

商汤科技AI伦理与治理宣传手册 | 2025年 |

AI伦理与治理委员会

2025年12月

主席致辞

各位同事、合作伙伴及各界朋友：

人工智能的发展正深刻影响着社会结构与治理体系，也对企业提出了前所未有的责任要求。技术的价值不仅取决于算法和算力，更取决于我们能否以稳健的治理框架与清晰的价值判断，引导它朝着有益于人类社会的方向演进。

在商汤科技，伦理治理是企业战略与日常运营的有机组成部分，而非仅是附加的约束。我们认为，负责任的创新需要在理念、制度与执行上形成闭环：从高层决策到研发设计，从数据管理到生态合作，都应当以对公共利益和长远价值的尊重为前提。只有这样，技术的发展才能在稳健安全与社会信任的土壤上扎根，并持续释放积极的经济与社会效益，最终实现科技普惠的愿景，让人工智能真正惠及更广泛的人群。

商汤伦理与治理委员会承担着协调公司整体AI伦理实践的使命。我们的目标是确保商汤在保持技术领先的同时，始终以全球视野和长远眼光审视自身行为，把安全性、透明性、公平性和可持续性嵌入业务的全生命周期。这既是企业对用户、合作伙伴和社会的责任，也是我们构建核心竞争力和品牌信任的基石。

这份《伦理宣传手册》是我们对这些理念的系统总结与实践蓝图。它不仅为每一位商汤人提供清晰的行动指引，也向合作伙伴传递我们的价值立场：尊重人类尊严，守护公共利益，促进技术与社会的良性互动。

未来，商汤伦理与治理委员会将继续秉持理性、审慎与开放的原则，不断提高伦理治理水平，支持公司管理层和团队在复杂环境中做出稳健、负责任的判断，携手合作伙伴，一起推动人工智能向更加可信赖、可持续的方向发展。

谢谢大家！

张 望

商汤科技 副总裁
AI伦理与治理委员会 主席

目录 CONTENTS

主席致辞----------1
商汤科技简介----------3
商汤科技AI伦理与治理委员会简介----------4
第一部分 我们的认识
1.1什么是人工智能伦理治理----------5
1.2人工智能时代的风险来源----------7
1.3人工智能治理的分层目标----------8
1.4平衡发展的伦理观----------9
1.5合乎伦理的设计原则----------10
第二部分 我们的实践
2.1起源与初期探索（2017-2019）----------12
2.2管理体系建设（2020-2021）----------13
2.3夯实治理能力（2022-2023）----------14
2.4深化与引领（2024-2025）----------15
第三部分 治理体系
3.1治理组织架构----------16
3.2制度框架----------17
组织与决策机制----------17
技术研发全流程规范----------17
风险防控与应急响应----------17
生态协同责任----------17
3.3 伦理审查机制与重点----------18
第四部分 生态构建与行业认可
4.1国内生态构建----------22
4.2国际生态构建----------30
4.3行业认可与实践影响----------33
第五部分 未来挑战与展望
5.1未来挑战----------35
5.2短期工作目标（2026-2027）----------37
5.3长期战略方向----------37
5.4总结呼吁----------38
联系我们----------39

商汤科技简介

商汤科技是亚洲知名的人工智能软件公司，以“坚持原创，让AI引领人类进步”为使命，长期投入于原创技术研究，不断增强行业领先的多模态、多任务通用人工智能能力。

于2014年底成立，2021年12月30日在香港交易所上市。

商汤打造新型人工智能基础设施——商汤AI大装置SenseCore，打通算力、算法和平台，并建立“商汤日日新SenseNova”大模型及研发体系，以低成本解锁AGI任务能力，推动高效率、低成本、规模化的AI创新和落地。

商汤科技业务涵盖生成式AI、视觉AI和创新业务，以高效率、低成本、规模化的AI创新和落地，打通商业价值闭环，引领人工智能进入工业化发展阶段。



商汤科技AI伦理与治理委员会简介

商汤科技人工智能伦理与治理委员会（以下简称“伦理委员会”）是公司为支撑战略发展、落实负责任人工智能理念而设立的专门机构。

委员会注重人工智能技术属性与社会属性的融合，前瞻性评估和应对AI技术及产品可能带来的社会问题，防范公司在发展过程中潜在的伦理风险。其设立旨在通过规范治理机制和明确职责权限，保障商汤AI研发与应用的合规性、透明性和社会责任，推动公司在技术创新与社会价值之间实现平衡。





第一部分 我们的认识

商汤科技AI伦理与治理委员会的使命是**协调和推动公司整体AI伦理治理实践，确保技术创新在安全、透明、公平和可持续的原则下服务于社会与公共利益。**

为了实现这一使命，首先需要对“治理”的内涵有深刻而系统的理解，既涵盖技术本身的规范和约束，也包括社会、行业与公共价值的协调。基于这一认识，我们在长期学习与实践中不断总结经验，逐步形成了一系列指导思想和行动路径。最终，我们致力于实现科技普惠的愿景，让人工智能真正惠及更广泛的人群。

1.1 什么是人工智能伦理治理

人工智能伦理治理，以相关国际法、《世界人权宣言》、《联合国宪章》和《2030年可持续发展议程》等国际规范为基础，遵循“和谐友好、公平公正、安全可控”等和谐社会基本原则，对AI系统的交互、规划、感知、生成等核心能力进行规范与引导，着力平衡其在经济、社会和环境三大维度的影响，确保技术发展符合人类共同价值观并服务于可持续发展目标。

[本段对“人工智能伦理治理”的界定系商汤科技AI伦理与治理委员会秘书处结合实际工作所作的阶段性总结，仅供参考。]



（图：人工智能伦理治理内涵框架）

附：国际组织最新定义/原则摘要（仅供参考）

Recommendation on the Ethics of Artificial Intelligence

（2021，193 个成员国采纳）

把AI伦理治理看作一个基于整体、跨文化、可演进（evolving）的价值/原则/行动框架，指导社会负责任地处理AI在已知与未知方面对人类、社会及环境的影响。其核心价值包括人权尊严、隐私、透明性、公正、责任、可持续性等。

OECD AI Principles

（2019 年发布，2024 年更新指南）

强调创新与可信AI（trustworthy AI），原则包括尊重人权与民主价值、公正、透明度与可解释性、安全 / 鲁棒性、责任与问责制，以及政策 / 治理建议，以指导国家和AI行为者落地实践。

《欧盟人工智能法案》

（AI Act）



法案采用风险分级管理（high-risk, prohibited, etc.），要求高风险AI系统在透明度、可解释性、安全性、人类监督等方面满足严格要求，并对数据使用、模型训练与部署等环节设定法律责任与标准。虽与“伦理治理”不完全同名，但其实质就是将伦理原则纳入法律强制性治理机制。

1.2 人工智能时代的风险来源

新技术的出现不仅带来了巨大的发展机遇，同时在其早期阶段也伴随着复杂而多样的风险挑战。作为一家以技术原创为核心使命的人工智能企业，我们在系统界定人工智能伦理治理的内涵，并分析其在经济、社会与环境三大维度的价值平衡后，意识到当前AI发展中仍存在政策与规范的空白。

基于对全球案例的深入研究，我们认为，人工智能时代的风险主要源自数据治理、算法设计与应用管理三个层面。具体而言：

数据层面的风险 **主要涉及隐私保护、数据质量与数据安全三个方面。其中：**

- 隐私保护风险是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，这一类问题是当前人工智能应用需要解决的关键问题之一；
- 数据质量风险是指用于人工智能的训练数据集以及采集的现场数据可能会存在的质量问题，以及导致相应的不良影响，这也是人工智能特有的一类数据风险问题；
- 数据安全风险是指人工智能开发及应用过程中，企业对持有数据的安全保护问题，其涉及数据采集、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。

算法层面的风险 **主要涉及算法决策、算法“黑箱”与算法安全三个方面。其中：**

- 算法决策风险是指因算法推理结果的不可预见性与人类自身的认知局限，导致无法预测智能系统做出的决策原因与产生的效果，归责问题便是此类风险的典型代表；
- 算法“黑箱”风险主要指因采用复杂神经网络的算法导致决策不透明与无法被充分解释，而造成的可解释性风险；
- 算法安全风险是指因模型参数泄露或被恶意修改、容错率与弹性不足引发的风险。

