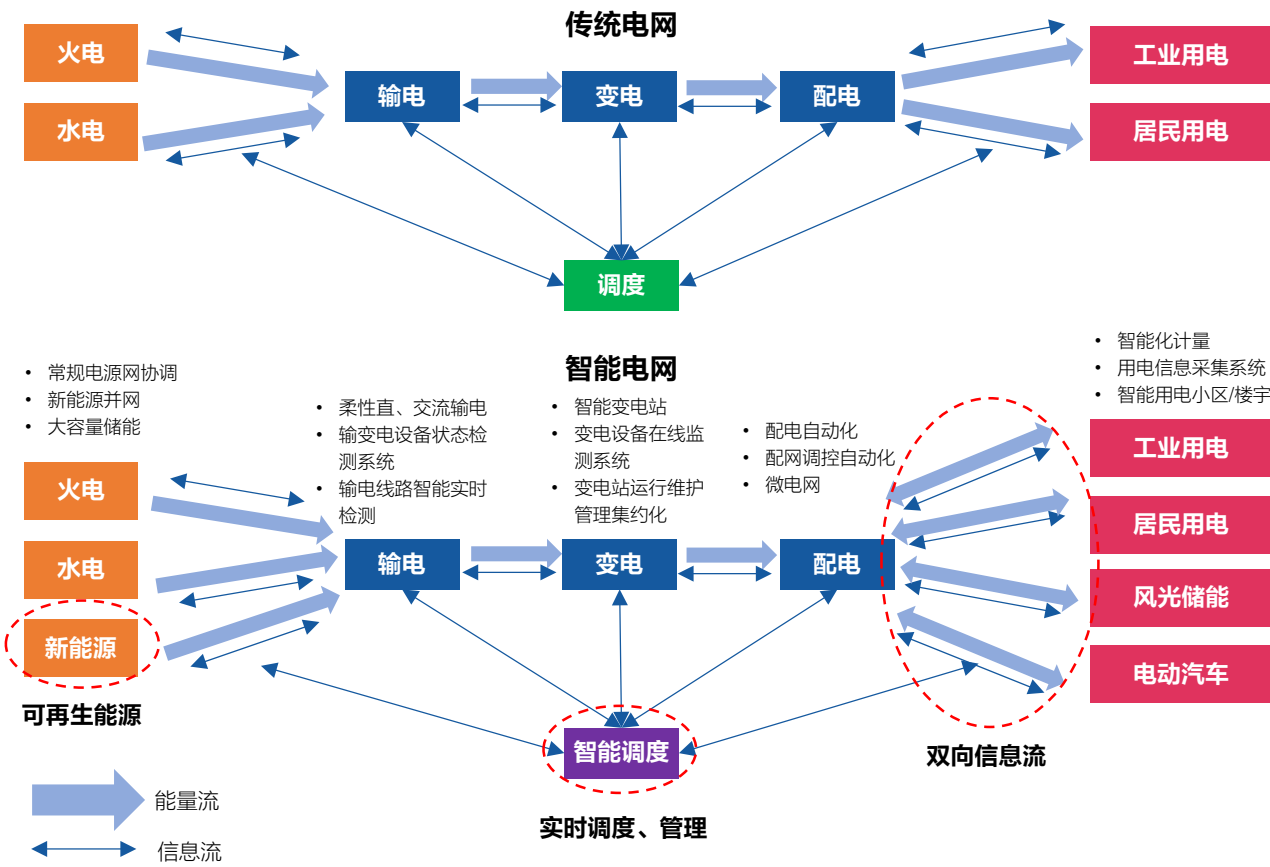


图 传统电网 VS 智能电网，商汤智能产业研究院整理

智能电网 AR 巡检——电网稳定运行

本轮能源变革的目标是通过科技创新，实现以低碳能源为核心的低碳经济。目前电力工业是全球最主要的 CO2 排放源（45% 的 CO2 来自电力生产）。因此，实现电力减排和清洁生产，降低电力输送损耗，全面优化电力生产、输送、消费全过程，将有助于推动低碳电力、低碳能源乃至低碳经济的发展。在此过程中，智能电网在推动电力清洁生产，促进电力高效利用以及保障可靠电力供应等方面将发挥重要作用，已经成为世界电网发展的必然趋势。

全球资源供应紧张，环境恶化等问题正在加剧，在我国郑重提出争取 2060 年前实现“碳中和”目标后，以风光电为主的可再生能源成为未来重点能源将成为主要趋势。从电源侧来看，风光电等新能源具有随机性、间歇性、波动性特征，大规模并网后，电力系统“双高”“双峰”的特性非常明显，电网安全稳定运行和电力电量平衡将面临极大考验。

为保障输电系统安全稳定运行,换流站运维工作尤为重要。商汤以 AI+AR 技术助力电网智能巡检,从而提高电力运维效率、节省时间和人力成本，有效保障在新能源大规模并网后的电网运行安全。

2020 年商汤科技与南方电网数字电网研究院达成合作，在超高压公司直流工程换流站进行了首个工业级视觉定位 AR 技术试点研究。前期商汤录入大量现场数据完成信息采集过程生成电站虚拟环境地图，在后期关联设备运行信息以及叠加外部环境信息后，通过识别现场环境便能对巡检人员实现精准定位，帮助其确定最优巡检路线，避免误入电站危险区域，既节省巡检时间，也从源头上保障巡检人员的安全。在巡检过程中，借助智能装备（AR 眼镜、平板电脑等），巡检人员可全程自动记录作业过程和巡检内容，借助 AI 技术后台系统自动分析并一键完成巡检报告，减轻巡检人员的工作负担，实现对换流站设备运维作业的智能辅助和管控。除此之外，AR 驱动将一线人员生产区域的设备及环境完整“复刻”到电脑上，技术人员或电网专家便可以在电脑前标记，将三维空间物体标识和检修说明推送到巡检人员的 AR 眼镜上，协同快速解决现场作业难题，完全突破空间限制，让技术交流更加顺畅便捷。

02 商汤实践：智慧医疗普惠全球

在医疗领域，商汤 SenseCare® AI 辅助诊疗系统在提升医疗工作者工作效率的同时帮助遏制新冠疫情的扩散，为实现联合国健康与福祉的可持续发展目标做出了贡献。

3

良好健康与福祉



SDG3

良好健康与福祉

导言：
联合国可持续发展《目标和具体分析》指出，企业可以提供创新的解决办法，让更多边远地区人口获得更高质量的健康服务实现 SDG3: 良好健康与福祉。当前，医疗资源分配不均问题是医疗领域面临的主要问题。商汤通过 AI 赋能医疗，借助 SenseCare® 智慧诊疗平台减少医生的机械性诊疗工作并降低漏检、误检率，此外还通过智慧医院运营平台降低病人的就诊成本，助力可持续发展目标实现。

2019 年 5 月 19 日

商汤科技推出 SenseCare 智慧诊疗平台

商汤科技 SenseCare 智慧诊疗平台具备高并发三维渲染能力，单台设备能支持超过 160 位医生同时进行高质量的三维渲染机交互，并供其进行便捷且实时的交互式分析，能让影像科及各临床方向医生更为便捷地进行诊、疗、愈全栈操作。借助商汤原创的深度学习平台，SenseCare 还可针对不同科室的需求和工作流灵活拓展含有丰富 AI 算法模块的临床应用，以满足不同临床方向诊疗愈的智能辅助。在 2019 商汤人工智能峰会上，上海第九人民医院的专家以骨肿瘤为例，介绍其借助商汤 SenseCare 智慧诊疗平台，极大地提高了骨肿瘤 3D 打印治疗规划的效率、精度和效果。

