舜宇光学科技（集团）有限公司

人工智能政策

一、 目的

为规范舜宇光学科技（集团）有限公司（以下简称“本集团”）人工智能相关业务活动，建立覆盖人工智能系统设计、开发、部署、使用、监测至退役全生命周期的治理框架，健全风险识别、评估与管控机制，明确治理职责与问责要求，确保人工智能应用恪守以人为本、安全可信、公平透明、尊重人权的核心原则，符合适用法律法规与国际通行标准，特制定本政策。本政策参照以下国际通行框架和标准制定：

（一）OECD AI Principles（经合组织人工智能原则）；

（二）UN Recommendation on the Ethics of Artificial Intelligence（联合国人工智能伦理建议书）；

（三）EU AI Act（欧盟人工智能法案）；

（四）ISO/IEC 42001 人工智能管理体系标准。

二、适用范围

本政策适用于本集团及其下属子公司，覆盖全体员工及各业务单元，涵盖人工智能系统设计、开发、采购、部署、使用、运营、监测、优化更新至退出管理的全生命周期各环节。适用的人工智能应用场景包括但不限于：

（一）为客户、合作伙伴开发人工智能产品、服务、系统或模型；

（二）采购、集成第三方人工智能模型、工具、平台及服务，并基于其开发产品与解决方案；

（三）在内部运营、生产制造、质量检测、供应链管理等业务场景部署、使用人工智能工具或服务；

（四）本集团内部开展的人工智能研发、落地应用、运营及管理类活动。

参与人工智能相关工作的员工及责任单位，应当根据岗位职责履行对应的合规、安全及伦理责任。本集团将负责任人工智能相关原则纳入供应商与合作伙伴管理，要求供应商、外包商、承揽商及其他价值链合作方，在与本集团有关的人工智能开发、服务交付、业务合作中，遵守本政策或同等效力规范，配合本集团开展风险管理与监督工作。

本政策与本集团现有信息安全、数据隐私、知识产权、商业道德、供应商管理等制度协同执行。本政策与本集团其他内部制度存在冲突时，以本政策为准；本政策与现行法律法规、监管要求存在冲突时，按其中要求更为严格的条款执行。

三、定义与术语

本政策中下列术语具有如下含义：

（一）人工智能系统：指利用机器学习、深度学习、自然语言处理、计算机视觉等人工智能技术，能够感知环境、学习知识、进行推理或辅助决策的系统、模型、工具或服务，包括但不限于生成式人工智能系统、深度合成系统、预测分析系统及自动化决策系统。

（二）关键人工智能系统：指涉及本集团核心业务流程、重大经营决策、客户重要权益、大量个人信息处理或可能对人身安全、财产安全产生直接影响的人工智能系统。

（三）高风险人工智能系统：指依据本政策第五章风险评估流程被评定为高风险等级，或属于法律法规明确列为高风险类别的人工智能系统。

（四）生成式人工智能：指基于算法、模型、规则生成文本、图片、音频、视频、代码等内容的人工智能技术。

（五）核心数据：指依据本集团数据分类分级管理制度被评定为核心级别的数据，包括但不限于未公开的核心技术参数、重大经营决策信息、关键客户保密信息等。

（六）重要数据：指依据本集团数据分类分级管理制度被评定为重要级别的数据，一旦泄露、篡改或丢失可能对本集团经营、客户权益或社会公共利益造成较大影响的数据。

四、风险管理

1. 风险评估时机

人工智能系统在以下节点必须开展风险评估：

（一）立项阶段：对拟开发或采购的人工智能系统进行初始风险评估；

（二）上线前：对已完成开发或集成的人工智能系统进行上线前风险评估；

（三）重大变更：系统功能、适用范围、数据来源或模型版本发生重大变更时进行风险评估；

2. 上线审批与退役管理

人工智能系统上线前须完成风险评估、合规审查与安全测试，提交审批流程，经对应权限审批后方可正式投入使用。未经审批的人工智能系统不得接入本集团生产环境或面向客户提供服务。

人工智能系统退役或下线时，责任部门应当提交下线申请，说明退役原因、数据处置方案与业务替代安排，经审批后执行。退役过程中须妥善完成数据归档或销毁、模型文件清理、相关权限回收，并更新系统登记台账。

