



商汤科技人工智能 伦理与治理前沿实践



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。

目录

【关于商汤科技】	05
前言.....	07
第一章 伦理治理组织建设.....	08
第二章 科技伦理联合研究.....	10
第三章 商汤科技伦理治理实践.....	12
第四章 科技伦理治理生态合作.....	17
结语.....	19
【关于本报告】	20

【关于商汤科技】

作为行业领先的人工智能软件公司，商汤集团以“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、社会和人类的发展，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。

商汤坚持“平衡发展”的伦理观，倡导“可持续发展、以人为本、技术可控”的伦理原则，实行严格的产品伦理风险审查机制，建设全面的 AI 伦理治理体系，并积极探索数据治理、算法治理相关的检测工具和技术手段，致力于将伦理原则嵌入到产品设计、开发、部署的全生命周期，发展负责任且可评估的人工智能。

目前，商汤集团（股票代码：0020.HK）已于香港交易所主板挂牌上市。商汤现已在香港、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，商汤在泰国、印度尼西亚、菲律宾等国家均有业务。

前言

近年来, 伴随人工智能技术的快速进步和广泛应用, 人工智能伦理和治理也受到广泛关注。今年 3 月, 中共中央办公厅、国务院办公厅印发《关于加强科技伦理治理的意见》。这是我国在科技伦理治理领域的第一个指导性文件, 其中讲到科技伦理是开展科学研究、技术开发等科技活动需要遵循的价值理念和行为规范, 是促进科技事业健康发展的重要保障, 并明确提到伦理先行、依法依规、敏捷治理、立足国情、开放合作五项治理要求。

今年 10 月, 科技部部长王志刚在《求是》杂志发表署名文章《完善科技伦理治理体系 保障科技创新健康发展》, 其中明确讲到, “进入新时代, 我国科技伦理治理已经摆在事关科技创新工作全局的重要位置”。

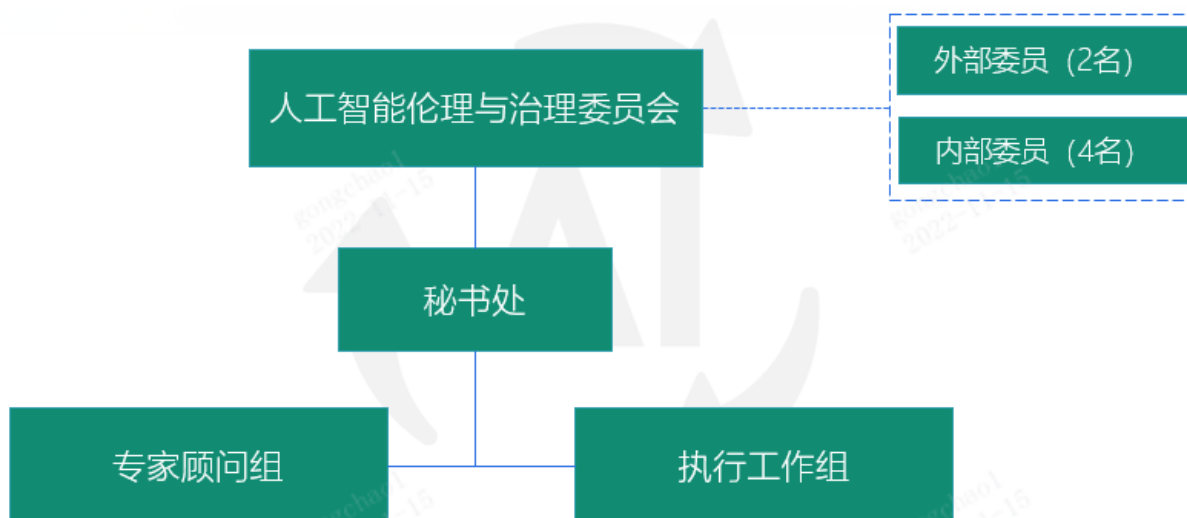
实际上, 无论是中国还是国际社会, 人工智能伦理治理的受重视程度近年来不断加强, 已成为文理交融、学科交叉的重点研究领域。商汤科技重视伦理治理, 并积极开展治理实践, 今年已经是第四个年头。本报告将从伦理治理组织建设、科技伦理联合研究、科技伦理治理实践、科技伦理生态合作四个方面, 分享商汤科技的 AI 伦理与治理实践情况。