2019 年 10 月

商汤与多方联合举办病理、放疗两大 MICCAI 国际挑战赛

国际顶级医学图像学术会议 MICCAI 2019 如期在中国深圳举办。商汤科技有七篇论文入选主会，同时，作为大会重磅活动之一，由商汤科技联合多方举办的 MICCAI 2019 消化道病理图像检测与分割、MICCAI 2019 放疗规划自动结构勾画两项国际挑战赛亦完成激烈角逐，并面向研究领域开放了大量专家支持的高质量标注脱敏数据和专家知识，包括大量具有专家标注的放疗规划 CT 数据以及行业第一个公开的消化系统病理图像数据集。

2019 年

RSNA（北美放射学会）大会 SenseCare 智慧诊疗平台收获多方赞誉

在 RSNA 2019 大会期间，商汤科技与医疗行业老牌强者 GPS（通用电气、飞利浦、西门子）同台争艳，展示了中国人工智能企业的实力。来自斯坦福大学、约翰霍普金斯大学、宾夕法尼亚大学等诸多美国顶尖高校教授和影像专家，和上海第九人民医院、北京大学第三医院等国内医院专家，以及三星、GE、西门子、日立、岛津等企业代表纷纷到访商汤科技展台进行深入交流。计算机视觉和医学图像领域的国际顶级专家 Dimitris Metaxas 教授及其放射领域合作伙伴们，也点赞 SenseCare 智慧诊疗平台：“SenseCare 不仅为影像科提供了很有效的工具，同时也为影像科和临床科室医生之间搭建起沟通的桥梁，让影像的价值得到了充分的延伸。”

2020 年 7 月 20 日

商汤 SenseCare 探索“联邦学习”入选计算机视觉顶会 ECCV

由于涉及隐私等问题，全球各国都针对医疗数据制定了相关的保护政策，使得多中心数据共同训练变得愈发困难，而这又是医疗 AI 模型开发迭代必须的步骤。联邦学习是一种分布式的机器学习方法，可以在不共享数据的情况下对多中心的数据进行联合建模，从技术上实现了保证数据安全的同时实现协作。依托深厚的学术研究底蕴以及对行业趋势的敏锐洞察，商汤携手美国罗格斯大学计算机系计算生物医学成像和建模研究中心，在全球顶级计算机视觉会议 ECCV（European Conference on Computer Vision，欧洲计算机视觉国际会议）上发表了一项新的研究成果，创新地使用基于分布式生成对抗网络（GAN）的结构来实现联邦学习，为打通 AI 医疗应用的最后一公里开辟出一条“新的道路”。

03 商汤实践：推广人工智能基础教育

在教育领域，商汤通过研发一系列教育平台产品及推行教师培训项目有效降低了 AI 数字鸿沟，为我国人工智能产业输送了一大批创新型人才，使更多孩子享受到平等的 AI 教育机会。



引言：
 秉承普及人工智能相关的科学知识、弘扬科学精神、传播科学思想、倡导科学方法的信念，商汤建立起一套纵向覆盖从小学到本科、横向兼顾“基础、拓展、科创”的 AI 教学体系，陆续出版了 AI 教材 11 册，并结合实验课程及机器人教具等 AI 实验配套设施辅助，构建理论与实践结合、衔接贯通的人工智能教学体系。同时，通过师资培训、科创竞赛等方式，帮助学校从 0 到 1 培训自己的 AI 教师队伍，激发学生对 AI 的兴趣，帮助国家培养和选拔面向未来的创新型 AI 人才。

截至目前，商汤已将 AI 基础教育推广至全国 30 多个城市，2700 多所中小学，共培训了 7200 多名一线任课教师。未来，随着商汤科技与更多城市和院校的深化合作，AI 基础教育将在更多地方形成燎原之势，更多孩子享受平等的 AI 教育机会。



全国首个地市级人工智能高中教育基地在山西晋中落地

2018 年 8 月 13 日，全国首个地市级人工智能高中教育基地在山西晋中落地，教育部课程中心陈云龙处长见证了本次签约仪式。他充分肯定了人工智能的重要性，并提出晋中市人工智能教育基地的落成对人工智能进入高中课程起到了示范及引领效应，创造了人工智能均衡教育的新典范。晋中市教育局党组书记、局长张海荣发表讲话并指出，此次晋中教育与商汤集团的战略合作，要在全市 35 所高中学校开设人工智能课程，打造全国首个市级人工智能教学基地。让晋中成为商汤集团中小学人工智能选修课程开设、产品开发应用、智慧教育建设的“试验良田”，把晋中打造成全国有影响、有地位、有成效的市级人工智能教学和人才培养高地。在“晋中模式”的示范引领下，商汤科技以《人工智能基础（高中版）》教材为基础，通过人工智能教师培训、课程内容统一部署与落地、实践项目等全方位布局，形成良好的人工智能基础教育体系，推动晋中人才高质量发展。

相关 SDGs：SDG1– 无贫穷；SDG10– 减少不平等

商汤发布《AI 教育白皮书：智能时代的教育变革与创新实践》为

回答人工智能教育面临的诸多严峻问题，商汤科技于 2020 年 12 月 16 日正式发布《AI 教育白皮书：智能时代的教育变革与创新实践》（以下简称白皮书）。《白皮书》回应了智能时代 AI 教育的新命题，从商汤教育落地山东青岛、山西晋中案例入手，剖析传统信息技术教育与人工智能教育的核心差异，探讨“普惠而均衡的 AI 教育”理念，呼吁建设开放共享的 AI 教育体系。

相关 SDGs：SDG1– 无贫穷；SDG10– 减少不平等

全球第一所商汤人工智能教育研究院落户青岛

2019 年 3 月 29 日，商汤科技全球布局的第一所人工智能教育研究院正式落户青岛。青岛教育局、崂山区政府分别与商汤科技签署《合作备忘录》，确定在“人工智能+教育”领域展开全面合作。2019 年 9 月起全市在 100 所人工智能试点中小学，至少 2 万名学生中全面开展人工智能教育。2019 年累计开展 10 场教师培训活动，参训教师 468 名；2020 年 8 月累计开展 3 场线上教师培训活动，参训教师 451 名。通过培养，青岛地区涌现出 67 名优秀骨干教师，积极发挥辐射带头作用，引领 AI 课程的建设。

上海市徐汇区中小学开始试点实施“人工智能”基础型课程

2019 年 10 月，在徐汇区 8 所高中、12 所初中和 10 所小学开始试点实施“人工智能”基础型课程。2020 年徐汇区全面开始人工智能基础型课程实施。依据国家版本的普通高中《信息技术课程标准》，商汤科技联合徐汇区教育局一起设计并编制了《徐汇区中小学人工智能课程实施意见》和《徐汇区中小学人工智能课程教学基本要求》。2019 年 4 月徐汇区举办了本区教师第一次人工智能课程培训。

上海市卢湾高级中学正式挂牌成为商汤科技实验学校

作为上海市实验性示范性高中，上海市卢湾高级中学以科学教育办学为特色，促进信息化时代人工智能教育课程的开发和学习方式的转变。2019 年正式挂牌成为商汤科技实验学校，共同打造人工智能标杆校，并设立人工智能奖学金、奖教金，奖励在人工智能教育及学习上表现优异的师生。学校已建设人工智能教学平台，普及推广人工智能课程；同时搭建科创平台，培养人工智能科创人才。在科学系统的科创辅导下，学校学生荣获第十八届“明日科技之星”称号。

上海交通大学附属中学与商汤合作建设人工智能实验室

作为全国百强校，上海高中四大名校之一，上海交通大学附属中学突出科技创新特色，创建“拔尖创新人才培养基地”，为了培养学生人工智能基础技能与意识、科技创新流程与技巧等能力，2020 年与商汤合作建设人工智能实验室，制定了科创课程内容和完整的授课计划。同年年底，在第三届上海市青少年人工智能挑战赛上，通过商汤科创辅导的 12 位同学参加了 AI 研究成果和 AI 大挑战两个比赛项目，斩获 12 个奖项（全市共 25 人获奖），含 3 个一等奖、4 个二等奖、5 个三等奖。