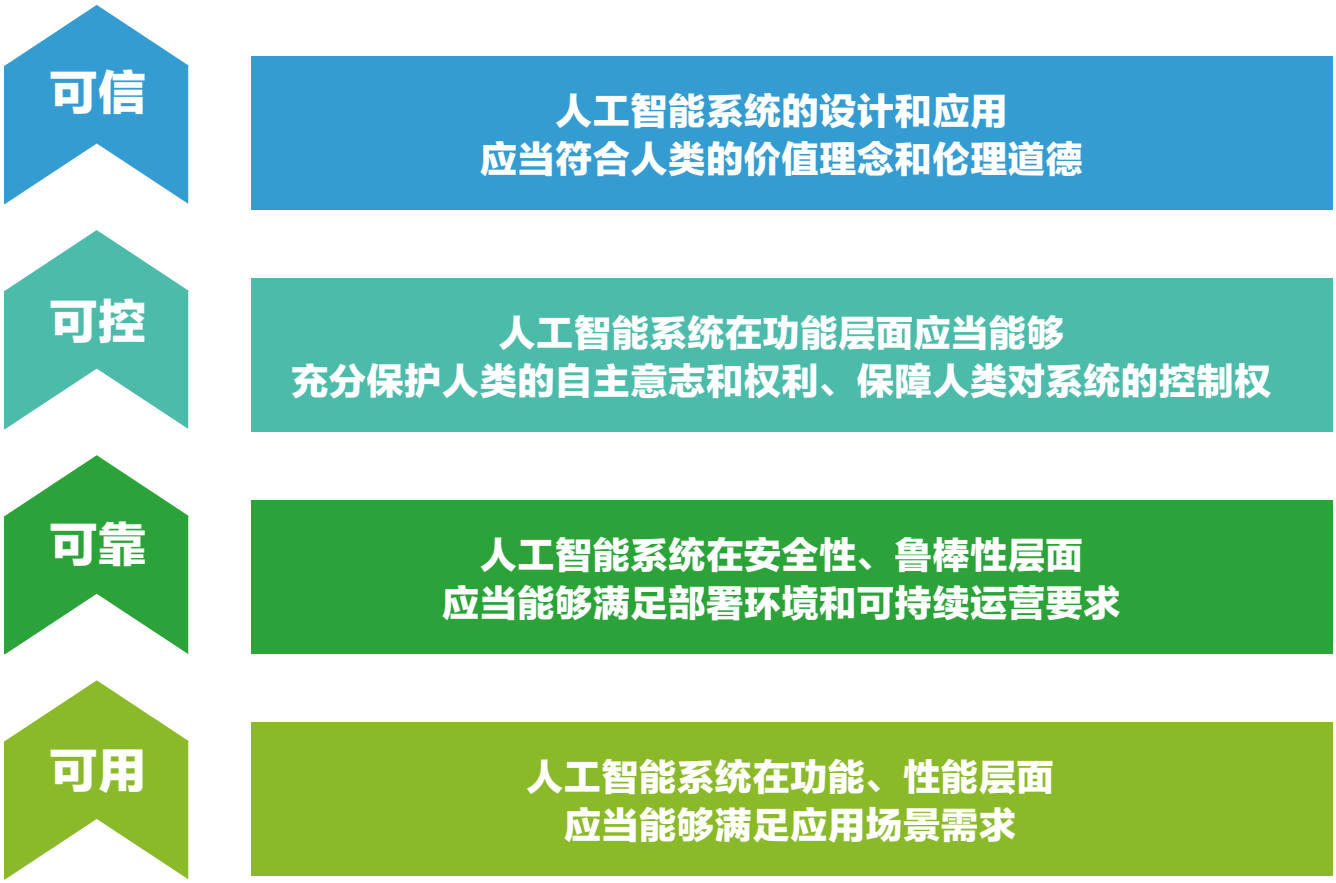
应用管理层面的风险 **涉及算法歧视、伦理冲突、劳动竞争等多个方面。例如：**

- 因人为原因、原始训练数据存在偏见性、机器在自我学习过程中输入了数据的多维不同特征，而造成算法将偏见引入决策过程的算法歧视风险；
- 因算法设计者出于自身的利益，对用户进行不良诱导、过度依赖算法本身、盲目扩大算法的应用范围而导致的算法滥用风险；
- 以及人工智能在长期对于就业的负面影响，加剧不正当竞争与市场支配地位，引起责任界定的困境与风险。

1.3 人工智能治理的分层目标

在明确了人工智能时代主要风险来源之后，问题的核心并不仅在于识别风险本身，而在于如何通过系统化的治理措施加以应对。风险的存在表明，单靠技术创新或市场力量难以确保人工智能的安全、可靠与可持续发展，因此，治理体系必须与风险管理紧密结合，形成从识别风险到防控风险、再到实现价值导向的闭环。基于此，我们提出人工智能治理的分层目标，将治理效果分为可用、可靠、可控与可信四个层次，既体现技术能力的保障，也强调价值引领与社会责任，为应对数据、算法和应用管理等各类风险提供明确方向。

我们对全球人工智能治理诉求进行了分析，并结合自身对技术和市场的理解，我们认为，人工智能治理应当是一个价值牵引、技术先行、多方参与、分层推进的动态进程。从治理效果看，人工智能治理的实现由低至高，可分为可用、可靠、可控、可信四个层次：



（图:人工智能的分层治理目标）

1.4 平衡发展的伦理观

作为行业领先的人工智能企业，商汤集团秉持“平衡发展”的伦理观，致力于在“治理”中推动人工智能技术助力各行各业实现更好的“发展”。

商汤集团遵循“以人为本、技术可控、可持续发展”的伦理原则，致力于发展负责任且可评估的人工智能。具体来说：



以人为本

应保护人类自由、尊严等基本权利，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，不损害人类的主体地位，确保人权和隐私保护，以及无偏见地应用技术；

技术可控

要确保人工智能由人类开发、为人类服务、受人类控制，相应地，其人工智能应用导致的伦理责任也应由其控制者（人类）承担；

可持续发展

要促进社会、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。

1.5 合乎伦理的设计原则

合乎伦理的设计原则是人工智能研发与应用的核心规范，遵循这些原则有助于确保AI技术在科学研究、产品开发和社会应用中既安全可靠，又符合社会价值与公共利益。这样做，不仅能够提升创新效率，更能在安全、可信和公平的前提下推动经济发展、社会进步和生态保护，实现“智能向善”的目标。具体原则如下：

保障数据质量

用于算法研究、模型训练的数据应具备充分的相关性、完整性、代表性、时效性和准确性；采取主动措施防止数据偏见。

保护个人权益

在算法研究、模型训练及部署应用等过程中应充分尊重个人依法享有的肖像权、名誉权、荣誉权、隐私权和个人信息权等权益；对个人信息的收集与使用应坚持最小必要原则，符合法律法规要求。

尊重知识产权

在算法研究、模型训练过程中使用受知识产权保护的数据应当符合法律法规要求；在产品开发和服务提供过程中积极探索知识产权保护机制。

确保可靠可控

在算法研究、模型训练及部署应用等过程中应确保算法模型具备充分的安全性和鲁棒性；在产品开发过程中，应确保在特定场景下，能够实现有效的人类干预和控制。

促进透明可释

在算法研究及模型训练过程中应优先采用可解释性强的算法模型，充分考虑不同利益相关方的信息诉求，建立必要的披露机制，明确告知用户产品局限及潜在风险，充分保障用户的知情权和选择权。

推动责任可溯

应具备相关机制实现对输出结果背后的模型和数据进行追溯；明确在研发、设计、部署、运营和服务等各环节的责任方及相关职责义务。

坚持智能向善

人工智能的研究和应用应符合人类社会的价值追求和发展目标。算法研究、模型开发以及部署应用应有利于推动经济发展、社会进步、民生改善和生态环境保护。



第二部分 我们的实践

商汤科技的 AI 伦理治理经历了从建立基础制度、完善评审流程，到将责任理念融入产品研发与行业共建的持续演进，逐步形成与国际接轨、覆盖全生命周期的治理体系。

起源与初期探索
（2017-2019）

管理体系建设
（2020-2021）

夯实治理能力
（2022-2023）

深化与引领
（2024-2025）

（注：展示案例为部分展示，兼具一定程度的里程碑意义）



2.1 起源与初期探索（2017–2019）

商汤自2014年创立之初便意识到，人工智能不仅是技术创新，更关乎社会责任。2016年国际上开始兴起关于人工智能伦理原则的讨论，商汤结合自身的洞察与实践，从早期算法研究到场景化应用，公司在探索中不断反思AI对公平、隐私和安全的影响，逐步形成“技术向善”的核心理念，并着手为后续治理体系打下价值根基。

（2017 – 2019 年商汤科技 AI 伦理治理探索案例，部分展示）

时间	案例	主要内容	伦理价值 / 意义
2017	发布企业价值观与 AI 伦理原则	在企业官网及内部文化中提出“以人为本、可持续、可控”等理念	为后续产品开发奠定负责任 AI 的基调
2018	SenseCare 智慧医疗平台	推出医学影像AI平台，支持放射科诊断，并强调数据安全与患者隐私	医疗场景中率先引入隐私保护和算法验证
2018 – 2019	金融级人脸识别与活体检测方案	推出防伪/活体检测技术，满足金融机构合规要求	在身份认证中关注欺诈防护与安全责任
2018 – 2019	参与 IEEE 人脸识别国际标准制定	参与 IEEE 2790 等标准草拟，将伦理考量纳入人脸识别标准	推动行业标准融入隐私、安全和公平等价值
2019 – 今	全球 AI 伦理风险库建设	建立商汤内部“全球AI伦理风险库”，收录百余项全球正负面 AI 伦理事件；在产品立项、发布、运营全生命周期设置伦理审核	系统化、全生命周期的 AI 伦理风险管理，确保产品开发与运营符合公司伦理标准和社会责任

2.2 管理体系建设（2020–2021）

随着业务规模与影响力的增长，商汤持续完善伦理治理框架：组建AI伦理与治理委员会，确立跨部门协作机制；制定行为准则与内部审查流程，覆盖研发、测试、部署全链路；建立专家顾问与外部评审机制，让治理思路与国际先进经验接轨。这一阶段，企业的伦理实践开始从理念走向制度化。