第一章 伦理治理组织建设

2019 年 7 月，中央发布《国家科技伦理委员会组建方案》，同年 10 月，又成立了国家科技伦理委员会。

为系统性应对数据、算法及应用等不同层面的人工智能伦理风险，响应国家政策，商汤科技于 2020 年 1 月正式成立“人工智能伦理与治理委员会”，统筹推进伦理治理工作体系建设。商汤科技是业界最先组建人工智能伦理委员会、并将伦理治理工作作为公司战略级方向的科技创新企业之一。但实际上，在 2019 年，商汤科技已经开始启动 AI 伦理研究工作，并着手建设了 AI 伦理风险案例库，为公司的创新活动提供参考。

从组织架构来看，人工智能伦理与治理委员会是决策主体，向集团董事会负责。委员会包括内、外部委员，内部委员由四名企业高管担任，他们分别来自技术、工程、法律、伦理等相关专业背景；外部委员由两名在相关领域内享有专业性权威的专家担任。



图表 1：商汤科技伦理与治理委员会架构图

同时，委员会下设秘书处、专家顾问组和执行工作组，细致分工、各司其职，为整个伦理治理体系的设计、建立和高效运转提供了专业的资源保障。

从制度建设来看，为确保委员会的高效运转、强化公司全体员工对人工智能伦理治理的贯彻落实，该委员会先后制定了《人工智能伦理与治理委员会管理章程》、《商汤集团伦理治理制度》等配套内部规章。

在此基础上，委员会统筹推进集团的伦理治理工作，日常工作职责主要可以概括为六个方面：

第一，制定伦理领域战略发展规划；

第二，产品伦理审核和风险把控；

第三，内部伦理通识培训与宣传；

第四，伦理治理生态体系建设；

第五，参与伦理标准讨论与制定；

第六，开展伦理联合研究。

第二章 科技伦理联合研究

商汤科技非常重视内外部联合伦理研究工作。从 2019 年开始，商汤科技每年会整理出大量的全球正面和负面科技伦理案例。综合全球科技伦理治理发展情况，立足中国国情和企业发展，商汤科技提出了平衡发展的伦理观，倡导“可持续发展、以人为本、技术可控”的伦理原则。



图表 2：平衡发展伦理观三原则

可持续发展原则是指促进社会、经济、文化及环境的可持续发展，崇尚开放及包容合作，积极探索创新及可持续的人工智能治理模式的应用。例如，商汤科技认为，应该建立起广泛的科技“全生命周期”可持续发展理念；应对气候变化是科技力量履行可持续发展使命的重中之重。

以人为本原则是指追求不同文化之间的道德共识，尊重、包容、平衡全球不同国家地区的历史、文化、社会、经济等方面的发展差异，确保人权和隐私保护，无偏见地应用技术。例如，商汤科技倡导科技“全生命周期”尊重人、保护人、改善人的生活；提倡“普惠科技”理念，最大限度规避科技应用鸿沟。

技术可控原则是指确保人工智能由人类开发、为人类服务、受人类控制，人工智能应用导致的伦理责任也应由其控制者（人类）承担。例如，商汤科技倡导技术可控应该是科技工作者的基本道德操守和行为底线；同时，要发展、用好技术是主方向，谨防遏制技术进步。

围绕“平衡发展”的伦理观和商汤治理实践，商汤科技每年出版可持续发展白皮书，以及一些细分领域的研究报告，把我们最新的伦理探索 and 治理实践向全社会分享。例如，商汤科技已连续发布《AI 可持续发展白皮书》《“平衡发展”的人工智能治理白皮书》《AI 抗疫“以人为本”案例集》《AI 教育白皮书》等。