江西赣州章贡区教育扶贫捐助

2020 年 9 月 7 日，商汤科技向章贡区政府捐赠了支持全区开展 AI 教育的基础设施和服务资源，推动 AI 基础教育在章贡的落地和普及，培养当地师生 AI 素养，实现 AI 基础教育的普惠和均衡发展。通过此次项目，商汤将有力帮助章贡区进一步提升教育水平，助力脱贫攻坚工作，为这一区域打造增长新动能。

相关 SDGs：SDG1– 无贫穷

2019 年开始连续三年举办国际青少年人工智能交流展示会（IAIF）

为了让同学们把在学校学习到的人工智能知识和 AI 创新精神学以致用，商汤教育于 2019 年开始连续三年举办国际青少年人工智能交流展示会，三届大赛累计下来的学生参与者达 4000 余人，覆盖到全球 14 个国家和地区，以及全国 21 个省份。

第三部分

以人为本实践

- » 商汤实践：商汤智慧医疗 SenseCare 平台，赋能前线医生
- » 商汤实践：赋能智慧文旅，迎接元宇宙世界到来
- » 商汤实践：智慧乘客服务平台，赋能成都城市地铁交通系统

可持续发展思想强调社会发展是以人为本的综合发展。它在促进人与社会、环境、经济、政治、文化协调发展的过程中，使人成为自然同时也成为社会的真正主人。可持续发展思想认为社会发展本质上是人多种价值要求的实现过程和人自身价值的实现过程。它要求在社会发展过程中，必须培养社会的人的一切属性，使每一个人能够把自己本身当作现有的、活生生的主体来对待。可持续发展教育力求把每个人培养成为自主、和谐与全面发展的人，也就是物质需求和精神需求达到高度统一的人，这正是可持续发展思想的终极价值目标，因此在人工智能发展过程中，也要注重“以人为本”的价值体系。

在“以人为本”的价值体系的引导下，商汤积极推进 AI 可持续发展目标的落实，以实际行动推动以人为本的价值观在不同产业领域的实践，履行其作为 AI 行业领导者的社会责任和愿景使命。

01 商汤实践：商汤智慧医疗 SenseCare 平台，赋能前线医生



图 商汤 AI 医疗影像团队疫情期间驰援各地医院

本次疫情中，肺部 CT 影像是新冠肺炎临床诊断中的重要参考。商汤联合湖北省多家医院，迅速开发出基于 SenseCare® 胸部 CT 智能临床解决方案的新冠肺炎功能升级。该产品凭借领先的深度学习算法辅助影像科医生快速完成对肺部 CT 图像的筛查，自动提示 CT 中的疑似新冠肺炎病灶位置，并在 2-3 秒内即可完成定量分析，助力尽早实施核酸检测、隔离等进一步措施，以实现早发现、早隔离、早治疗，为患者争取宝贵时间、最大程度避免交叉感染。

作为战“疫”最前线，为确保“应收尽收”，湖北省此前曾将“疑似病例具有肺炎影像学特征者”纳入临床诊断病例，这意味着只要患者具备临床表现和肺炎 CT 影像学特征，便被列入新冠肺炎的新增病例，也导致影像医生工作任务的艰巨性在短时间内大幅增加。

为了以最快的速度缓解前线医生的压力，商汤将 SenseCare® 胸部 CT 智能临床解决方案以云服务的模式接入到合作伙伴的医学影像阅片平台，通过远程的方式快速驰援十余家前线医院及多地医院和医疗机构，为医生提供免费的 AI 辅助分析。

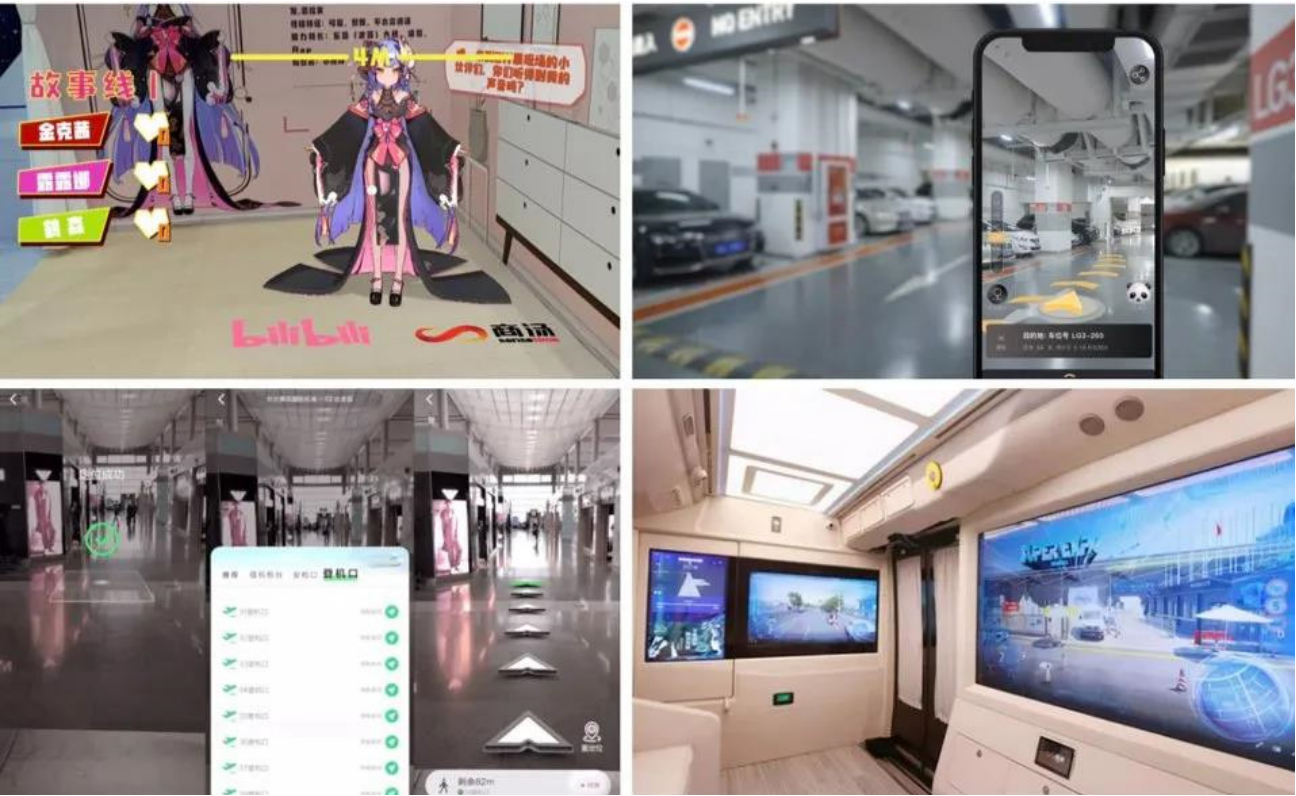
疫情期间，青岛西海岸新区人民医院作为青岛市指定新冠肺炎患者收治单位，肩负着重要的筛查工作职责。商汤智慧医疗团队在 2020 年 2 月疫情最紧张的阶段，紧急协调服务器资源，克服大雪天气、物流不畅等重重困难连夜赶往医院进行装机，确保了医生第二天即使用上有迹可循、有据可依的 AI 智能辅助。

在各地人员开始返程复工期间，青岛西海岸新区人民医院曾接到一个批次就 120 人的复工 CT 体检任务，在 AI 的辅助下，医生仅仅用了 20 分钟就把诊断报告全部发出，极大提升了复工筛查效率，避免人员长时间滞留医院，降低交叉感染风险。

02 商汤实践：赋能智慧文旅，迎接元宇宙世界到来

旅游产业在疫情期间受到严重冲击，但从过往数据和长远发展看，旅游早已成为中国居民主要的闲余活动和生活方式，并在消费升级的引领下迎来新的发展机遇。据文旅部统计，从 2013 年到 2019 年，国内旅游人数从 32.62 亿人次增长到 60.06 亿人次，将近翻了一番，年均增速超过 10%，无论是出游规模还是出游人次，我国都已进入了大众旅游时代。