（2020 – 2021 年商汤科技 AI 伦理治理体系建设案例，部分展示）

时间	案例	主要内容	伦理价值 / 意义
2020	组建 AI 伦理与治理委员会	组建AI伦理与治理委员会，明确跨部门协作机制，确保产品、技术、合规、法律、运营等部门在伦理决策中联动	使伦理管理制度化、责任落实全链路、风险可控，同时增强企业社会信任和行业示范作用
2021	推动在线AI伦理风险控制系统的建构	商汤科技的伦理风险控制系统涵盖了从项目审批、产品发布到持续运营监控及产品迭代的整个生命周期。在评估过程中，企业可选择驳回新产品方案、中断进行中的产品开发项目，或停止不符合商汤原则及标准的现有产品	这一管理体系确保了伦理风险的早期识别和有效控制，体现了商汤科技对人工智能伦理治理的高度重视
2021	构建人工智能系统伦理风险管理指标体系	基于商汤实践和 AI 伦理治理基本价值理念研制人工智能伦理风险管理指标体系	将抽象理念转化为可操作的量化指标，使伦理治理不再依赖个体判断，保证一致性与可执行性

2.3 夯实治理能力（2022–2023）

在制度逐渐成熟的基础上，商汤着力将治理能力融入日常运营：推出面向不同员工群体的伦理培训课程，提升伦理意识和风险识别能力；推动重点业务线开展“责任AI”试点，形成可复制的管理模板；建设伦理风险监测与预警工具，为产品和服务提供全周期保障。伦理治理不再是“附加项”，而是与创新并行的核心能力。

（2022 - 2023 年商汤科技 AI 伦理治理探索案例，部分展示）

时间	案例	主要内容	伦理价值 / 意义
2022	开设面向不同员工群体的AI伦理培训课程	重点针对管理层、产研侧、销售侧员工群体打造具有针对性的务实伦理培训课程	把抽象的价值要求转化为各岗位可执行的行为规范，强化了全员对责任研发和合规应用的理解与承诺，从而巩固企业AI伦理治理的文化基础
2022	建立“伦理风险分级分类管理机制”	依据产品对最终产品安全、个人权益、市场公平、公共安全和生态安全的影响程度，将伦理风险由低至高划分为E0至E4这五个等级进行评审管理	建立分级评审机制，使治理资源与风险相匹配，提升了AI产品全生命周期的责任管理与社会保障水平
2022	设立“伦理风险审核小组”、“伦理风险联络人”管理机制	构建起覆盖业务一线与决策层的双向沟通与把关体系，确保伦理风险在产品全生命周期中被及时识别、评估与处置	打通前线业务与治理决策的通道，实现对伦理风险的持续监测与快速响应，提升了企业AI研发与应用的透明度、责任感和防范能力
2022	联合国内外行业伙伴发起“亚洲科技促进可持续发展目标联盟（Tech4SDG）”	致力于推动人工智能等前沿技术服务联合国可持续发展目标，促进行业责任创新与区域合作	这一举措通过搭建面向亚洲的伦理交流与标准协作平台，推动人工智能等新兴技术在国际对话、规则共建与普惠应用中形成合力，强化了企业对全球科技治理与可持续发展的价值引领
2023	推出生成式人工智能治理基础设施–商汤“SenseTrust”工具体系	推出覆盖数据处理、模型训练、模型部署，以及推理服务等AI系统全生命周期的治理工具体系。为行业提供AI治理公益技术服务，推动建设安全可信的人工智能产业生态	使潜在风险能够被量化识别、持续监测和及时修正，强化了责任研发和安全应用的透明度与可追溯性，推动行业形成稳健、可信的技术治理文化
2023	加入 AI.Verify Foundation	作为成员单位，加入新加坡“人工智能验证基金会”（AI. Verify Foundation），携手全球相关方共同推动可信AI生态建设	参与全球多方合作，推动可信AI标准和验证机制建设，强化了商汤在国际AI伦理治理与责任技术实践中的引领作用。

2.4 深化与引领（2024–2025）

（2024 – 2025 年商汤科技 AI 伦理治理探索案例，部分展示）

时间	案例	主要内容	伦理价值 / 意义
2024	推动可信数字人行业标准建立	通过推动可信数字人行业标准的建立，旨在制定技术规范与伦理准则，确保数字人产品在内容安全、隐私保护、身份识别和交互可靠性等方面达到可验证的信任标准	从行业视角纳入伦理可信框架，强化行业责任意识，推动新兴技术在社会可接受范围内健康应用
2024	受邀成为中国人工智能产业发展联盟（AIIA）科技伦理工作组副组长单位	作为副组长单位，商汤科技参与制定行业伦理规范、推动科技伦理政策落地，并协调成员单位在AI研发、应用和治理实践中落实伦理标准与最佳实践	体现了其在推动行业伦理治理规范、参与制定标准与政策中的核心引领地位
2024	商汤金融大模型案例成功入选国际电信联盟全球首批“AI for Good”案例集	商汤金融大模型案例通过在金融场景中应用AI技术实现风险识别、合规管理和决策辅助，被国际电信联盟评选为全球首批“AI for Good”案例，展示了AI在社会责任和可持续发展中的实际应用价值	通过入选国际电信联盟“AI for Good”案例，展示了商汤在金融AI应用中将伦理原则、合规管理和风险控制融入全球可评估标准的国际示范价值
2025	科技伦理（审查）委员会管理体系认证试点	商汤科技作为人工智能领域首批受邀参与试点的企业，该认证工作由科技部与国家市场监管总局联合发起	通过参与科技伦理委员会管理体系认证试点，将企业治理行为纳入官方认可的伦理标准框架，彰显了商汤在AI伦理合规性和责任落实上的示范作用
2025	担任国家人工智能标准化总体组人工智能治理专题组联合组长	作为国家人工智能标准化总体组人工智能治理专题组联合组长，商汤科技参与制定AI治理相关国家标准，推动伦理规范、技术安全和责任落实在产业中的系统化应用	这一角色使商汤能够在国家层面推动并引领AI治理标准与伦理规范的制定与落地，强化产业责任意识并提升人工智能技术的安全性、可控性和社会信任度
2025	开拓伦理跨区域合作	在2025WAIC上，商汤科技AI伦理与治理委员会与印尼国家政府率先签署关于“人工智能伦理、教育与创新”的合作框架	推动跨国伦理治理对话与实践落地，成为全球AI治理领域标杆性跨国政企实践，彰显了商汤在负责任AI时代的卓越领导力与担当



第三部分 治理体系

3.1 治理组织架构

商汤人工智能伦理与治理委员会（AI Ethics and Governance Committee）是公司AI伦理治理的核心决策机构，采用“委员会+工作组”的协同管理模式，确保伦理治理的科学性、专业性与全球视野。

委员会的组织架构主要包括三个层级：

1.委员会主席与内外部委员

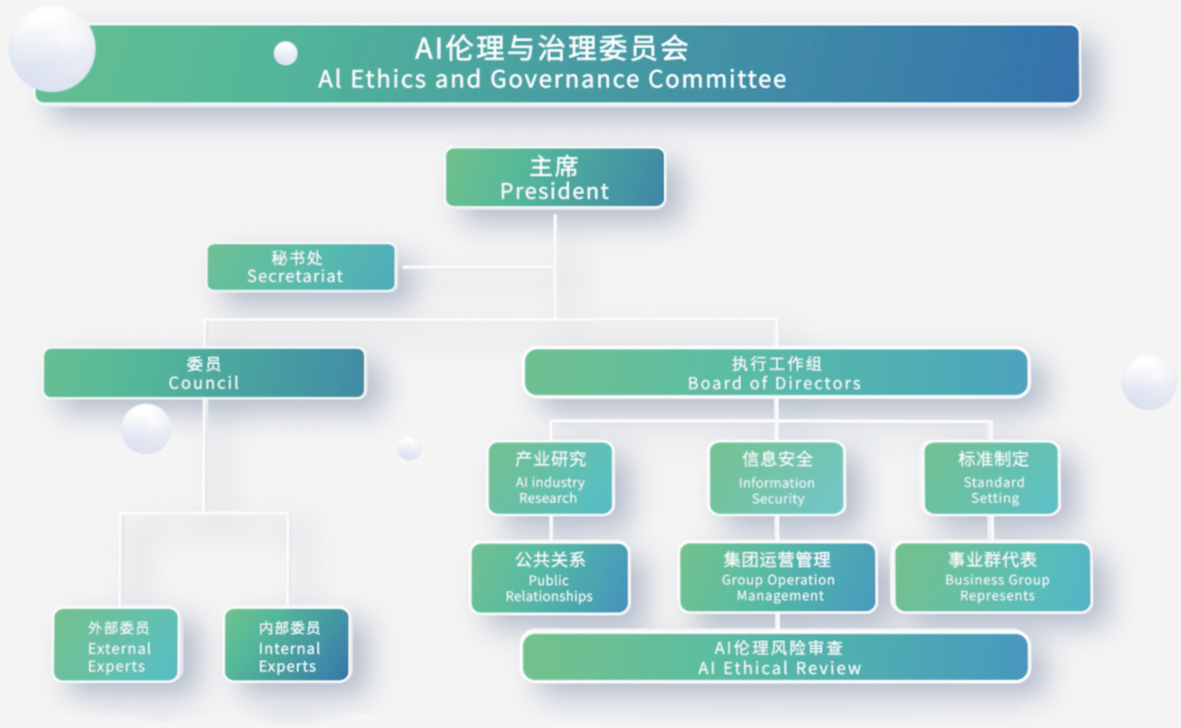
由公司高管和行业/学术专家组成，负责战略方向、重大决策和伦理治理政策的审批与指导。

