

聚焦 博物馆数字化

AI信息幻觉对博物馆业务的影响与应对策略

截至2025年，生成式人工智能（Generative AI）已深度渗透至文化领域的各个层面。博物馆作为人类文明记忆的殿堂与知识传播的权威机构，正积极探索利用大型语言模型（LLM）等技术提升学术研究效率与公众体验，从辅助学者整理海量文献，到为展览自动生成解说词，AI的应用前景广阔。然而，技术赋能的背后潜藏着严峻的挑战，其中“信息幻觉”（Information Hallucination）——即AI模型生成看似合理但与事实不符、凭空捏造或存在逻辑谬误的内容，构成了对博物馆价值的负面影响。

博物馆的核心价值建立在“信息真实性”之上，AI信息幻觉直接冲击这一基石。博物馆的学术研究、陈列展览、信息传播与社会教育四大核心业务，对信息真实性与专业性有高要求，也成为AI信息幻觉的高发场景，且错误形态呈现领域差异化特征。AI信息幻觉类型虽因业务领域而异，但其负面影响已渗透至机构功能实现、公信力维护与公共价值传递的全链条，形成多维度危害。

博物馆的公信力基石在于对“真实性”的坚守。无论是展柜中历经沧桑的文物，还是展墙上严谨考证的文字，都承载着机构的学术权威和公众的信任。AI生成文本中一旦出现信息幻觉，哪怕是细微的错误，都可能误导观众、扭曲历史叙事，并最终侵蚀作为知识权威机构的博物馆的社会地位，必须从信息素养、数据伦理和风险规制的角度进行系统性规避。

博物馆核心业务领域的AI信息幻觉影响

学术研究领域的隐蔽性错误，侵蚀机构学术权威与公信力

学术研究作为博物馆知识生产的基础，AI信息幻觉以高度专业化形式隐蔽存在，对研究严谨性构成系统性威胁。学术研究领域的虚假文献与错误数据，若被纳入机构研究成果发布，将导致学术声誉受损。

其一，虚构学术资源与数据，在文物考古报告整理、文化遗产研究综述撰写中，AI会生成格式规范却完全不存在的学术文献，或编造数据支撑错误结论，导致研究成果根基失效。

其二，错误关联与归因谬误，在文化遗产谱系构建中，AI会基于概率关联而非史实逻辑，生成文物与历史事件的虚假联系。

其三，文化遗产修复方案的误导性建议，AI若缺乏专业修复数据训练，会生成看似科学实则有害的方案，直接威胁文化遗产本体安全。

陈列展览领域的公共性错误，偏离客观事实损害公众信任

陈列展览作为博物馆面向社会公众的核心展示窗口，AI信息幻觉所产生的影响呈现出直接性与广泛性特征，陈列展览的信息错误与叙事矛盾，会直接误导公众认知，公众对机构“专业把关能力”的信任度会下降，对机构公信力构成损害。

其一，展品信息事实性失真，AI在生成展品标签、单元解说词时，易出现关键信息错误。美国纳什尔博物馆早期AI策展实验已证实此类风险，AI曾生成包含“非馆藏”虚构艺术品的展览清单，暴露基础信息失控隐患。

其二，文化语境歪曲与误读，AI因缺乏对特定文化的深度理解，易简化或扭曲文物文化内涵，造成文化扁平化解读，进而消解文化多样性价值。

其三，展览叙事逻辑断裂，在大型特展叙事框架构建中，AI会生成前后矛盾的内容，导致叙事框架混乱，损害展览专业性。

信息传播领域的扩散性错误，扭曲文化传承与历史记忆

博物馆承担着文化传承与历史记忆建构的使命，AI信息幻觉会导致文化意义的误读与历史事实的篡改。AI的网络传播会将错误的历史信息引入，数字化传播的即时性与互动性会加剧AI信息幻觉的影响扩散、扩大错误信息的传播范围。