尽管各类 APP 在旅游支付、行程管理上提供了使用便捷，但供给侧的旅行配套服务不到位、景区大颗粒度式管理，以及景观内容缺乏创新、千篇一律，与实际消费需求之间严重脱节，是现阶段掣肘旅游产业发展的主要矛盾，“强管理、优服务、重体验、多创新”是旅游业转型的主要目标；与此同时，在文旅部推动下，“文化”与“旅游”的跨界融合已成为旅游业供给侧改革的新趋势。



随着 5G、物联网、云计算、人工智能、大数据、AR/VR、高清视频等技术的成熟与应用，新一代信息技术成为促进文旅融合、产业转型的新引擎，以文化为内涵、旅游为载体、技术为动力，推动旅游产业向“智慧文旅”方向升级。

围绕“强管理、优服务、重体验、多创新”目标，智慧文旅的建设与发展主要体现在三个方面：

一是面向文旅企业（To B），提升文旅服务和产品的智慧化。包括 AR 导游、智慧景区、智慧游乐、智慧文娱、智慧文博、虚拟旅游、大数据精准营销等全方位的服务。

二是面向政府和景区（To G），强化景区精细化管理，提升公共服务的智慧化。包括景区动态监测、信息监管、大数据分析、景区人流引导、安全预警等内容，全面了解行业和游客需求、旅游目的地动态、投诉建议等内容，推动传统旅游管理方式向现代管理方式转变，实现政府的科学决策和管理。

三是面向消费者（To C），提升创新文旅体验和消费服务的智慧化。推动文化旅游资源与多样化、个性化消费需求的高效对接，实现精准的线上线下无缝服务和深度文化体验等实时、互动、个性化服务，提升用户体验。

商汤科技便凭借智能感知、智能内容生成和智能内容增强等原创 AI 技术的积累，推出了 SenseME 水星智能移动终端平台及 SenseMARS 火星混合现实平台，成为目前最大的元宇宙赋能平台之一。依托 SenseCore 商汤 AI 大装置，SenseMARS 火星混合现实平台目前已涵盖超过 3,500 个模型，拥有城市级大尺度高精三维数字化空间构建、全场域厘米级端云协同空间定位，以及跨平台的终端空间感知计算等驱动空间元宇宙的多种关键能力，可高效构建与现实世界对应的数字孪生世界，实现现实世界和虚拟世界的连接。

在成都 IFS 国际金融中心，商汤科技不仅实现了高精度 AR 导航，更通过导航串联品牌营销活动，促进顾客活跃度，以空间元宇宙为线下零售业态打造了全新的流量入口和营销闭环。

2021 年 7 月，商汤科技还与 Bilibili 合作，在其线下嘉年华活动——BilibiliWorld 中共同打造虚拟体验馆，将平日里只出现在弹幕视频里的二维虚拟主播，以鲜活的三维形象来到玩家眼前，与玩家展开跨越次元壁的虚实互动，以空间元宇宙开创了互动娱乐新模式。

商汤还充分发挥跨行业优势，将 AR 与智能驾驶技术融合应用，创新性的打造 SenseAuto AR-Robobus 绝影自动驾驶 AR 小巴，将车窗变为一块屏幕，让乘客在虚实融合的视觉奇景中，一览城市、人文等景观，以空间元宇宙创造园区、景区游览新体验。



成都地铁以中心化方式部署服务器，能极大降低运维成本，但也给算法和网络传输带来巨大挑战。对此，SenseFoundry Traffic 商汤睿途智慧交通平台创新采用异构并行计算分布式架构，基于支付级百万分之一误识率的超高精度原创算法，通过各站点常客底库与中心服务器完整底库的并行检索比对方式，实现响应速度的大幅提升，并显著简化了线网部署的复杂度。实际应用中，从信息获取、识别、比对、支付到票务平台的整个流程处理只需不到 200ms，乘客过闸时没有任何延迟感。

闸机终端更采用定制化商汤 SensePass Pro 一体机，在复杂光照、多人通行等场景下，也能精准完成比对和活体检测。同时乘客应用“无感通行”进站时无需取下口罩，并可自动检测口罩佩戴状态，保障疫情常态化防控阶段的乘车安全。目前，市民乘客通过成都地铁 App 已经可自主选择开通此功能，即可享受全线“无感通行”。未来，在 SenseCore 商汤 AI 大装置的支持下，商汤还将针对车站管理中的更多长尾问题，自动化、高效率地量产各类高质量算法，为乘客打造安全与效率并举的乘车体验。

03 商汤实践：智慧乘客服务平台，赋能成都城市地铁交通系统

2020 年 3 月，中共中央政治局常务委员会召开会议，研究决定要推动新型基础设施建设，其中城市轨道交通和人工智能是新基建的重要方向。2020 年 4 月，发改委进一步明确“新基建”概念内涵，将新型基础设施划分为创新基础设施、信息基础设施和融合基础设施三个方面，人工智能赋能地铁是融合基础设施中的重要内容。地铁已经成为城市交通出行不可或缺的方式，随着地铁 规模增大，安全保障、运营效率、服务质量等多方面需求相互交织，都迫切需要打造智慧地铁大脑，为地铁赋能实现智能化管理。

商汤科技携手智元汇共同打造的“智慧乘客服务平台”，涵盖智慧票务、智慧安检等多种系统，依托 SenseFoundry Traffic 商汤睿途智慧交通平台多项原创领先的 AI 技术，为成都地铁票检工作降本增效。智元汇智慧乘客服务平台项目实施负责人蒲磊表示：“依托商汤科技在通用视觉领域的核心技术优势以及分布式架构等软硬件协同能力，平台在乘客体验上带来了突破和创新。未来将与商汤科技不断深化合作，用 AI 技术为智慧轨道交通建设按下‘加速键’。”

第四部分

技术可控实践

- » 商汤实践：数据隐私保护
- » 商汤实践：AI 伦理风控与审核
- » 商汤实践：AI 伦理高校联合研究

当前，随着人工智能技术应用的广度和深度不断拓展，正在成为信息基础设施的重要组成，但在这个过程中也不断暴露出一些风险隐患，包括数据滥用导致的隐私泄露风险、算法安全导致的应用风险、系统决策复杂导致责任主体难以界定等。面对人工智能技术引发的风险，推动可信人工智能技术发展，从而实现人工智能的技术可控成为了全球共识。

我们认为，“技术可控”既指应用层面的合规可控（例如数据隐私保护），更指核心技术的自主可控，即以自主可控为跟进，通过理论创新、技术突破形成核心竞争力是关键。本部分通过讨论数据隐私保护、人工智能训练开发平台的技术可控以及通过创新型技术的研究课题来实现技术可控的案例，介绍了商汤在相关领域的实践如何体现技术可控的原则。

01 商汤实践：数据隐私保护

商汤是第一家获得全部三项 ISO/IEC 隐私信息管理系统、信息安全管理和个人身份信息保护认证的人工智能公司。我们高度重视数据安全及保护，我们已采取标准保护措施，包括保密分频、访问控制、数据加密及脱敏，以防止未经授权访问、泄露、不当使用或修改、损害或丢失数据及个人信息。数据安全及个人信息保护管理委员会监督数据安全及个人信息保护工作。我们已建立全面的个人信息管理系统，并制订一系列技术标准及规范，以确保数据及个人信息于整个生命周期内的安全。

下文陈列我们数据及个人信息保护措施的主要举措：

（1）数据的授权存取及使用

在数据的授权存取及使用方面，我们要求客户确认其已从合法来源获取该数据的使用权且终端用户已同意协议中指定的使用目的。我们仅使用数据作客户明确授权的用途，例如身份核实、记录保存及统计数据，不会将数据用于未经事先批准及同意的用途。我们持续监控与第三方的数据处理合作，并定期审查该等合作的内容、合作协议的范围及该等协议的执行情况，以确保遵守相关法律法规。我们高度重视数据安全及保护，已采取标准保护措施，包括保密分类、访问控制、数据加密及脱敏，以防止未经授权访问、泄露、不当使用或修改、损害或丢失数据及个人数据。