2.秘书处

作为日常管理和协调机构，承担委员会日常运作、项目推进、跨部门沟通以及对外合作协调工作。

3.专业工作组/研究团队

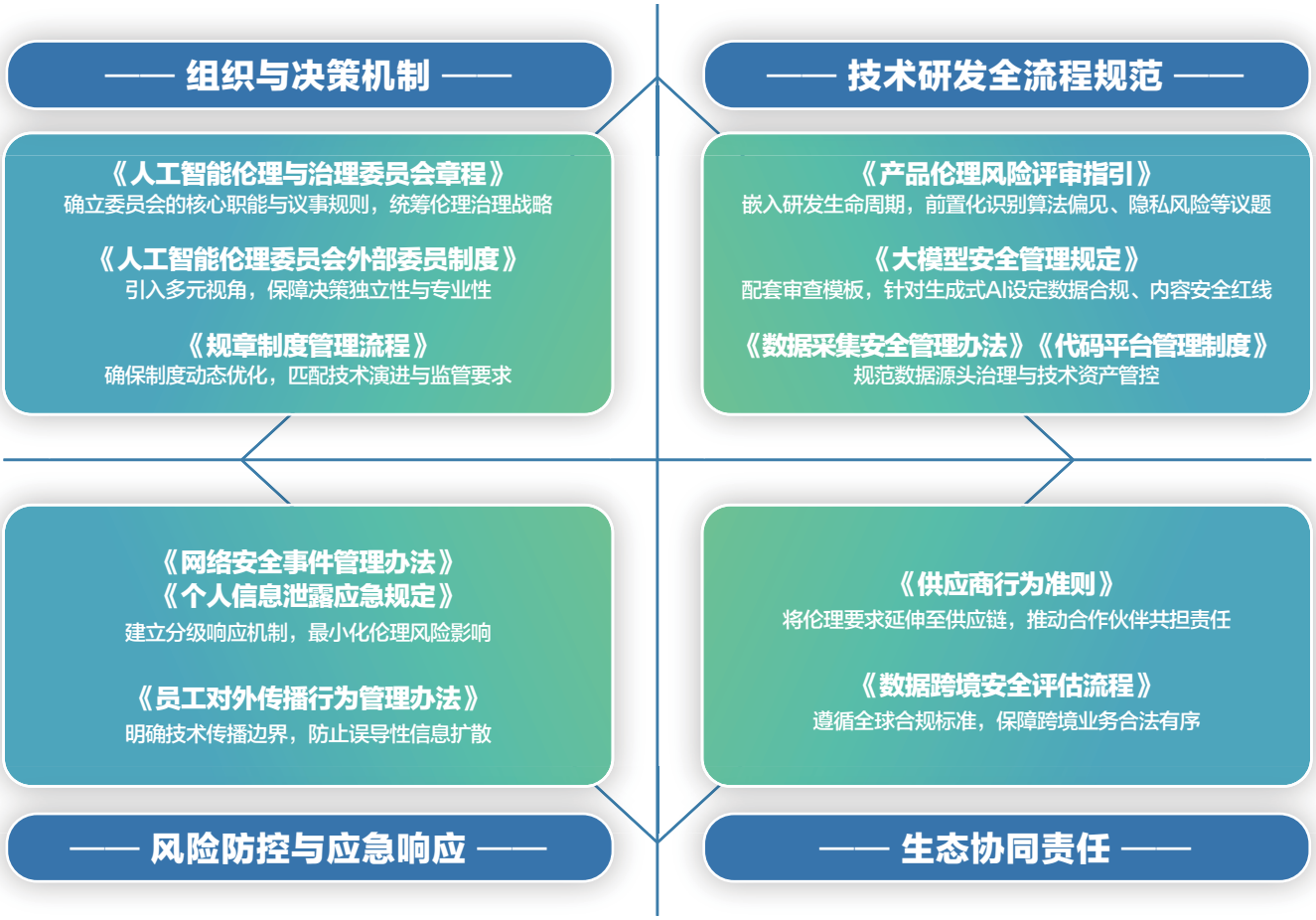
负责具体项目执行、研究案例开发、工具与方法制定，并形成对外输出成果。



这一架构体现了战略决策—日常管理—专业执行的三级治理模式，确保商汤科技在AI研发、产品部署及国内外合作中能够系统落实伦理与治理要求。

3.2 制度框架

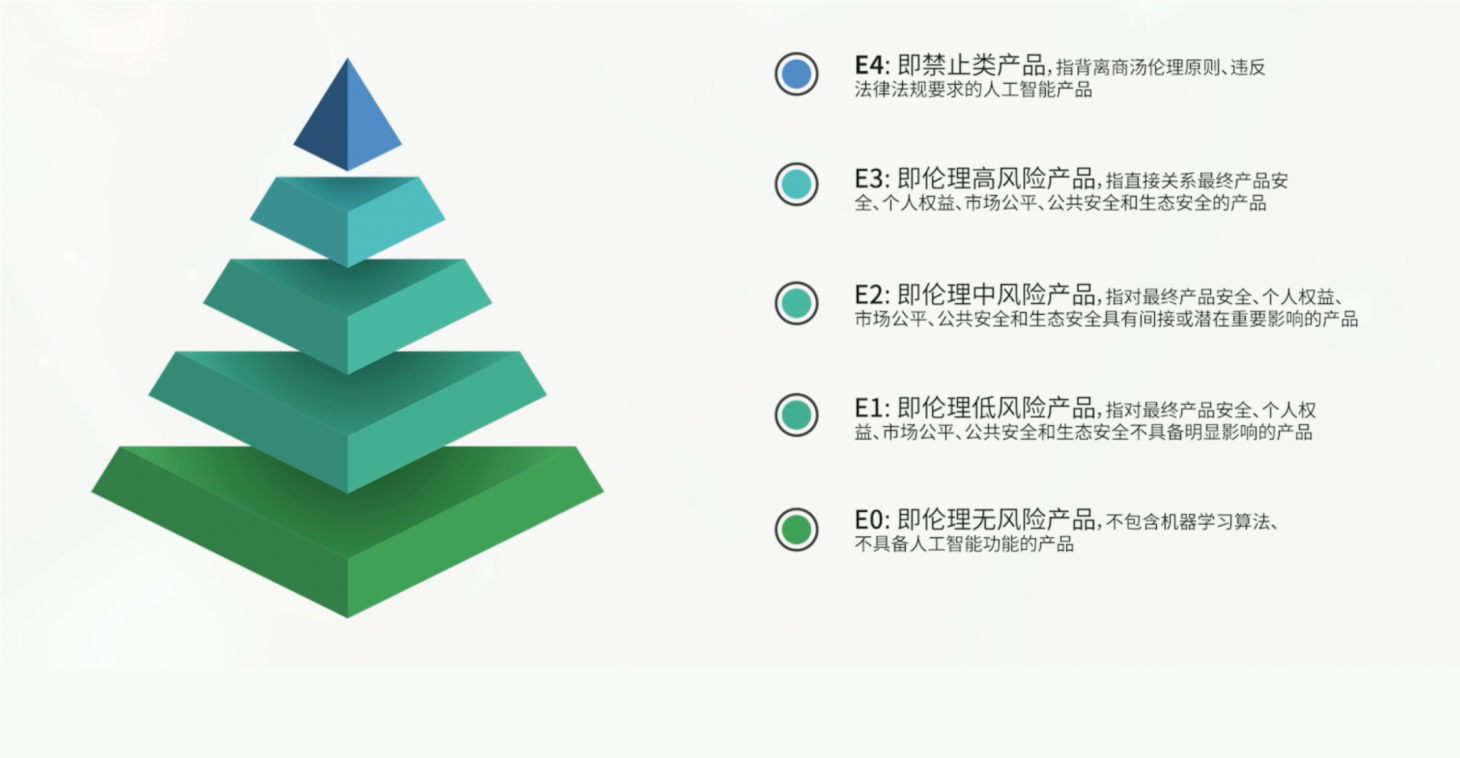
商汤集团以“以人为本，技术可控，可持续发展”为核心治理原则，构建了系统化、多层次的伦理治理制度体系，确保技术发展与社会价值相协调。本框架涵盖组织管理、技术规范、风险管控及合作伙伴协同四大维度，为伦理治理提供全面指引：