图表 3：商汤科技伦理与治理部分研究成果

同时，考虑到自身是一家市场主体，而非专业伦理研究机构，因此商汤科技在伦理研究方面非常重视外部力量的支持。因此，商汤科技与上海交通大学、清华大学建立起密切的伦理研究合作机制，积极开展合作课题，推动 AI 技术向负责任且可持续的方向发展。今年三季度，商汤科技也与相关高校就自动驾驶数据保护、亚太地区自动驾驶立法、以及元宇宙产业发展与伦理等课题展开合作。

第三章 商汤科技伦理治理实践

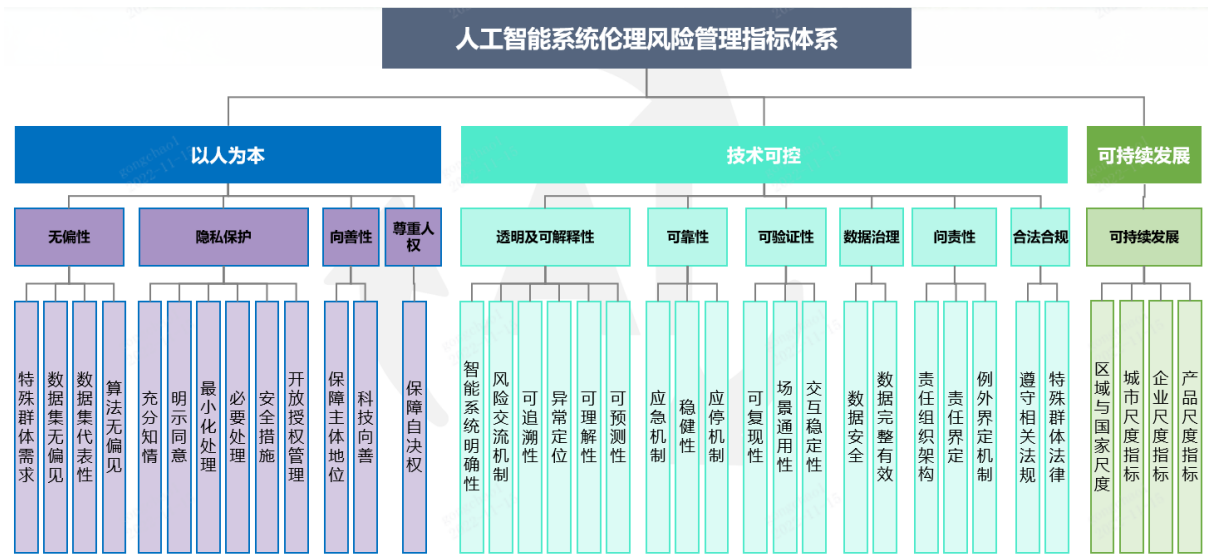
IBM 曾做过一项研究发现，虽然超过半数接受调研的组织发布了或公开认可了人工智能伦理的共同原则，但只有不到四分之一的组织能付诸实施。商汤科技坚持“平衡发展”的伦理观，把伦理原则落实到实际工作中，近四年来不断推动伦理治理工作迭代升级。

商汤科技在伦理委员会正式成立之前，就已经开始对全球的 AI 伦理案例进行研究归纳，作为内部伦理研究和治理的参考。2020 年正式成立人工智能伦理与治理委员会，商汤科技开始更加规范化、机制化、系统性的伦理治理实践。2021 年，商汤科技建立起全面覆盖的产品伦理风险审查机制，建立了人工智能系统伦理风险管理指标体系。2022 年，商汤科技伦理与治理委员会的一个核心工作就是再次升级伦理治理，着手建立 AI 治理技术和管理工具箱。

基于对全球风险案例的深入分析，商汤科技认为人工智能时代的风险挑战主要来自数据、算法和应用管理三个层面。商汤科技实行了严格的产品伦理风险审查机制，初步建立了 AI 伦理治理风险管理指标体系，在可持续发展、以人为本、技术可控这三项一级指标下，打造了包括无偏见、隐私保护、向善性、尊重人权、透明及可解释性、可靠性、可验证性、数据治理、问责性、合法合规、可持续发展在内的 11 个二级指标，以及 36 个三级指标的系统伦理风险管理指标体系。