（2）独立数据库及安全服务器系统

我们拥有自己独立的数据库，并且不会与其他第三方共享客户或终端用户的个人信息。我们的服务器系统已受增强的安全级别保护。我们定期对服务器运营推行用户账户审核及监控。一旦发现若干服务器系统出现安全问题，我们将立即升级有关系统，以确保服务器系统及应用的安全。我们已加强系统中敏感个人信息的加密力度，以确保数据保密性。我们拥有全面的个人信息安全及管理系统，涵盖我们数据、源代码、个人信息、第三方人员、网络安全事件及基础设施的安全管理。

（3）全面的数据及个人信息安全管理政策

我们已实施全面的雇员保密政策、数据使用审批程序及数据追踪机制，以确保我们数据库的安全性。我们已根据相关规则及法规制订相应的工作流程。作为数据处理者，我们已实施多项数据保护及网络安全措施，以确保我们妥善处理敏感数据，包括我们用于所有数据培训活动的数据脱敏技术。通过对技术进步的持续投资，我们提高整体安全能力。同时，我们获得信息安全管理体系认证（ISO/IEC 27001:2013）、个人身份信息保护管理体系认证（ISO/IEC 29151:2017）及隐私信息管理系统体系认证（ISO/IEC 27701:2019）等多项认证，重点产品已通过信息安全登记保护评测。根据海外业务的需要，我们已正式启动国际隐私标准 BS10012（对标欧盟 GDPR 个人信息保护法案）的认证工作，预计 2022 年年初获取证书。此外，在产品的设计方面我们高度重视数据安全，且产品在推出或交付给客户前必须通过数据隐私评估及安全测试。

（4）数据平台的隐私处理

数据安全及隐私保护被视为我们训练数据平台的基本要求，我们力求遵循最高标准的信息安全及数据隐私。具体而言，我们通过系统性的访问控制及数据加密来保护数据并通过自动脱敏保护数据隐私。我们隐去个人及敏感信息，确保数据用于训练时不附带任何个人信息，在 DCP 技术平台中，我们开发了数据网关，确保数据进入集群的合规性。针对汽车等特定场景，我们也制订了《汽车传输视频及图像脱敏技术要求与方法》这样的团体标准来实现更好的数据合规。

（5）定期审核及应急计划

我们每年对数据安全合规状况进行专项审核。2021 年度共计现场访谈了 BI 产线、智慧城市产品线、智能驾驶产线、SenseID 产线、互娱产品线的研发 & 测试 & 交付全流程。专项工作组针对高危展项进行了排查，推动了高危展项隐私的整改落地，并制定了展厅展项安全审核流程。我们已制定网络安全应急计划，并每年进行培训及安全演示，为任何紧急网络安全事件做好准备。若安全措施受到影响，我们将根据相关法律法规向主管部门报告，并及时通知受影响的用户。

02 商汤实践：算法 / 芯片评测

商汤通过基于真实场景下的数据集的算法验证与评测，实现 AI 技术的真正可信与可控。算法测评的不仅仅保障了 AI 算法应用的可靠与可信，也使得整个算法实施的结果能够在产品开发的全流程中实现应用的技术可控。通过算法测评实现技术可信可控的人工智能的意义在于将人工智能原则转化为了具体实践，落实到了技术、产品与应用中去，是 AI 企业回应社会关切、防范 AI 伦理风险、解决突出的应用实践矛盾的重要路径和必然选择，也是关系到人工智能技术长远发展的重要工作。

具体来说，一方面算法评测可以帮助我们对 AI 算法的行为与结果进行充分、完整和一致性的测试，保障算法的结果是完全是可以被预测和控制的，从而实现了在结果层面的稳定与可信；另一方面，我们通过专业的测评，保障算法能够被完整、充分和正确的被定义与使用，其定义的规则及其后果可以通过数据集反复验证得到可靠的结果，实现了 AI 技术的可信与可控。

通过算法测评的方式，可以通过真实的数据集全面的对算法真实的应用场景进行测评，围绕“AI 算法的数据隐私保护”、“AI 算法的鲁棒性”、“AI 算法的公平性”等多个方面，对 AI 产品与开放的全流程进行测评，包括不限于数据处理、模型构建、模型部署以及支撑与服务相关的产品流程，通过对人工智能算法研发和使用过的整个生命周期过程中贯彻“技术可控”的要求，从产品源头确保了人工智能算法的可信品质。

除此之外，我们还关注芯片的测评工作，联合产业并联合产业生态上下游形成智算联盟，来更好的实现技术可控的目标，具体举措如下：

（1）针对商业化算法场景的全面评测

通过对主要场景和长尾场景的算法测评的全覆盖，我们实现了所有商业化算法的可信与可控。为了保障算法评测覆盖的全面性，我们将算法评测的对象按大类分为主算法和长尾算法两个部分，主算法是人脸、人体和结构化属性，车辆车牌识别和结构化属性等，长尾算法是指特定场景的专用算法，涉及智慧城市，智慧轨道交通和智慧车站等。其中主算法的测评包括静态场景算法评测与动态场景算法评测，而长尾算法则根据不同的场景进行全面的细节评测。

（2）全方位多数据集的算法评测

为了保障算法能够被完整、正确与充分的使用，在测评过程中我们通过足够充分的数据集以及全方面的评测方案对算法进行充分的评测。主算法的评测以图片解析为主，会按照学术的精准率、召回率、准确率、F 值计算；长尾算法大多数以事件检测为主，使用带有事件的视频，有的是以算法精度测试有漏报和误报来计算准确率的指标，有的是以事件发生的时间区间维度的准确率指标，也有按照视频内目标对象的事件维度准确率指标。以 SenseCore 大装置为例，其支持 108 类算法的部署，87 类不同的精度评测方案以及 1279 个测试数据集。

（3）推动算法测评的成熟与创新

随着算法测评的工作实践逐步成熟，我们逐步推动行业算法测评标准的形成，以推动人工智能技术在普遍意义上的可信与可控。我们将目前所使用的主算法评测的方式与实践提炼出来，打造了一系列行业标准。而在长尾场景的测评中，我们所提出很多创新的测试方法。例如使用最先进的逐帧标注方法，目前正在申请 8 项基于事件的算法测评专利，而业界目前没有从这个角度去做任何创新的测评方法，商汤首先在产业实践中去应用了该方法提升算法测评的可靠性。

（4）组建智算联盟，推动多家厂商的芯片评测落实

2021 年 7 月 9 日，商汤科技联合中国电子技术标准化研究院，中国信息通信研究院华东分院、清华大学、上海交通大学、复旦大学，及华为、寒武纪、燧原科技等多个行业伙伴成立人工智能算力产业生态联盟（智算联盟），旨在不断探索 AI 算力创新应用，推动国产 AI 算力的落地应用，建立自主可控的 AI 算力开发生态，促进中国自主 AI 算力的可持续发展。

目前我们基本完成寒武纪和海光芯片的适配，并且开始规模化采购和推进商业化算力。正在进行中的华为、天数智芯、燧原等自主可控的 AI 芯片的摸底测试。另外，积极的推进自主可控 AI 芯片的编译生态，推出长尾算子的 benchmark，以智算联盟为基础，探索自主可控的编译框架。

（5）建立联合实验室，加强芯片测评工作体系化进程

工信部标准化研究院和商汤科技合作共建“CESI-SenseTime 人工智能算力及芯片评测联合实验室”，共同开展人工智能芯片评测方法与相关标准研究，促进人工智能处理器评测技术及标准化等领域的推广，分享标准化领域的科研成果转化经验，进一步推动各方在人工智能芯片相关方面的标准研究和制定。