此外，我们也将宣贯培训、意识提升、流程改进等方面也纳入了制度与纪律方面的规范和要求。为加强员工对人工智能伦理治理的理解与践行，商汤科技高度重视系统化的宣贯与培训工作。集团通过多维度、全方位的知识宣传，将伦理理念融入日常工作，并结合不同员工群体的职责特点，开设针对性培训课程——覆盖研发、产品、销售及管理层等关键岗位，确保各类岗位能够高效提升伦理意识与责任认知。同时，公司定期举办专题研讨会，邀请内外部专家开展交流与培训，促进员工对伦理治理最新理念与实践的理解与应用，推动人工智能技术在安全、透明、公平和可持续的原则下健康发展。

在制度执行层面，我们持续优化内部工作流程，将伦理要求贯穿产品研发、测试、发布及运营的全生命周期。通过梳理关键环节、明确责任分工、建立标准化操作规范，推动各业务单元在实际工作中自觉遵循伦理原则。同时，公司定期评估和优化流程，结合实践案例与监管动态，提升整体治理效率与响应能力，确保技术创新与社会责任并行不悖。

3.3 伦理审查机制与重点



在人工智能研发与应用过程中，风险不仅可能对用户权益、社会公共利益产生影响，也可能引发法律和声誉风险。因此，建立系统化的伦理审查机制显得尤为重要。伦理审查不仅是对潜在风险的前置防控，更是确保技术开发与应用符合社会价值和伦理原则的关键手段。它能够将伦理治理从理念落实到操作层面，实现技术创新与社会责任的有机结合，为企业在快速发展的同时保持可持续性和可信赖性奠定基础。

在此背景下，商汤科技的AI伦理审查机制通过分级分类与全生命周期嵌入，实现对产品研发、发布和售后的全程管控。机制核心包括：依据风险金字塔（E0－E4）对产品进行分级，从无风险到禁止类逐级管控；在设计、发布和售后阶段设立审查节点，结合《伦理自查清单》和平台化协同工具，实现业务侧与伦理侧高效互动和全程可追溯；审查重点聚焦安全与权益、公共影响及法律合规，并根据风险等级采取差异化管控策略，低风险自查、中风险抽检、高风险全流程审查；机制同时通过动态优化，将售后反馈纳入迭代评估标准，确保形成闭环。整体特点是预防性、精准性和闭环性，为商汤“负责任AI”战略提供了系统化支撑。

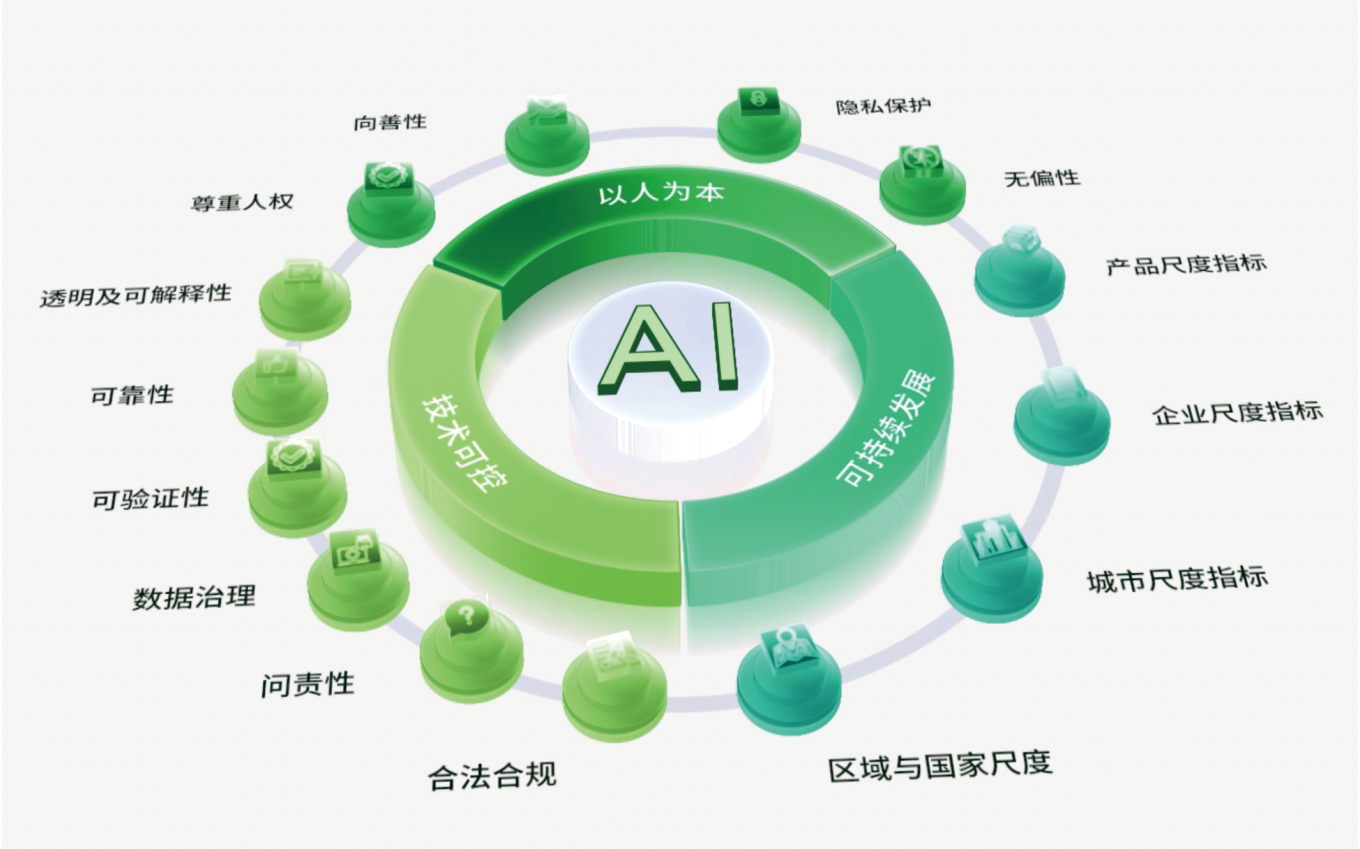
在明确了AI伦理审查机制的核心理念与运作特点后，下一步便是将这些原则落实到具体操作流程中。商汤科技针对AI产品开发和应用设计了全生命周期的审查管控流程，确保机制从理念到实践的有效落地。该流程分为产品立项阶段、产品发布阶段和产品售后阶段三个阶段：

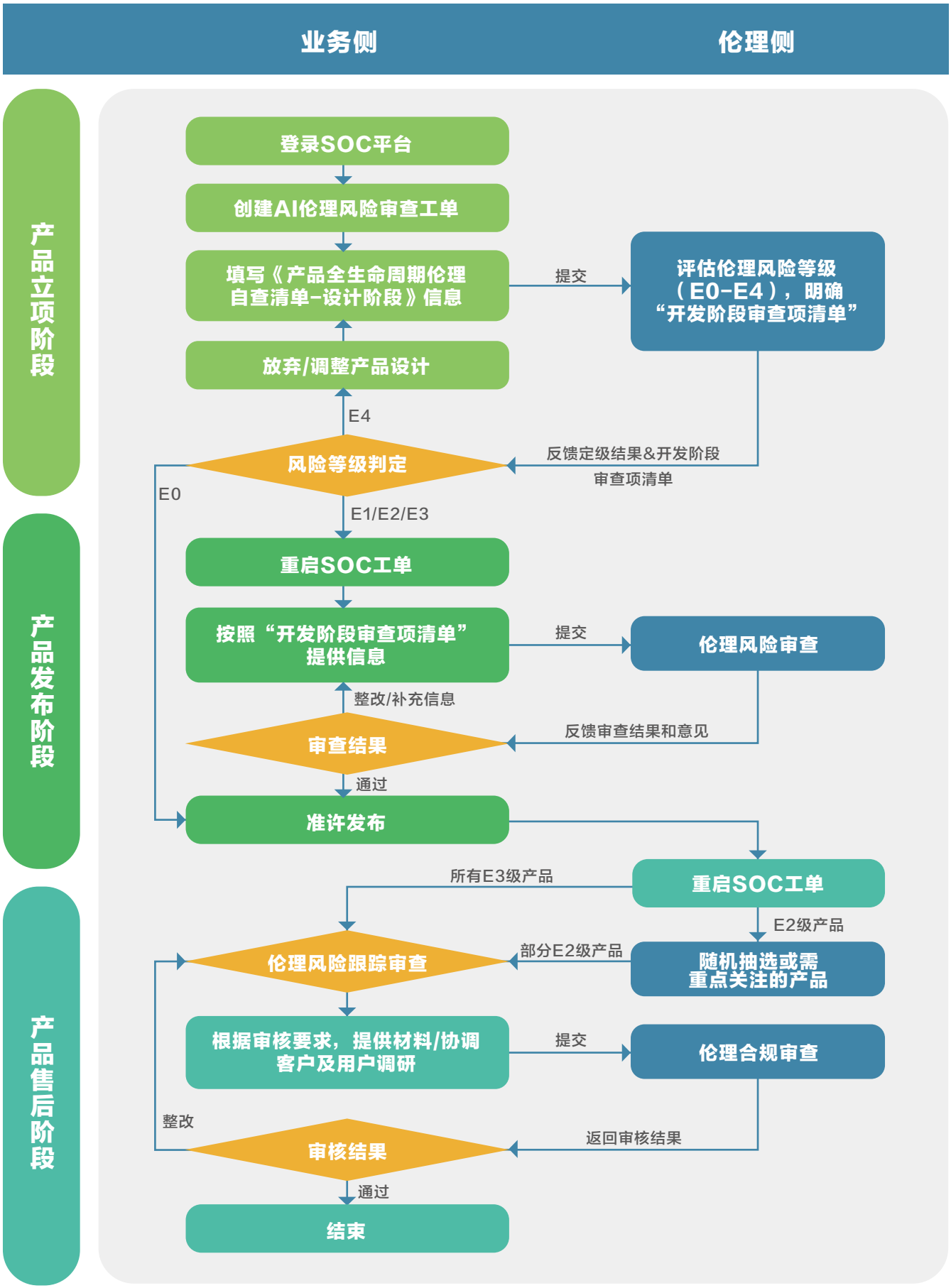
1.产品立项阶段：业务侧在SOC平台创建AI伦理风险审查工单，填写产品生命周期的伦理自查信息，提交给伦理侧进行风险等级评估（E0－E4）。高风险（E4）产品需调整或放弃设计，其余产品根据等级判定后进入后续流程。