这些指标体系丰富地涵盖了当前社会对人工智能伦理高关切度的一些事项。例如保护隐私、数据治理、合法合规等。在这些高关注度事项下面，

通过设置更细化的三级指标，来保证能够准确评估这一事项是否存在风险。



图表 4：人工智能伦理治理风险管理指标体系

商汤科技将伦理原则嵌入到产品设计、开发、部署的全生命周期，并于今年在业界首次提出，发展“负责任且可评估”的人工智能，将其作为开展人工智能治理的愿景目标。具体而言，“负责任且可评估”的人工智能具体包括七个方面的要求：对人负责、对社会负责、对环境负责、效果可评估、过程可评估、风险可评估、责任可溯源。



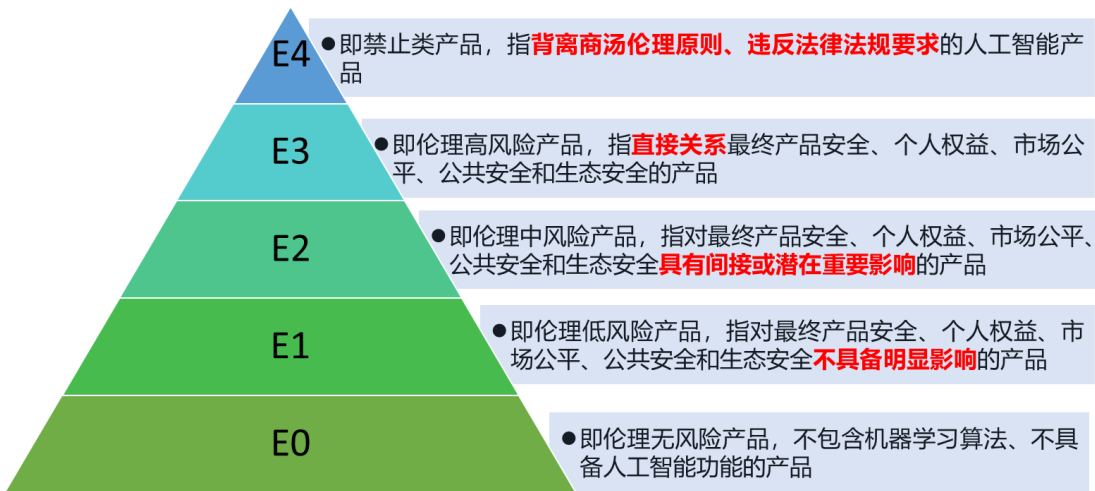
图表 5：商汤科技倡导发展“负责任且可评估”的人工智能

为了实现这一远景目标，商汤科技的治理实践可以分为三个方面：建立覆盖产品全生命周期的风险控制机制；开发了一系列流程工具和技术平台；做好公司内部的科技伦理宣贯。

第一，建立覆盖产品全生命周期的风险控制机制，针对人工智能领域面临的数据、算法及应用管理风险，初步形成了治理闭环。

具体而言，在应对算法风险上，建立了“算法定级备案管理与安全评估机制”，成立了“算法安全管理工作组”，依据算法的数据类型、业务场景、数据质量、存储现状、用户规模、对用户行为的干预程度等，将算法进行分类分级备案管理，从技术局限、算法设计、软件缺陷、数据安全、框架安全等维度对算法风险进行安全评估。