3. 模型运行监控与漂移管理

人工智能系统上线后，责任部门须建立持续性能监控机制，定期检测模型输出准确性、稳定性和可靠性。当检测到模型性能退化或数据分布偏移（模型漂移）时，须及时启动纠正措施，包括但不限于：模型重训练、参数调整、数据集更新或功能降级运行。

关键人工智能系统须建立自动化监控指标和预警阈值，当关键性能指标低于预设标准时自动触发告警和人工审查流程。监控记录、漂移事件及处置措施须完整归档留存。

五、 伦理规范

本集团恪守公平、公正、非歧视、以人为本的基本原则，尊重人格尊严、合法权利与人的自主地位，保证人类对人工智能系统具备监督、干预、管控能力。

在人工智能系统全生命周期内持续识别、评估、缓释偏见风险。通过优化数据质量，提升数据样本的代表性与多样性，减少由数据、模型、算法带来的不公平输出结果，避免复制、加剧、放大既有偏见。

对于关键人工智能系统、生成式人工智能系统以及第三方人工智能系统，一旦输出不公平、带有歧视或不合理结果，须及时采取修正优化、功能限制、暂停使用等处置措施；必要时为受影响主体提供反馈、申诉、复核渠道。

六、 人工智能治理架构

本集团建立分层负责的人工智能治理架构，确保 AI 治理决策、执行和监督职责清晰明确。

1. 决策层

本集团董事会（或董事会下设的 ESG/可持续发展委员会）对人工智能治理承担监督职责，负责审批 AI 政策、审议重大 AI 风险事项、监督 AI 合规执行情况。

2. 管理层

本集团管理层（由信息安全管理部门牵头，联合技术中心、法务部、人力资源部等相关部门）负责 AI 政策的落地实施、风险评估组织、培训推进和持续改进。

3. 执行层

各业务单元设立人工智能合规联络员，负责本单元 AI 系统登记、风险评估申报、日常合规检查和事件报告。

七、决策管控

人工智能系统仅承担辅助决策角色，其输出可能存在错误、偏差及非预期结果。本集团结合业务场景、风险等级、自动化程度、对个人权益的潜在影响，建立完备的人工监督与干预控制机制。

针对员工招聘筛选、绩效考评、供应商准入评估、客户信用评估等涉及个人或组织重大权益的自动化决策场景，保障决策过程透明、结果可解释。自动化生成的决策必须开放人工复核通道，相关个人有权要求对决策作出说明，有权拒绝仅依靠自动化生成的决定。定期开展算法公平性评估，防范训练数据偏差引发算法歧视问题。

在生产制造场景中，人工智能系统用于质量检测判定、工艺参数推荐、设备故障预警等环节时，关键判定结果须经人工复核确认后方可执行；涉及安全联锁的人工智能输出不得直接触发停机或危险操作，须设置人工确认与安全冗余机制。

八、透明度与可解释性

本集团持续提升人工智能系统的透明度与可解释性。

（一）透明度：清晰公示人工智能的使用场景、能力边界、局限性、数据来源与潜在风险；

（二）可解释性：结合不同利益相关方诉求，针对人工智能输出提供通俗易懂的依据说明。

当用户交互对象为人工智能而非自然人，或涉及生成式人工智能、深度合成等容易造成认知混淆的内容时，本集团按照《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》等法律法规及场景要求，履行说明、标识、提示义务，降低误导、混淆及不当传播风险。

关键人工智能系统须建立完整文档留存，记录系统能力、预期用途、适用范围、局限性、潜在偏见与风险。如为本集团自主研发训练的人工智能系统，还应当记录数据来源、

数据采集标注、算法逻辑、训练、测试验证全流程；系统功能、适用范围、版本发生重大变更时同步更新文档。

在适用范围内，向终端用户、受影响个人及责任岗位人员说明人工智能输出或决策的生成逻辑，包含输出依据、数据来源、置信水平、关键影响因素等。关键人工智能系统需要支持结果溯源，可定位至输入数据、模型行为、核心决策逻辑，用于监控、复核、纠错与问责。