2.产品发布阶段：业务侧根据“开发阶段审查项清单”提供信息，伦理侧进行风险审查并反馈意见。通过审查的产品方可准许发布，未通过的需整改或补充信息后复审。

3.产品售后阶段：针对已发布产品，伦理侧开展风险跟踪审查，对E2及E3类产品实施重点关注或抽查。业务侧根据审核要求提供材料或协助用户调研，伦理侧完成合规审查并反馈结果，形成闭环管理。

整个流程体现了分级分类、全生命周期嵌入、业务与伦理侧协同、闭环追踪的特点，确保AI产品在设计、开发、发布及应用全阶段均符合伦理与合规要求。





（图：产品全生命周期伦理评审流程）



第四部分 生态构建与行业认可

随着人工智能技术在全球范围的深度应用，商汤科技始终将伦理治理视为企业长期发展的基石。公司不仅在内部持续完善制度与流程，也在国内外积极参与行业生态建设，并凭借稳健又不失创新的实践获得多项权威认可，逐步形成以“标准引领、开放协作、责任实践”为核心的伦理治理格局。





4.1 国内生态构建

商汤科技AI伦理与治理委员会在推动公司内部伦理治理的同时，注重构建开放、合作的行业生态。我们通过多种方式加强与学术机构、行业组织、政府监管机构及社会各界的互动，包括：

产学研合作： 与顶尖高校、科研机构建立联合研究项目，推动伦理与技术融合研究。

标准制定与参与： 积极参与国内外人工智能伦理标准与规范的研讨和制定，贡献实践经验和技术视角。

跨界交流平台： 组织或参与行业论坛、工作坊、线上研讨等活动，分享AI伦理治理经验，促进多方交流与共识形成。

在国内，商汤科技始终以系统化、标准化的建设为核心，持续探索具有行业示范效应的治理路径。我们在产学研合作、伦理研究与标准体系建设等方面不断取得重要进展，积极推动负责任人工智能理念在国家与行业层面的落实。

A. 产学合作与研究平台建设

1、2021年3月，与上海交通大学共同成立“计算法学与AI伦理研究中心”，推动AI伦理研究的交叉创新；

2、同年4月，参与清华大学人工智能国际治理研究院首届学术委员会，商汤科技受邀成为其战略合作与发展委员会副理事长单位，助力构建中国AI治理学术与政策研究网络。

3、同年7月，商汤联创杨帆先生，受聘担任上海国家新一代人工智能创新发展试验区专家咨询委员会治理工作组顾问。



（图：2021年3月，与上海交通大学共同成立“计算法学与AI伦理研究中心”。）



（图：2021年4月，商汤受邀成为清华大学人工智能国际治理研究院战略合作与发展委员会副理事长单位。）



（图：2021年7月，商汤联创杨帆先生，受聘担任上海国家新一代人工智能创新发展试验区专家咨询委员会治理工作组顾问。）

B. 标准化与制度化布局

1、主导与参与国内外数十项人工智能伦理与治理相关标准制定，积极推动行业规范建设；

2、2024年，受邀担任中国人工智能产业发展联盟科技伦理工作组副组长单位，系统参与行业伦理治理议题和行动计划；

3、同年，委员会主席受邀担任工信部人工智能标准化技术委员会委员；

4、2025年，作为人工智能领域首批企业之一，参与科技部与国家市场监督管理总局联合发起的科技伦理（审查）委员会管理体系认证试点，为形成国家级人工智能伦理认证框架贡献可复制经验；

5、同年，委员会主席受聘担任国家人工智能标准化总体组人工智能治理专题组联合组长，在国家层面推动并引领AI治理标准与伦理规范的制定与落地，强化产业责任意识，提升人工智能技术的安全性、可控性与社会信任度。

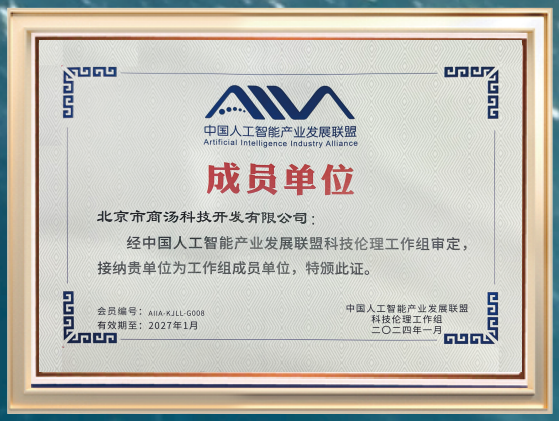


（图：2023年3月，商汤领衔参与AI伦理治理国家标准指南。）

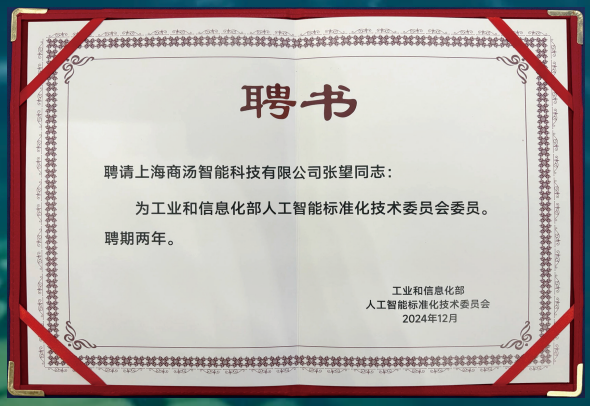


（图：2023年12月，商汤与信通院联合编制大模型可信赖报告。）





（图：2024年1月，商汤受邀成为AIIA科技伦理工作组副组长成员。）



（图：2024年12月，商汤AI伦理委员会主席受邀成为中国工信部AI标准化技术委员会委员。）



（图：2025年1月，商汤作为科技伦理（审查）委员会试点单位，率先通过科技伦理（审查）委员会管理体系认证。）



（图：2025年6月，商汤伦理委员会主席受聘担任国家人工智能标准化总体组联合组长。）

跨界交流



（图：2021年7月，薛澜教授对话商汤科技徐立：伦理治理给人工智能更大创新空间。）



（图：2021年10月，商汤员工进行伦理主题线上交流。）



（图：2021年11月，商汤CEO在第九届岭南论坛作治理主题发言。）



（图：2022年8月，商汤伦理委员会成员与中国科技伦理政策制定者们交流分享。）



（图：2022年9月，商汤科技与上海交通大学计算法学与AI伦理研究中心2022-2023年度课题开题沟通会顺利举行。）



（图：2022年12月，商汤伦理委员会主席与清华学者和优秀企业代表共论人工智能产业发展与治理。）



（图：2023年1月，商汤伦理委员会举办内外部成员线上工作研讨会。）



（图：2023年3月，商汤伦理委员会邀请上海交通大学郑戈教授为商汤内部全员作关于自动驾驶伦理主题的学术分享。）



（图：2023年7月，商汤发布SenseTrust治理平台，护航生成式AI可信发展。）



（图：2023年9月，商汤伦理委员会成员到访清华大学人工智能国际治理研究院交流会谈。）



（图：2023年7月，在上海世界人工智能大会（WAIC）商汤论坛圆桌环节，与各界重磅嘉宾们共论AI伦理。）

人物从左向右依次是：上海交通大学文科资深教授、中国法与社会研究院院长、中国计算机学会（CCF）计算法学分会会长、计算法学与AI伦理研究中心主任季卫东；清华大学公共管理学院院长、清华大学科技发展与治理研究中心主任朱旭峰；AIII人工智能国际研究院创始人与院长翁家良；三井住友保险（中国）有限公司副总经理兼首席财务官、首席投资官郑永强。）



（图：2024年6月，北京互联网法院姜颖院长一行莅临商汤指导并交流AI伦理与法律方面的工作。）



（图：2024年7月，在上海世界人工智能大会（WAIC）展区，商汤布置了面向公众的AI伦理主题展台。）





（图：2024年10月，商汤在香港举办成立十周年论坛，有幸邀请到联合国大学校长马瓦拉（Tshilidzi Marwala）博士致辞，并与国际嘉宾共论AI伦理治理。）