在应对数据风险上，建立了“个人信息及隐私数据影响评估机制”，并设置“数据安全与个人信息保护委员会”，确保产品遵循默认符合个人信息保护的设计要求。



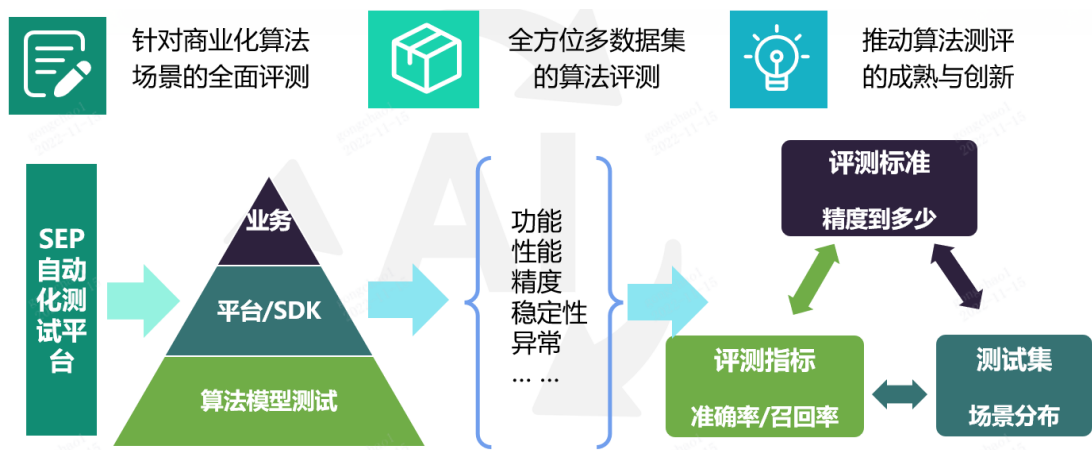
图表 6：产品伦理风险分级分类管理评估标准

在应对应用管理风险上，建立了“伦理风险分级分类管理机制”，设立

“伦理风险审核小组”，对产品设计、开发、部署、运营的全生命周期实施分阶段、分目标的分级分类管理。商汤科技在学习归纳国内外相关风险分级分类管理机制的基础上，结合我国伦理风险实际情况，以及实际工作中的心得体会，将伦理风险由低至高划分为 E0 至 E4 五个等级：分别是伦理无风险产品、伦理低风险产品、伦理中风险产品、伦理高风险产品、禁止类产品。

第二，在流程工具和技术平台方面，商汤科技已经开发了一系列覆盖数据治理、算法测评、模型体检、伦理审查的流程工具和技术平台。

具体而言，在应对算法风险上，算法黑箱是目前算法不可信任和不可解释的重要风险来源。在设计模型时，我们会将各种信息在代码中输出，当算法决策出错时，可以实现快速追溯原因。同时，我们通过建立模型体检平台，对模型进行推理攻击和逆向攻击测试。



图表 7：商汤科技建立基于真实场景数据的算法测评体系

另外，商汤科技基于真实场景下的数据集的算法验证与评测，开发了“算法测评工具”：一是，通过对主要场景和长尾场景的算法测评的全覆盖，实现了商业化算法的可信与可控。二是，设计多元化的测评项目和指标体系。三是，通过足够充分的数据集以及全方面的评测方案，对算法进行充分评测。

四是，将目前所使用的主算法评测方式与实践提炼出来，积极参与建立行业标准。

在应对数据风险上，开发了统一的数据治理平台以及规范的数据采集流程。在数据处理过程中，通过开发、使用自动化标注工具，减少人工接触的数据量，在模型训练的源头降低人类偏见的风险。而且，数据标注平台具备访问控制和身份验证功能，只有经认证的数据标注人员才能进行访问。

目前，商汤科技通过了多项网络安全及数据安全国际性权威资质。例如，信息安全管理体系国际认证、个人隐私信息管理体系标准认证、个人信息安全管理体系认证等，商汤所售的产品也获得了重要信息系统等级保护三级认证、可信人脸认证专项检测证书等。

在应对应用管理风险上，商汤科技结合产品伦理风险审查的不同阶段设计了一整套伦理风险自查工具，并开发了产品伦理风险审查平台。目前，所有商汤科技人工智能产品在立项、发布，再到上线运营的不同阶段均需通过产品伦理风险审查平台进行伦理风险审查。