其一，社交媒体内容“张冠李戴”，AI处理海量信息时易出现基础错误，如混淆同名历史文化名人、配错文物图片等。

其二，语音导览与虚拟讲解的虚构叙事，AI为增强内容吸引力，会在历史背景描述中“添油加醋”，使公众将文学性叙事当作历史事实记忆。历史叙事的虚构，篡改了真实历史，易导致年轻一代形成错误的历史认知。

其三，VR/AR展项中的“幽灵展品”，在沉浸式虚拟展览中，AI为填充场景，会自动生成真实存在的事物，模糊“历史真实”与“技术再现”的边界，使公众难以区分文化遗产的原始意义与技术生成的附加意义，最终导致文化传承的“失真”。

社会教育领域的长效性错误，加剧公共文化服务不平等

博物馆的社会教育是其公共价值的重要体现，AI信息幻觉的影响具有持续性，直接作用于受众认知建构。AI信息幻觉会通过“信息差”进一步加剧公共文化服务的不平等。

其一，低信息素养群体（如老年人、青少年）更易轻信AI生成的错误内容，如青少年依赖AI完成历史作业，因幻觉内容出现学术错误；老年人通过非专业机构的AI语音导览获取文化信息，误将虚构传说当作史实。

其二，文化边缘化群体会因AI训练数据的偏见，遭遇

更严重的信息误导，如少数民族文化在AI生成内容中被简化或曲解，使其群体成员难以获取本民族文化的准确信息，进一步加剧文化知识鸿沟。

其三，教育材料的文化偏见固化，在全球文化课程中，AI会因训练数据的西方中心倾向，生成不平等的内容结构，强化文化霸权认知。

博物馆AI信息幻觉的应对策略

针对AI信息幻觉的多维度风险，博物馆需构建“技术防控—流程规范—制度保障—素养提升”的系统性应对体系，实现风险的全生命周期管控。

技术层面：构建领域化可信数据与模型优化机制

技术防控是从源头降低信息幻觉风险的核心，需聚焦机构业务特性优化AI应用技术路径。

其一，构建垂直领域可信知识库，将博物馆馆藏信息、研究成果、权威文献等进行数字化、结构化处理，形成闭环的专业知识库。通过检索增强生成（RAG）技术，强制AI模型在生成内容时仅从该库获取信息，切断对不可靠公共数据的依赖。

其二，精细化提示工程，培训工作人员掌握专业提示词设计技巧，通过明确约束条件减少模型“创造空间”。有效的提示词应包含具体上下文、输出要求，针对高风险场景开发提示词模板，固定约束性条款。

其三，多模型交叉验证，对关键内容采用2至3个不同架构的AI模型进行平行生成，若结果差异率超过预设阈值，则启动人工介入机制，避免单一模型的偏见影响结论。

流程层面：建立人机协同的审核闭环

流程规范是确保技术防控落地的关键，需通过刚性机制明确AI与人类的权责边界。

其一，确立“AI辅助—人工终审”的工作流程。在学术研究领域，AI生成的文献综述需经研究员交叉验证文献真实性与数据逻辑性；陈列展览领域，解说词需经策展人审核并与馆藏档案比对；社会教育领域，课程内容需经教育专家评估适切性。以展陈文本为例，规范流程为：策展人提供核心资料→AI生成初稿→策展助理对照馆藏档案校验事实→策展人审核叙事与语境→总策展人批准。

其二，分级审核机制，根据内容重要性实施差异化审核：高风险内容经2名以上专家交叉审核；中风险内容1名专业人员审核并记录；低风险内容可采用“自动化校验+抽样人工审核”。

其三，错误追溯与快速响应机制，建立AI生成内容的错误反馈渠道，鼓励公众、员工、专家报告问题，收到反馈后启动核查，完成更正并公开说明，同时记录错误案例作为模型优化与培训素材。