（图：2024年11月，商汤伦理委员会主席参加在日本东京举办的第六届中日企业法务论坛。）



（图：2024年11月，商汤伦理委员会代表参与由财新网发起的“人本智能：发展与治理倡议。）



（图：2025年4月，商汤举办网络治理开放日活动。）

4.2 国际生态构建

在全球人工智能快速发展的背景下，商汤科技持续深化与国际机构、学术界和行业伙伴的合作，积极推动跨国AI伦理治理生态建设。

2022年，联合国内外行业伙伴发起“亚洲科技促进可持续发展目标联盟（Tech4SDG）”，致力于推动人工智能等前沿技术服务联合国可持续发展目标，促进行业责任创新与区域合作。联盟目前已有40多家智库高校、企业加入，涵盖亚洲地区的9个国家和地区。

2023年，作为中国唯二的企业成员单位之一，商汤科技获颁AI Verify Foundation创始成员证书，由新加坡资讯通信媒体发展（IMDA）发起设立。同年，商汤科技AI伦理与治理委员会参与第二届博鳌亚太经济安全论坛，并联合上海交通大学承办其中的“AIGC时代可持续发展与风险管理”分论坛；

2024，联合国秘书长技术特使在外交部的陪同下来访商汤与委员会作伦理治理实践方面的交流；

2025年，公司在世界人工智能大会（WAIC）上与印度尼西亚国家政府签署《人工智能伦理、教育与创新合作框架》，开创跨国AI伦理教育与治理协作的新模式；

同年10月，我们支持联合国大学澳门研究所（UNU Macau）发布《人工智能安全治理互操作性：伦理、法规与标准》的国际政策报告。UNU校长马瓦拉（Tshilidzi Marwala）先生并向商汤科技人工智能伦理与治理委员会主席颁发纪念证书，以表彰商汤在全球AI治理与伦理研究中的贡献。



（图：2023年，商汤科技获颁 AI Verify Foundation 创始成员证书。）



（图：2023年10月，Tech4SDG联盟承办博鳌亚洲论坛全球经安论坛“AIGC时代可持续发展与风险管理”分论坛。）

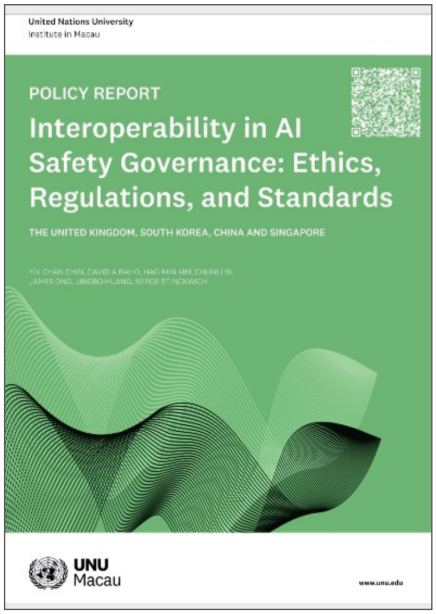


（图：2024年4月，联合国秘书长技术特使阿曼迪普·辛格·吉尔到访商汤科技并与伦理团队展开交流。）



（图：2025年7月，印尼高等教育与科技部副部长蒋黛兰（Prof. Stella Christie），印尼共和国副总统特别顾问艾哈迈德•阿迪蒂亚（Dr. Achmad Adhitya），商汤科技副总裁、人工智能伦理与治理委员会主席张望，商汤科技副总裁兼亚太业务事业群总裁史军共同签约。）





（图：2025年10月，联合国大学校长提希利迪·马尔瓦拉（Tshilidzi Marwala）先生向商汤科技人工智能伦理与治理委员会主席颁发纪念证书，以表彰商汤在全球AI治理与伦理研究中的贡献。）

4.3 行业认可与实践影响

商汤科技在AI伦理治理实践中获得行业认可，既体现了企业的责任意识，也彰显了技术与治理结合的实践价值：



（图：2021年，商汤获《哈佛商业评论》拉姆·查兰奖 奖项案例《商汤科技：人工智能的发展伦理观管理理念与实践》。）



（图：2021年，商汤获《界面》ESG先锋60-年度公司治理优秀奖。）



（图：2022年，商汤获上海日报颁发的“公司治理卓越成就奖”。）





（图：2022年，商汤的人工智能伦理治理体系和算法测评工具获AIIA联盟颁发的可信AI实践优秀案例奖。）



（图：2022年，商汤“平衡发展“的AI伦理治理实践上榜华尔街见闻公司治理榜。）



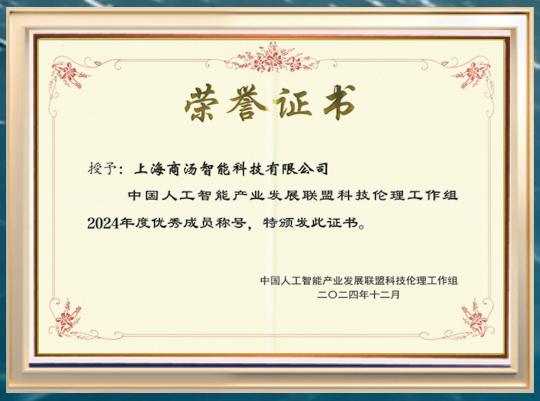
（图：2025年，商汤科技荣获“中欧企业 ESG 最佳实践大奖——数字转型最佳案例”奖。）



（图：2025年，商汤伦理治理实践获颁AI可持续发展十大卓越案例，由全国信标委与全球工业人工智能卓越中心联合颁发。）



（图：2023年，商汤可信人工智能基础设施获颁AI融合发展与安全应用典型案例奖。）



（图：2024年，商汤获颁中国人工智能产业发展联盟伦理工作组优秀成员称号。）

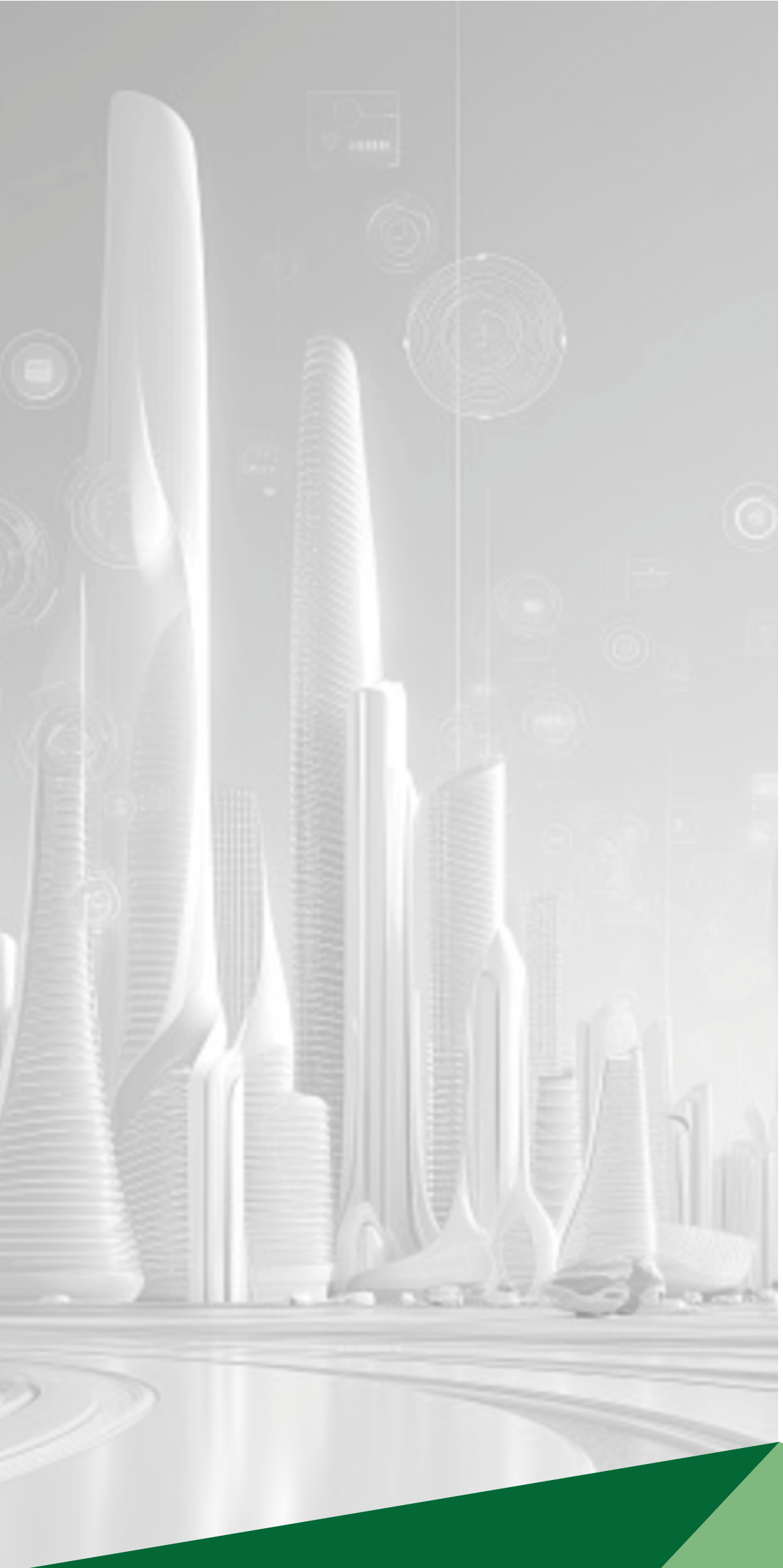


（图：2024年，商汤科技荣获联合国国际电信联盟（ITU）“AI for Good 全球峰会”“Innovate for Impact Use Case”大奖证书。）