第三，做好内部宣贯，推动加强员工对人工智能治理要求的关注理解，并在工作全流程中践行。商汤科技深刻认识到发展“负责任且可评估”人工智能的关键在于集团层面形成伦理治理的组织文化。因此，商汤科技建立了定期宣传和培训机制，每周都会向全体员工发送人工智能治理相关的重要动态，并会定期组织研讨会、邀请内外部专家对员工进行伦理治理培训。同时，强调“分类宣贯”，针对核心的产品人员，风险联络人，以及管理层，公司全员都配有针对性的宣传内容及方案，全方位让伦理贴近大家的工作视野。

第四章 科技伦理治理生态合作

商汤科技的伦理治理生态合作主要包括两个方面。一方面，商汤科技与高校、科研院所、政府机构等开展常态化交流。其中，通过与来自上海交通大学、清华大学、中国科学院大学、中国科学技术战略研究院、新加坡国立大学等机构的学者，针对相关科技伦理议题召开或参与研讨会，商汤科技充分吸收学者们最前沿科技伦理研究成果，并听取相关工作建议。



图表 8：商汤科技与高校召开伦理研讨会

同时，商汤科技将其伦理工作的机制建设、研究成果等与学者们分享，为学者提供一线研究案例。另外，商汤科技也会向有关政府部门及时反映关于人工智能伦理治理的一些最新研究成果和探索做法。目前，商汤科技也正在和多所高校联合申报国家关于科技伦理的课题。

另一方面，积极参与相关国内外组织活动。商汤科技积极参与国家信标委、国际 IEEE 等标准组织，并在多个工作组担任组长或副组长。到目前为止，商汤科技半年时间参与了十几项国内/国际 AI 伦理相关标准制定。

商汤科技也与上海交通大学、新加坡“人工智能国际研究院”等国内外机构联合发起了“亚洲科技促进可持续发展目标联盟”(Tech4SDG)。Tech4SDG 联盟现在涵盖中国、新加坡、印度、沙特等在内的亚洲 9 个国家和地区，机构成员超过 40 家。这个联盟旨在广泛开展科技伦理的研究合作，推动科技助力可持续发展目标的实现。



图表 9：亚洲科技促进可持续发展目标联盟成立

结语

本报告以上内容是对商汤科技近四年来，在人工智能伦理与治理上的一些实践情况的总结分享。近四年的探索，商汤科技初步建立了人工智能伦理治理体系，力求实现从理论到实践、从组织到机制、从研发到产品的全生命周期的多维度闭环。

商汤科技的伦理与治理工作也获得了一些认可，其发布的首份《AI 可持续发展白皮书》被联合国《人工智能战略资源指南》收录。商汤科技的人工智能伦理治理体系和相关技术工具，多次获得第三方机构的积极评价，并先后入选哈佛商业评论及中国人工智能产业发展联盟“可信 AI”的优秀案例。

当然，人工智能伦理与治理工作的道路还很漫长，还需要继续研究、实践、探索，商汤科技也欢迎各界提出宝贵意见，协同实现人工智能伦理与治理取得新进展。

【关于本报告】

商汤科技人工智能伦理与治理委员会持续推进伦理研究与治理，并及时发布相关报告，向社会介绍最新探索成果。

如需获取更多研究成果，或有任何建议、意见，请登录官网网址：
<https://www.sensetime.com/cn/ethic>

或发送邮件至：aiethics.committee@sensetime.com

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
徐立 商汤科技联合创始人、董事长兼首席执行官
张望 商汤科技副总裁、人工智能伦理与治理委员会主席
杨帆 商汤科技联合创始人、副总裁
骆静 商汤科技副总裁、首席运营官
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
林洁敏 商汤科技副总裁
孙大鹏 商汤科技副总裁

作者



田 丰

商汤科技
智能产业研究院院长
tianfeng@sensetime.com



宫 超

商汤科技智研院科技
政策与伦理研究主任
gongchao1@sensetime.com



胡正坤

商汤科技人工智能
伦理与治理研究主任
huzhengkun@sensetime.com

鸣谢

梅莹、闫欣桐、李玥璐、刘志毅、吴晶彧、成瑾、龚柳婷、
綦伟良、梁鼎、吴一超、王义飞、何聪辉、辛昱辰



商汤

senseTime