03 商汤实践：科研课题研究

为了保障 AI 技术的自主可控，需要在前瞻性技术方向进行科技研究。商汤科技自 2019 年便开始了对联邦学习的前瞻性研究，并携手美国罗格斯大学计算机系计算生物医学成像和建模研究中心，在全球顶级计算机视觉会议 ECCV（欧洲计算机视觉国际会议）上发表了一项新的研究成果，创新地使用基于分布式生成对抗网络（GAN）的结构来实现联邦学习，为打通 AI 医疗应用的最后一公里开辟出一条“新的道路”。

该研究通过将位于多个数据孤岛的分布式异步鉴别器和一个中心生成器组成对抗网络，让中心生成器在不接触原始隐私数据的情况下，也能进行合成训练，从而能够生成与各数据孤岛中原始数据相近似的合成数据样本，供下游任务使用。由于避免了对原始数据的直接访问，这套研究方法秉承了联邦学习的核心优势，很好地解决了医疗数据的隐私保护问题。

不仅如此，相较于传统的联邦学习，由于在实现方法上“另辟蹊径”，该研究成果还有效减少了中心与数据孤岛之间的通信数据量，仅需传输合成图像数据和反馈误差而非整个模型的所有参数数据，而且各数据孤岛之间无需交换任何数据或参数，因而可显著降低医疗机构之间通过联邦学习进行研究的成本，加快研究效率和 AI 模型的生产速度。

此外，这种创新、低成本的联邦学习模式还可以推动低效率的分散型数据中心向高效的集约型数据网络升级，从而更好地助力地区性数据中心或行业标准数据库的建立。这正契合了国家“新基建”的战略指导方针，加速“数据智能”的基础设施建设，为医疗乃至更多行业节约成本，创造价值。

除了上述研究，我们还与上海清源研究院共同完成人工智能医学图像的可解释的神经网络的课题，目前这两个联合研究都发表了专业的论文，成为重要的学术成果，具体如下：

（1）2020 年入选 ECCV 论文一篇，主题是创新地使用基于分布式生成对抗网络（GAN）的结构来实现联邦学习，在保障数据隐私和数据安全的前提下去训练模型；

（2）2021 年入选 TMI 学术论文一篇，主要是针对人工智能医学图像的可解释的神经网络的课题，用于对医疗领域的 AI 算法的可解释性进行研究。

合作方	论文主题	入选刊物
美国格罗斯大学计算机系	创新地使用基于分布式生成对抗网络（GAN）的结构来实现联邦学习	European Conference on Computer Vision
上海交大清源研究院	人工智能医学图像领域可解释性的神经网络研究	IEEE Transactions on Medical Imaging (TMI)

第五部分

AI 伦理与治理实践

- » 商汤实践：组建 AI 伦理与治理委员会
- » 商汤实践：人工智能伦理与治理风险控制机制
- » 商汤实践：AI 伦理标准制定
- » 商汤实践：AI 伦理外部联合研究



在可持续发展原则、以人为本原则与技术可控原则的指引下，商汤在以下几个方面开展 AI 伦理与治理的工作：

01 商汤实践：组建 AI 伦理与治理委员会

随着 AI 产业的发展，AI 伦理已经成为产业各方都关注的重要课题之一，也是 AI 企业发展过程中面对的巨大挑战。AI 伦理不仅仅是 AI 企业承担社会责任的重要因素，也是客户信任的基石。

商汤作为 AI 企业的领军者，也承担着更重要的行业责任和社会表率作用，商汤的伦理工作的核心，就是推动商汤所有员工共同参与到 AI 伦理的落实和负责任的 AI 企业建设过程之中来，推动商汤在 AI 伦理层面成为业界的标杆，推动商汤作为“可持续发展 AI”的价值提倡者的落实。为了充分落实商汤在伦理工作的原则，实现商汤伦理工作的“知行合一”，将伦理的原则落实在具体工作实践中，推动 AI 伦理工作能够独立、专业的在企业内部进行，因此需要成立专业的伦理委员会来推动相应的工作落实。

2020 年 1 月，商汤人工智能伦理与治理委员会正式成立，人工智能伦理与治理委员会由六名成员组成，包括两名外部顾问（其为人工智能伦理领域的学术专家）及四名高级管理层成员。作为集团在人工智能伦理方面的第一责任机构，人工智能伦理与治理委员会负责确定及实施人工智能伦理相关原则、战略及具体政策举措，在公司内部设计并推动整体组织践行 AI 伦理规范，推动商汤落实 AI 的可持续发展战略，同时发挥商汤作为行业领袖的影响力，在政府、产业、生态圈中倡导负责任的、可持续的 AI 发展观。

考虑到构建健康的 AI 伦理生态，不能是一家之言，需要 AI 企业、各类研究机构、政府和教育机构多方共同参与，兼听并收，商汤伦理委员会由商汤内部高层和外部的的人工智能权威专家组成，并且硬性规定外部委员不能少于 1/3 席位，以此来吸纳不同背景、不同视角、不同观点，综合兼顾，以确保委员会及其运作的专业性、中立性及客观性。

商汤 AI 伦理委员会外部委员则是伦理委员会的重要组成部分，是指不在公司担任除委员以外的其他职务，并与公司及伦理委员会其他委员间不存在可能妨碍其进行独立客观判断的关系，外部委员需要在不存在利益冲突的前提下对商汤的伦理与治理相关事宜进行判断和辅导。外部委员主要来自于从事 AI 伦理的高校、智库以及其他第三方机构。

日常工作职责



制定伦理领域
战略发展规划



产品伦理审核
与风险把控



伦理通识培训
与宣传



伦理治理
生态建设



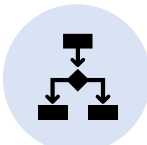
参与伦理标准
讨论与制定



广泛开展
联合研究

重要工作内容

对公司及伦理委员会重大事项（详见下）发表独立意见



公司伦理战略重大事项
的决策制定



伦理委员会发展规划
的设定与调整



提名、任免伦理委员
会专家顾问



有关法律、行政法规、部门
规章、规范性文件、及公司
章程规定的其他事项

02 商汤实践：人工智能伦理与治理风险控制机制

我们已建立人工智能伦理审查程序，以审查及监测我们产品及服务的潜在伦理风险，涵盖从项目审批、产品及服务的发布，到持续运营监控，及其后的产品及解决方案迭代的整个生命周期。我们的伦理审查程序根据符合我们上述三项人工智能伦理原则的标准评估项目及服务。在评估过程中，我们可选择驳回新产品方案，中断进行中的产品开发项目，或停止不符合我们原则及标准的现有产品。我们亦非常重视数据隐私保护，并会对我们所有的项目及解决方案的数据隐私相关事项进行全面审查。有关我们对数据隐私保护措施的审查流程，请参阅《数据隐私及个人信息保护》。我们亦聘请外部顾问为我们的审查程序及风险管理模型提供建议，以确保我们的产品及服务符合伦理规范。

从 2019 年至今，为紧密跟踪全球人工智能产业违法事件、伦理挑战，已在商汤内部建立起“全球 AI 伦理风险库”，包括涵盖百余项全球 AI 伦理正面及负面事件、案例分析，我们从产品立项、发布、运营全生命周期设置 AI 伦理审核环节，自发布日起审核商汤内部 100% 的新产品研发立项，并根据伦理准则拒绝了 10% 的产品立项提案。