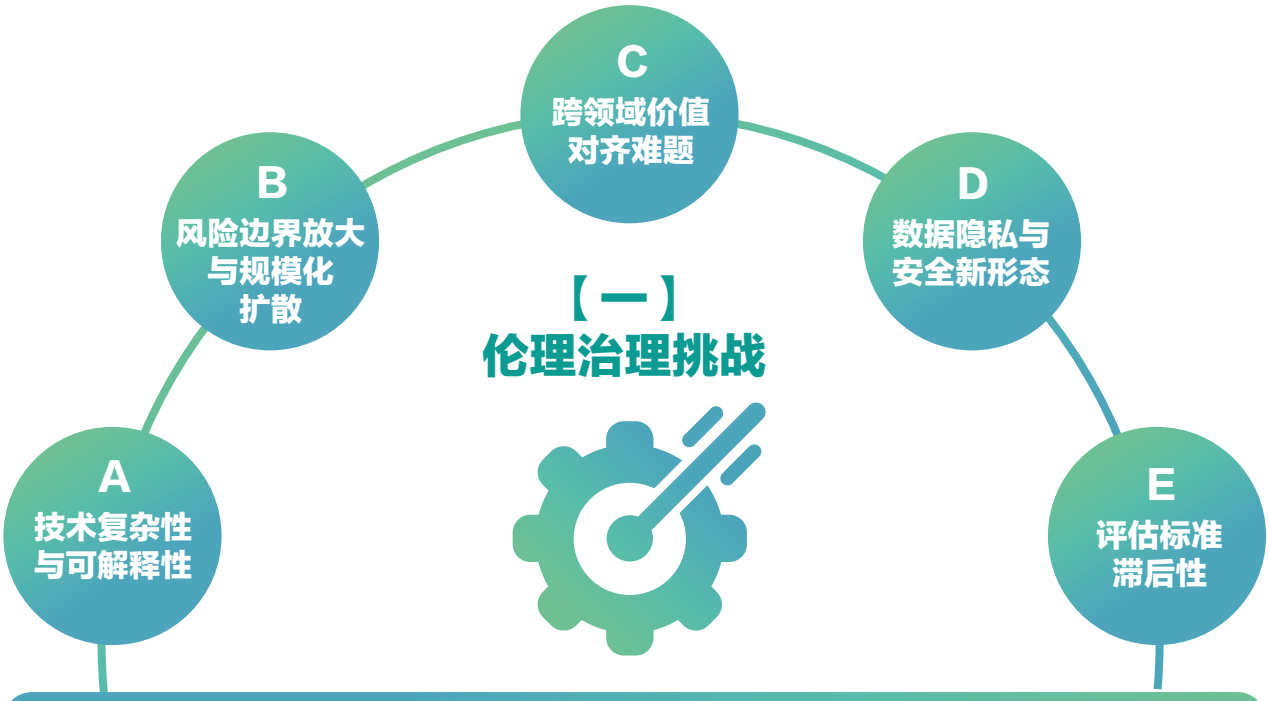
（图：2023年，商汤的产品应用获工信部重点实验室颁发的可信AI技术应用优秀案例奖。）



第五部分 未来挑战与展望

5.1 未来挑战

人工智能技术的快速发展给社会带来了巨大的机遇，同时也带来前所未有的伦理与治理挑战。在大模型时代，生成式AI的广泛应用使得风险类型更加复杂、风险边界持续扩大。商汤科技AI伦理与治理委员会在总结经验与对政策趋势的展望基础上，提出未来主要关注的挑战与发展方向：



A.技术复杂性与可解释性：随着大模型向多模态、跨领域方向发展，算法决策过程呈现"黑箱化"特征。特别是在生成式AI2.0阶段，模型从封闭场景转向开放任务、单一模态转向跨模态，使得风险防不胜防，安全机制与模型性能的平衡成为核心难题。

B.风险边界放大与规模化扩散：大模型应用场景（如创作型文本生成、广告创意、文生图等）的多元化，导致滥用风险指数级上升。风险类型从传统的隐私侵权、歧视偏见扩展到虚假误导、负面引导等新型风险，且影响范围呈规模化扩散特征。

C.跨领域价值对齐难题：在医疗、金融、军事等专业领域，大模型需要同时满足"基本价值一致性"（社会普适伦理）和"领域价值一致性"（专业伦理规范），这种双重标准使得伦理评估体系更加复杂。

D.数据隐私与安全新形态：大模型训练涉及海量多源数据，面临数据泄露、混淆攻击、投毒攻击等新型安全威胁。同时，生成内容可能包含隐含的个人隐私信息，传统的隐私保护机制面临挑战。

E.评估标准滞后性：现有伦理评测体系难以完全覆盖大模型的新型风险，特别是在虚假信息生成、诱导攻击等领域的评估标准亟待建立。

未来的人工智能治理不仅要直面技术风险的复杂性和扩散性，还需要不断更新治理理念与工具体系，为下一阶段的发展方向提供明确指引。



A.标准与法规不一致：不同国家和地区在AI伦理标准与法律框架上存在差异，导致跨境数据流动、产品应用和合规要求存在障碍。特别是在生成式AI领域，部分国家已出台规范，而另一些国家仍在探索，如何推动国际协调与互认，避免“治理碎片化”，成为全球合作的关键。

B.多方参与机制建设：企业、政府、学术界和社会组织之间的合作尚不充分，治理角色和责任边界模糊。面对快速演化的新型风险，单一主体难以独立应对，亟需搭建跨部门、跨行业的协作平台，形成技术开发者、政策制定者和社会监督力量的合力。

C.公众理解与信任：社会大众对AI的认知差异显著，部分群体存在焦虑与误解，这不仅制约了应用推广，也影响社会接受度。建立长期信任需要加强伦理教育、强化透明沟通与完善责任追溯机制，让公众更好地理解并信赖AI技术。

D.社会公平与可及性：人工智能的应用存在显著的地域与群体差异，弱势群体可能因数字鸿沟被边缘化。如何确保技术在不同国家、地区与群体间普惠共享，避免因资源不均而加剧社会不平等，是行业与社会协作需要面对的长期课题。

总体而言，行业与社会层面的治理协作需要在全球范围内形成更加统一的标准、开放的合作机制和广泛的社会共识，为人工智能的负责任发展营造良好环境。

5.2 近期应对目标（2026-2027）

制度完善与能力建设

建立并不断优化公司层面的AI伦理治理框架，细化原则、流程和审查标准，确保覆盖研发、测试、部署及应用的全生命周期。继续设立专项培训项目，提高全体员工及合作伙伴的伦理意识与风险防控能力。

场景化实践与示范

在重点业务领域（智慧城市、医疗、教育等）推出可复制的伦理治理试点，形成“从原则到落地”的成功样板。与国内外学术、行业及监管机构建立合作机制，共同推动负责任AI实践标准。

沟通与透明

发布AI伦理审查与管理报告，增强对外透明度，分享实践经验与挑战。打造公开交流平台，与社会、客户及研究社区持续对话，提升公众信任。

5.3 长期战略方向

全球化治理与标准引领

主动参与并引领国际人工智能伦理标准与治理框架的制定，贡献中国科技企业的责任实践。推动公司解决方案在全球范围内形成“安全、可信、可解释”的行业标杆。

技术与社会价值深度融合

推动前沿研究与社会责任同频共振，在算法公平性、隐私保护、环境可持续等议题上形成领先优势。通过开放生态和跨界合作，支持AI服务社会福祉，助力教育、医疗、环境等领域的长期价值创造。

韧性治理体系建设

建立具备前瞻性与自适应能力的治理体系，应对技术迭代和政策演变中的复杂风险。通过伦理风险预警和审查机制，为未来AI产品和服务提供持续的责任护航。

5.4 总结呼吁

人工智能的发展既带来巨大的机遇，也提出前所未有的伦理与治理挑战。面对技术复杂性、跨领域应用风险、数据隐私保护以及全球标准差异等多重考验，商汤科技AI伦理与治理委员会持续推动制度完善、场景化实践和社会沟通。

展望未来，公司将通过全球化治理、技术与社会价值融合以及韧性治理体系建设，不仅引领行业标准和负责任AI实践，也致力于构建开放协作的国际生态。

商汤呼吁业界、学术界、监管机构及社会各方携手，共同应对挑战、共享经验、推动人工智能在创新中守护权益、在应用中造福社会，实现技术发展与社会责任的良性共振。



商汤
sensetime

以创新驱动AI普惠
以伦理护航可信未来

商汤科技 sensetime
AI伦理与治理委员会

商汤科技AI伦理与治理委员会秘书处 梅莹
M：185 0059 5009 E：meiying@sensetime.com