制度层面：完善AI应用治理框架

制度保障是应对策略长效性的基础，需从伦理、责任、透明度等维度构建治理体系。

其一，制定《AI使用规范指南》，明确AI应用的伦理边界（如禁止生成可能引发文化误解的内容）、操作规范（如知识库更新频率）、责任划分，特别规定敏感文化议题（如宗教、民族、革命历史）的AI应用需额外经过伦理审查。

其二，建立AI伦理审查委员会，由机构业务骨干、外部专家组成，定期评估AI应用的潜在风险，审议高争议性内容生成方案，为指南修订提供建议。

其三，透明度管理制度，向公众公开AI应用范围与审核流程，在AI生成的内容中标注“本内容由AI辅助生成，已通过专业审核”，并提供反馈入口，体现机构负责任态度，同时培养公众批判性认知。

素养层面：提升全员AI风险认知与应对能力

人员素养是风险防控的最后防线，需通过分层培训，强化员工与公众的风险意识。

其一，员工分层培训，对研究人员重点培训文献真实性核查方法、多模型验证技巧；对策展人员强化文化语境解读能力、叙事逻辑校验能力；对教育人员提升儿童内容适切性判断、偏见识别能力，定期开展案例研讨，增强风险敏感度。

其二，公众AI素养教育，通过机构官网、展厅提示等渠道，告知公众“AI内容可能存在误差”，引导其结合权威资料交叉验证；开展“AI与文化解读”科普活动，教授识别信息幻觉的方法。

结语

今天，生成式人工智能正以前所未有的力量重塑知识生产与传播的范式。生成式AI为博物馆的创新发展提供了技术红利，但信息幻觉风险对机构核心价值构成挑战。从学术研究的隐蔽性错误到社会教育的长效性影响，从公信力侵蚀到文化传承失真，AI信息幻觉的危害已超越技术层面，触及机构公共使命的本质。对博物馆而言，这既是活化馆藏资源、提升公共服务能力的巨大机遇，也是一场关乎机构公信力的严峻考验。

应对这一挑战的出路绝非因噎废食地拒绝技术，而是必须采取审慎、主动、系统化的风险规避策略，其核心在于博物馆需摒弃“技术万能”的认知误区，在拥抱技术的同时坚守“真实性”底线，通过技术层面构建垂直领域知识库和优化模型应用方式，结合治理层面制定清晰的权责清单、工作流程和伦理指南，构建“人机协同、以人为本”的风险防控体系，在享受AI带来的效率提升的同时，将信息幻觉的风险控制在可接受的最低水平。

技术的角色是增强人类智慧，而非取代人类判断，研究者的严谨、策展人的智慧以及文化遗产相关专业人士对于历史与文化的深刻洞察，是AI无法替代的。唯有如此，才能在智能时代充分发挥AI的赋能作用，持续履行文化传承、历史记忆建构与公共教育的核心使命，为公众提供高质量、可信的文化服务。

（作者单位：九江市博物馆
赣州市博物馆）



仅增强视觉节奏，也通过色彩的情感属性深化观众对展览内容的理解。

注重整体协调统一，强化细节表现，凸显展览主题与叙事结构

展览陈列通过色相选择、明度调控和区域配色的系统规划，有助于构建具有逻辑与节奏感的视觉叙事。中国共产党历史展览馆举办的“为了共同的事业——从西班牙战场到中国抗日战场”等专题展览，在色彩色调实践运用方面做了积极尝试与创新。

结合展陈主题主旨，明确视觉焦点，强化视觉冲击与引导

色彩由色相、明度、饱和度三大属性构成，其组合与对比可形成多样化色调系统。在博物馆陈列实践中，色彩色调的表现力决定着视觉焦点的清晰度，能够产生强大的视觉冲击力和感染力，需要注意的是，这种聚焦效果不仅依赖单一色彩属性，更在于系统性的配置与关系构建。高对比度的色彩组合（如红与黄、蓝与白）能够通过强烈的视觉张力迅速捕获观众的注意力，并引导视线流动。同时，主色调的确立、辅助色的配合以及点缀色的搭配使用，共同建立起清晰的视觉层级与节奏秩序，使核心参展项在无需外界提示的情况下仍能被自然感知。例如，2021年7月，在中国共产党历史展览馆举办的“不忘初心、牢记使命”中国共产党历史展览，以红色为主调、搭配金黄色党徽与立体文字的视觉策略，既强化了视觉中心，也隐喻了中国共产党的政治底色与精神血脉。