我们已经实现了严格的伦理领域的风险控制体系和专业的伦理风险控制的机制，关注以下三个方面的风险控制：



数据风险
包括数据隐私、数据治理与数据质量三方面问题

数据隐私
指在人工智能的开发、测试、运行过程中存在的隐私侵犯问题，这一类问题当前是人工智能应用需要解决的关键问题之一。

数据质量
主要用于人工智能的训练数据集以及采集的现场数据潜在存在的质量问题，以及可能导致的后果，这是人工智能特有的一类数据安全问题。

数据保护
主要指人工智能开发及应用企业对持有数据的安全保护问题，涉及数据采集、传输、存储、使用、流转等全生命周期，以及人工智能开发和应用等各个环节。



算法风险
（包含了应用风险与系统风险）
基于算法的使用过程，从算法的自主性特征、应用性场景和归责性困难三个方面理解算法的伦理问题。

算法的自主性在提高决策效率和质量的同时，也将不确定性和不透明性引入了决策过程。

应用风险：算法在应用性场景下的伦理问题，主要包括算法偏见、算法歧视和个人隐私保护等。

系统风险：算法的归责性困境问题。
算法能够自主地调整操作参数和规则,这种调整被比喻为“黑箱操作”，由此也就将不确定性引入了决策过程,从而对算法的可控性提出挑战



社会风险
由于人工智能是一种与传统完全不同的技术体系，因此，人工智能势必对人类社会产生深刻的从内而外的改变并引发相应的风险。

劳动竞争

主体多元

边界模糊

能力溢出

惩罚无效

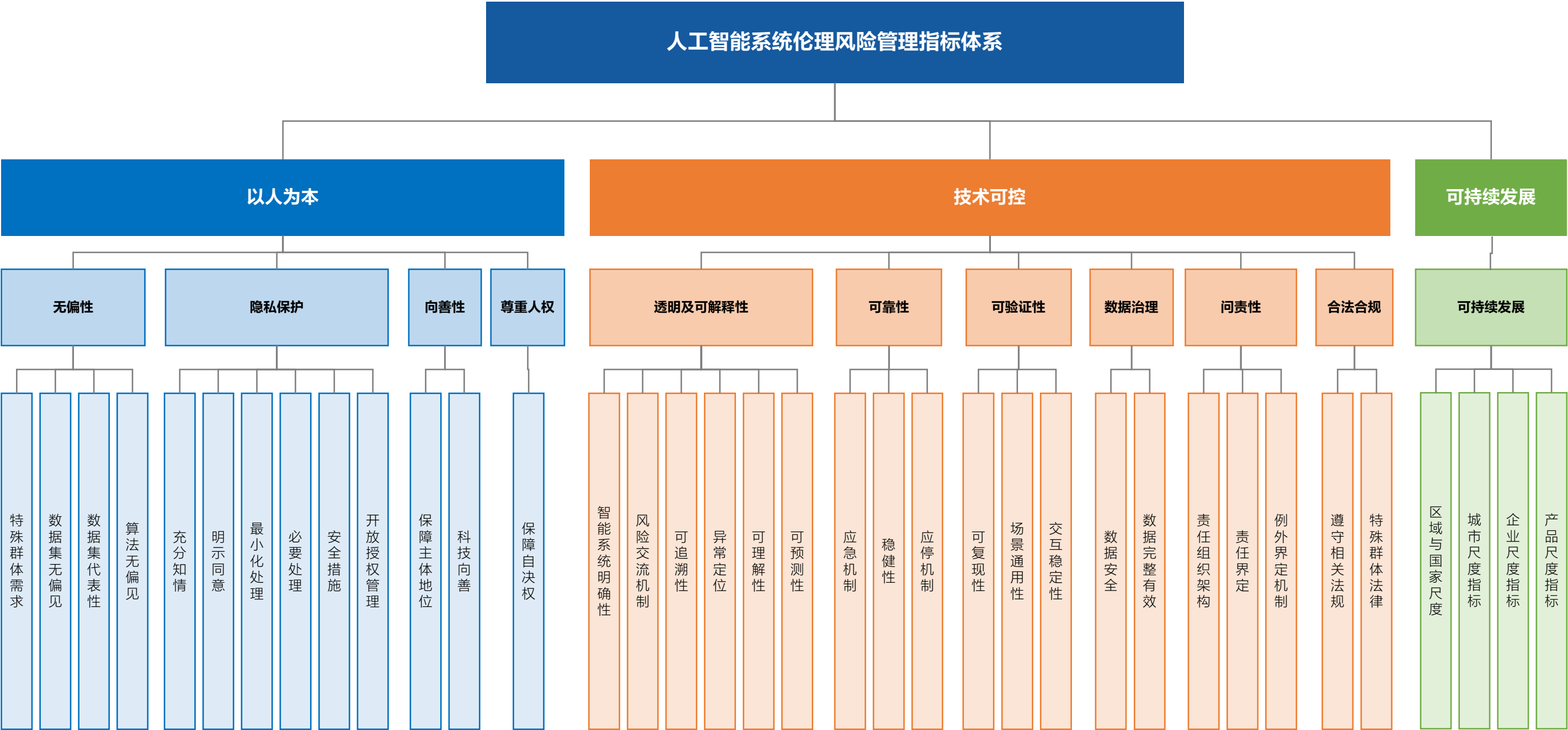
伦理冲突

暴力扩散

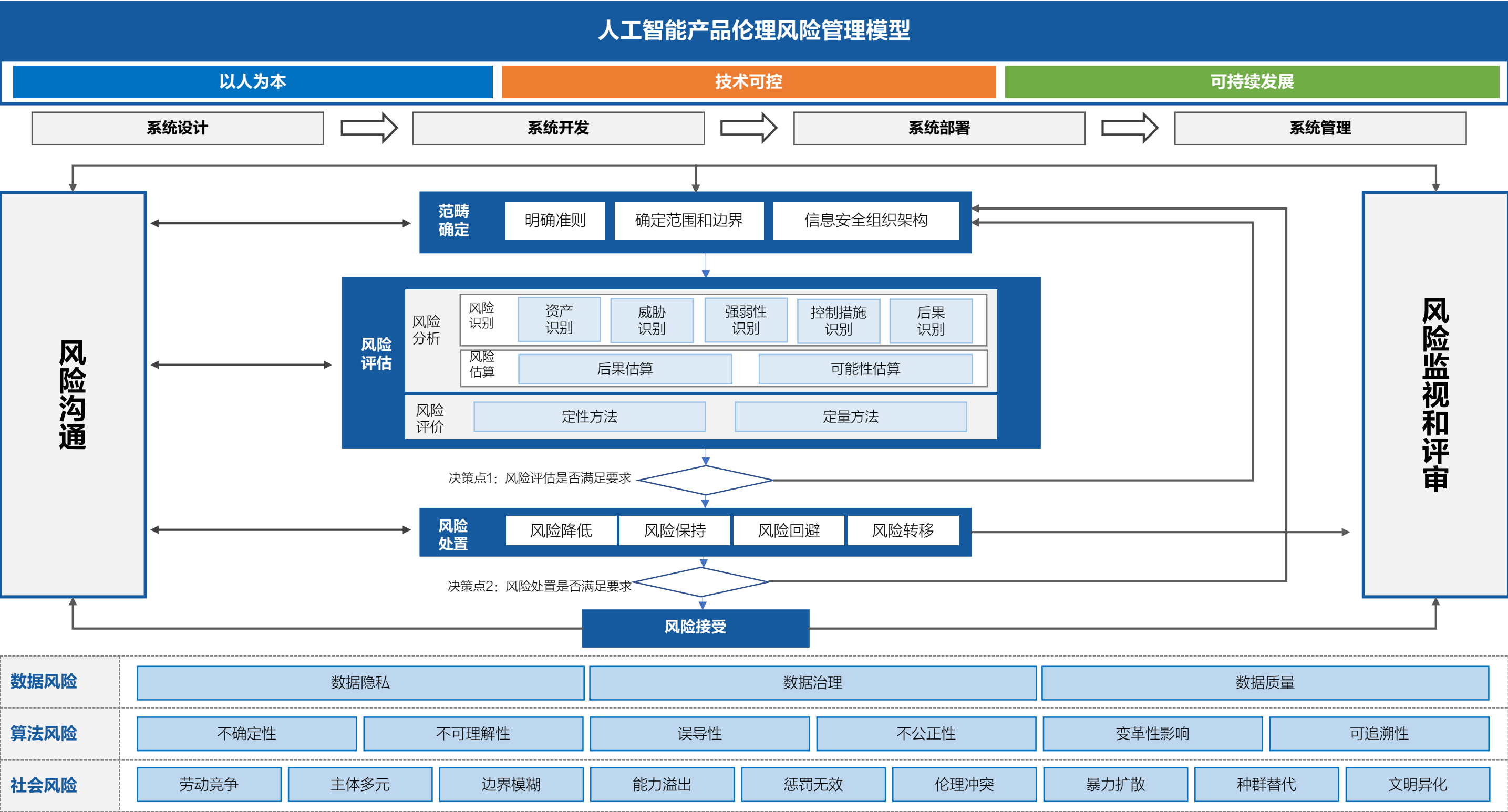
种群替代

文明异化

2021 年商汤开始重点推动在线风险控制系统的建构以提升 AI 伦理审核的深度，通过引入外部专家对整个审核过程和风控模型的提升提供专业指导，基于商汤的实践实现基于以人为本、技术可控、可持续发展体系的伦理标准工作，具体指标如下：