九、责任与问责

本集团明确人工智能全生命周期内各参与主体的角色与权责，人工智能输出结果、风险处置、合规事项均落实责任归属，责任可追溯至具体责任部门、人员或法律实体。

建立监控反馈、事件调查、纠正处置及补救机制。及时识别、调查人工智能带来的错误、损害及负面影响；针对受人工智能影响的决策，厘清法律责任与管理责任，结合影响严重程度采取整改、风险补救、追责问责等手段。

选用第三方人工智能产品、模型、工具、服务时，应当评估第三方机构的专业能力、合规水平、安全保障能力与风险承担机制，在合同中明确双方权责边界，搭建监督、纠偏、补救、业务退出机制。

十、人工智能数据安全防护

本集团按照保密性、可用性、完整性要求，对人工智能数据资产分级管控。数据分类分级标准遵照本集团《数据分类分级管理制度》执行。

针对人工智能数据安全，执行以下专项管理规则：

（一）严禁将核心数据、重要数据投入公有云人工智能服务器开展训练与交互；

（二）训练数据必须来源于合法授权渠道；外部引入数据需要完成安全检测，规避恶意数据污染；

（三）采集用户数据严格落实最小必要原则，不收集业务无关的个人信息；

（四）不得非法留存用户输入数据；

（五）数据出境严格遵守《数据出境安全评估办法》《个人信息保护法》等相关规定。

本集团在人工智能系统设计、开发、测试、部署、运营和退役全生命周期中，严格遵守《个人信息保护法》《数据安全法》及业务经营地适用的数据保护法律法规。在 AI 系统数据处理活动中，落实以下隐私保护要求：

（一）开展数据隐私影响评估（DPIA），在 AI 系统立项阶段识别个人信息处理活动对个人权益的影响；

（二）建立个人信息处理活动的合法依据审查机制，确保 AI 系统处理个人信息具有明确的法律基础；

（三）保障数据主体权利，包括知情权、访问权、更正权、删除权和反对自动化决策权；

（四）对涉及敏感个人信息的 AI 应用，实施强化保护措施，包括数据脱敏、匿名化/假名化处理和访问权限严格管控。

本集团对人工智能系统实施以下网络安全专项措施：

（一）定期开展渗透测试，识别和修复 AI 系统及接口的网络安全漏洞；

（二）建立 AI 安全事件响应机制，制定应急响应预案，明确事件分级、报告流程和处置时限；

（三）对 AI 训练数据、模型文件和 API 接口实施加密保护，采用安全编码实践防范对抗性攻击、数据投毒和模型逆向工程；

（四）部署 AI 内容安全防护机制，包括提示词防护盾（prompt shield）等，防止有害或不当内容生成。

十一、能力界定与应用限制

本集团明确各人工智能系统授权使用范围、能力边界与固有局限，文档化记录预期用途、决策边界、输入输出约束、技术限制；建立使用许可清单、禁止应用清单、安全护栏、能力约束、版本升级管控节点，规避系统误用、业务越界、非预期场景调用，确保人工智能运行不超出预设范围。

本集团定期评估更新各业务板块人工智能许可及禁止应用范围，对每个人工智能系统书面明确预期用途。未经评估、审批及授权，不得将人工智能系统用于原定用途之外、未完成验证或存在重大风险的场景；配套持续监控与责任追溯机制。

本集团对面部识别、情绪分析、生物特征识别、大规模监控等敏感 AI 能力实施严格访问控制：

（一）建立敏感 AI 能力清单，明确各项敏感技术的部署审批权限和监督机制；

（二）敏感 AI 能力的使用须经专项风险评估和高层审批，未经授权不得部署；

（三）对敏感 AI 能力的使用人员实施权限管控和操作审计，确保可追溯。

十二、禁止行为

本集团禁止研发、部署、使用以下类别的人工智能系统（参照 EU AI Act 及国际通行标准禁止类别），防范侵害个人权利、人身自由、安全及社会公平，严格遵守中国及业务经营地关于人工智能的法律法规，包括但不限于《网络安全法》《数据安全法》《个人信息保护法》《生成式人工智能服务管理暂行办法》《互联网信息服务深度合成管理规定》等。