今年8月举办的“为了共同的事业”专题展览，反映西班牙反法西斯斗争开始后，在中国共产党积极倡导和支持下，旅欧、美、亚等地100多名共产主义战士，组成国际纵队中国支队，与各国勇士并肩作战，以及在西班牙反法西斯斗争结束、中国全民族抗战爆发后，部分国际纵队成员奔赴中国抗日战场，同中国人民一道，以顽强的意志和英勇的斗争，彻底打败日本法西斯的光辉历程。该展览的叙事时间轴为1936年至1945年，主题脉络为国际纵队中国支队以及国际友人辗转西班牙战场和中国抗日战场共同抗击法西斯的伟大壮举。基于此，策展团队以“中国共产党”“国际纵队中国支队”和“国际友人”作为核心视觉焦点，并分别赋予红色与深蓝色作为焦点标识色。这两种色彩不仅在地域象征（中/西）上形成强烈对比，也强化了在民族独立、人民解放这一“共同事业”中的多元力量参与、携手团结主题。

打破单一色彩范式，优化区域配色，唤起情感共鸣

色彩不仅作用于视觉，更深入影响心理与情绪。暖色调常引发积极热烈的情感，冷色调导向冷静沉思；高饱和色彩带来兴奋，低饱和色调伴随庄重。此外，色彩作为文化符号，还能唤起集体记忆与文化联想，成为情感传达的有效媒介。在展陈语境中，合理运用色彩色调，能够塑造特定的情感氛围，增强展览的表现力与感染力，使观众在情感认识中深化对展览内容的理解。

“为了共同的事业”专题展览分为四个部分：国际纵队与西班牙人民并肩奋战、国际纵队中的中国支队、从“西战场”到“东战场”奔赴中国抗日前线、永远铭记。从逻辑体例上看，四个部分既相对独立又环环相扣，为了在视觉上增强展览的层次感，丰富观众感官体验，策展团队在色彩搭配上打破单一色彩范式，实施“分区配色、渐进过渡”的策略，以贴合叙事逻辑并增强情感表达：第一部分以深蓝色为主进行渲染，力求再现89年前西班牙人民反抗法西斯的情境，引导观众感受世界正义和平力量的汇聚；第二部分融合红色与深咖色，凸显中国共产党在大敌当前的情况下仍具胸怀天下的伟大气魄，体现中华儿女奋不顾身的国际主义精神；第三部分借助土黄与深蓝，分别喻示中国战场的艰苦与国际友人的援手；第四部分以亮黄色为主，展示胜利的喜悦，传递出友谊与铭记。如此分区不



本文配图：“为了共同的事业——从西班牙战场到中国抗日战场”专题展场景

色彩色调在陈列展览中的运用

赵骄健

值得一提的是，展览中还根据内容归属（如国际背景、中共行动、日军暴行）分别匹配不同的色彩。这一细致的色彩信息系统不仅优化了观众认知流线，凸显了中国共产党在世界反法西斯战争中的历史地位与主要作用，也强化了展览的主题一致性与叙事深度。

色彩色调作为一种核心的展陈语言，兼具视觉表达与情感传达的双重功能。在博物馆展览实践中，应立足于主题叙事与情感目标，通过对色彩属性的科学配置与艺术化处理，构建层次丰富、逻辑清晰且情感充沛的视觉场景，从而实现阐释、导览与共鸣的多重展陈使命。

（作者单位：中国共产党历史展览馆）