按照以上指标，我们在内部实现了人工智能产品伦理风险管理模型，如图：



03 商汤实践：AI 伦理标准制定

我们担任人工智能可信赖国家标准工作组的副主席。我们参与起草由人工智能可信赖国家标准工作组发起的 13 项国家或组织人工智能伦理标准。该等标准包括人工智能——风险评估模型、人工智能——伦理风险评估指南等。我们亦担任上海市人工智能标准化技术委员会副会长单位，使我们能够密切参与制订多个领域的行业标准，如负责任的人工智能、人工智能伦理及数据安全。

我们参与了《人工智能风险评估模型》、《机器学习算法安全评估规范》、《人工智能伦理与治理规范》、《信息技术生物特征识别 人脸识别系统应用要求》、《人脸识别系统可信要求》、《人工智能数据采集及标注安全规范指引》、《可信人工智能研发管理指南》、《人工智能可信赖白皮书》等十余项与伦理相关的标准项目。

除了积极参与可信人工智能（AI 伦理与数据安全）领域的多个标准工作的落实，商汤科技通过担任人脸识别国家标准工作组组长，AI 可信赖国家标准组副组长，以及上海市人工智能标准化技术委员会副主任委员，联合工作组成员，推进人工智能技术、产品的标准体系建设，制定具体标准计划，全面开展国家标准制修订工作，促进行业更健康发展，为全球标准化事业贡献商汤智慧。

＜ 伦理 / 安全相关会议 ＞		
序号	组织	承担角色
1	人脸识别国家标准工作组	组长
2	IEEE 人脸识别工作组	主席
3	AI 可信赖国家标准工作组	副组长
4	ISO / IEC JTC1 WG3 AI 可信赖	中国代表专家
5	全国信息安全标委会	委员
6	信通院“互联计划专项”	主编之一
7	清华大学智能社会治理标准项目	成员
8	新一代人工智能联盟（AITISA）数据隐私保护专题组	成员
9	CCSA / TC8 网络与信息安全 / WG3 安全管理	成员

04 商汤实践：AI 伦理外部联合研究

我们与第三方机构及智库的合作使我们能够实时了解人工智能伦理领域的最新发展，并在实施人工智能伦理实践时保持中立及客观。该等合作亦巩固了我们作为人工智能生态体系中的行业领导者的地位，并进一步提升我们在倡导负责任及可持续人工智能方面的影响力。

为推进我们在人工智能治理领域以人为本的原则，我们承担清华大学人工智能国际治理研究院的副理事长单位角色，该研究院是中国人工智能治理领域的领先学术机构，由国内外知名学者代表。我们与该等学者合作，推动诸如敏捷治理等创新及可持续的治理模式的研究。

我们在人工智能及合规领域倡导技术可控原则，并与上海交通大学共同成立计算法学与人工智能伦理研究中心，在数据安全、隐私保护、无偏算法等领域推行研究。我们亦与其他机构合作，对人工智能算法的可解释性进行研究。

为进一步推动我们可持续发展原则在全球发展，我们与联合国及其他国际组织进行深入合作。在联合国于 2021 年 6 月发布的《人工智能战略资源指南》（Resource Guide on Artificial Intelligence (AI) Strategies）中，《AI 可持续发展白皮书》（Code of Ethics for Sustainable AI Development）被选定为私营部门的主要参考出版物之一。

商汤在人工智能伦理的研究涉及诸多课题，包括但不限于人工智能伦理及法律、人工智能治理及可持续人工智能。我们认为相关研究工作巩固了我们人工智能伦理原则的基础，并将我们的产业实践与人工智能伦理原则结合起来。商汤积极开展人工智能领域对外合作研究，并与全球人工智能学术联盟、清华大学、上海交通大学高校团队进行人工智能相关课题的联合研究。

目前与高校智库合作的课题数量为 10 个，其中上海交大的研究课题为 7 个，清华大学人工智能国际研究院的合作课题方向为 2 个，上海科学学所研究课题 1 个，具体题目如下（附表）：

合作方	课题名称
上海交大凯原法学院	人工智能伦理及司法案例研究
	AI 伦理与立法的比较研究
上海交大计算法学与人工智能伦理研究中心	大数据证据在刑事司法中的应用与价值组
	数字经济产业国际规则研究
	后疫情时代城市技术治理的风险及防范
	矛盾纠纷源头治理中的智慧司法研究
	数据的法律属性及权属研究
上海市科学学所	数字城市生物信息特征社会实验
清华大学国际人工智能研究院	后疫情时代的数智化治理报告
	AI 可持续发展理念的企业社会责任报告

结语

本报告是商汤第二次发出《AI 可持续发展报告》，相对于去年的报告我们在拓展了“可持续发展”的内涵和外延的基础上，将过去一年多商汤在不同方向的企业实践放在了新的报告中。

商汤科技的使命是坚持原创，让 AI 引领人类进步，让 AI 推动经济、社会和人类的发展，建设更美好的未来，这促使商汤不断将人工智能用于造福大众、增强行业能力和包容开放创新。在这个过程中，坚持不断的推动 AI 伦理的价值观落实，推动可持续发展的人工智能的实践，这是我们坚持“知行合一”，推动企业愿景的重要支撑与举措，也是我们认为要落实“科技为善”的重要方式。正如清华大学苏世民学院教授、院长薛澜所说，“由于技术和发展如此之快，AI 的伦理治理目标在不断地变化。但我们欣慰地看到，商汤正在采取措施，与广泛的产业生态圈开展合作，为人工智能的治理努力做出贡献，并努力引导着人工智能向更加向善、普惠的方向发展”。

展望未来，人工智能技术的长期可持续发展将立足于伦理与道德的准则，以及它不断造福人类社会的能力。商汤作为 AI 领域的龙头企业，将不断落实相关的实践工作，并通过与产业生态的广泛合作连接，共同推动更加美好、更加幸福的未来生活的到来。

指导委员会

薛 澜	清华大学人工智能国际治理研究院院长
季卫东	上海交通大学中国法与社会研究院院长
杨 帆	商汤科技联合创始人、副总裁
张 望	商汤科技副总裁
林洁敏	商汤科技副总裁
金 俊	商汤科技首席营销官
张少霆	商汤科技副总裁、研究院副院长
骆 静	商汤科技副总裁

作者



刘志毅
商汤智能产业研究院
主任
liuzhiyi@sensetime.com



田 丰
商汤智能产业研究院
院长
tianfeng@sensetime.com



吕宝粮
上海交通大学清源研究院
执行院长
blu@sjtu.edu.cn



翁家良
AIII 人工智能国际研究院
院长
james.ong@aiii.global



杨 燕
商汤智能产业研究院
战略生态主任
yangyan@sensetime.com



毛翔铭
商汤智能产业研究院
助理研究员
maoxiangming@sensetime.com

鸣谢

吴晶斌、闫欣桐、吴怡帆、何聪辉、余锋伟、聂海斌、胡夕祝、郭俊翔、吴王韵