禁止部署的 AI 系统类别包括：

（一）使用潜意识、操纵性或欺骗性技术，对个人行为产生实质性扭曲影响的 AI 系统；

（二）利用个人因年龄、残疾或特定社会经济状况的脆弱性，对个人行为产生实质性扭曲影响的 AI 系统；

（三）用于对社会行为或个人特征进行随时间评估或分类，导致在不相关情境下受到不利对待的社会评分 AI 系统；

（四）在公共场所实时远程生物特征识别系统（未经授权的生物识别监控）。

1. 内容生成类禁止

禁止利用人工智能系统制作、复制、发布、传播含有以下内容的信息：

（一）违反宪法基本原则，危害国家安全、荣誉与利益；

（二）煽动颠覆国家政权、推翻社会主义制度，煽动分裂国家、破坏国家统一；

（三）宣扬恐怖主义、极端主义，鼓吹民族仇恨、民族歧视；

（四）传播暴力、淫秽色情信息，编造、散布虚假不实信息。

2. 数据与权益类禁止

禁止利用人工智能系统从事以下行为：

（一）侵害他人名誉权、隐私权、知识产权以及其他合法权益；

（二）泄露国家秘密、工作秘密、商业秘密；

（三）诱导、骗取国家秘密、工作秘密、商业秘密或者个人敏感信息；

（四）法律法规明令禁止的其他情形。

若人工智能系统出现上述禁止情形，应及时限制、暂停乃至终止使用，将该系统移出人工智能管理体系，同步开展风险评估、整改处置及责任追踪。

十三、 意识与素养

本集团推进人工智能文化建设，将人工智能伦理、安全、风险意识纳入整体治理与培训体系，持续提升员工及合作相关方对人工智能社会影响、技术风险、安全实践的认知水平。

本集团将人工智能数据隐私保护、科技伦理设置为全员必修课程，依托内部平台完成学习、考核与档案留存；结合内外部科技伦理与风险案例常态化宣导，强化全员风险意识与责任观念。

推行分层分类人工智能培训，匹配不同岗位、应用场景、专业能力。从事人工智能开发、采购、部署、运营、管理以及高频使用的人员，必须接受岗位匹配的专项培训，经考核合格后方可履职。

定期复盘人工智能培训落地情况，做好记录归档；持续迭代更新培训内容，适配人工智能技术演进、伦理标准迭代以及法律法规监管要求。

十四、 申诉、事件管理与纠正措施

本集团对违背人工智能科技伦理、违反安全使用要求的行为秉持零容忍。发现疑似违规行为、风险事件或不良影响时，立即启动内部调查、原因溯源、影响评估；根据事件严重程度采取暂停使用、访问限制、风险缓释、整改、追责等措施；重大事项提交本集团管理层审议。

搭建透明、易触达、公平高效的申诉澄清流程。受到人工智能决策、输出结果不利影响的个人或组织，对结果、处置流程存在疑问、异议的，可通过本集团指定渠道发起质疑、申诉、澄清或者人工复核申请。本集团依法受理处置，保障相关主体拥有意见表达、申请复核的权利。

任何单位或个人发现涉嫌违法处理数据、侵犯隐私、违背科技伦理、不当使用人工智能的行为，均可向本集团职能部门投诉举报或申诉。本集团遵照法律法规与内部制度处理诉求，原则上自收到申请之日起 15 个工作日内响应并反馈处理结果或进展；情况复杂的可延长至 30 个工作日，并将延长理由告知申请人。

申诉邮箱：bgs@sunnyoptical.com

十五、 人工智能环境可持续性

本集团高度重视智能化业务快速发展带来的资源环境挑战，始终坚持绿色低碳发展理念。我们通过推动基础设施能效优化、加强运营环节环境数据监测、完善内部评估改进机

制，持续降低业务生态足迹。同时，本集团将环境保护要求延伸至供应链管理，在合作伙伴筛选中关注其环境表现。本集团致力于将生态责任融入日常运营与长期战略，实现经济效益与气候责任的平衡统一。

十六、 附则

（一）本政策未尽事宜，遵照国家法律法规、外部行业规范以及本集团内部制度执行。

（二）本政策由本集团信息安全管理部门负责解释。

（三）本政策经 CEO 批准后正式实施；政策修订亦按本流程执行。

（四）本集团积极推进人工智能管理体系标准化建设，计划依据 ISO/IEC 42001 标准建立 AI 管理体系，并适时申请第三方认证。