

天问研究所

TIANWEN INSTITUTE · BY NANKONG

触见名画

AI无障碍视觉描述实验室

TOUCH · LISTEN · UNDERSTAND

天问研究所出品 · 高中生AI向善研究与实践

# 触见名画

AI无障碍视觉描述实验室  
完整研究、产品与实施方案

作者：Nankong · 非博物馆官方服务

Copyright © 2026 Nankong. 保留所有权利

2026年8月

# 不是让AI“看图说话”，而是让每句话可核验

本项目研究一个具体问题：经过来源锁定、事实与解释分离、分层重写和人工审核后，AI辅助生成的视觉描述，能否比普通展签更清楚地传达名画中的空间、动作、色彩和关键细节？

用户靠近作品旁的NFC标签，手机识别统一馆藏号，打开对应作品；经审核的讲解以“15秒定位、45秒观察、90秒理解”三层呈现，并用本地原创女声朗读。

6

件开放版权名画

3层

渐进式视觉描述

1套

AI发布质量门禁

## 项目成果

- 可安装、可离线的无障碍网页应用，支持安卓Web NFC和通用深链接。
- 六件真实公共领域馆藏、六段本地原创语音、结构焦点与馆藏来源。
- 普通展签与分层描述的被试内A/B实验、匿名数据导出和分析模板。
- 可迁移到博物馆App、支付宝小程序或馆方服务端的讲解重构Skill。

**研究边界：**未邀请真实盲人或低视力用户参与前，不宣称产品已经验证视障使用效果；普通参与者实验只测理解度，不替代残障者共同设计。

# 切入点：讲解内容存在，但“看见方式”仍不平等

## 普通展签的限制

作者、年代和媒介有助于识别作品，却很少告诉无法直接看清画面的参观者：人物在哪里、谁在做什么、视线如何移动、最关键的细节是什么。

## 生成式AI的机会

AI可以快速整理信息、改写长度、生成多种版本和语音，但它也可能把心理、身份、象征和情节猜成事实。对无障碍导览而言，这类错误不是“文风问题”，而是信息可靠性问题。

### 联合国教科文组织

2026年发布《中国博物馆无障碍指南》，强调从实体环境、信息传播和服务设计等方面改善文化参与。

### WCAG 2.2 AA

网页以键盘可操作、文字替代、清晰焦点、足够对比、可调整显示和减少动态为目标。

### 博物馆开放获取

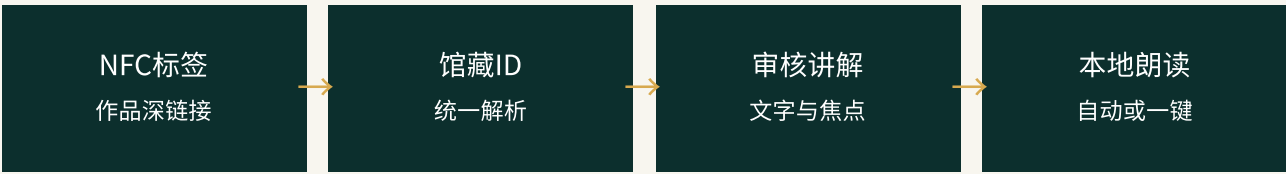
首版只使用大都会艺术博物馆明确标注公共领域的作品图像与可追溯馆藏页。

**研究假设：**与只提供作者、年代、媒介的普通展签相比，分层视觉描述将提高空间关系与关键信息理解正确率，并降低主观理解负担。

## 三个可检验问题

1. 分层描述是否改善关键信息回忆和空间关系判断？
2. AI原稿经过质量门禁后，事实错误、无依据解释和遗漏是否减少？
3. 触碰导览是否能减少“查找作品—寻找讲解—启动播放”的操作步骤？

# “碰一碰” 导览的真实系统链路



## 用户旅程

- 1 首次进入：**用户打开PWA并主动开启触碰导览， 安卓Chrome请求一次NFC权限， 同时解锁网页音频。
- 2 靠近作品：**系统读取NDEF中的链接或馆藏号， 统一解析为MET馆藏ID。
- 3 自动切换：**页面打开对应作品， 优先播放已审核的本地音频， 同时显示同步文字。
- 4 继续探索：**用户可切换15秒、45秒、90秒描述， 或进入下一件作品。

## 不同设备的可实现边界

| 环境                    | 能力                                | 回退方式              |
|-----------------------|-----------------------------------|-------------------|
| 安卓Chrome / PWA        | 用户授权且页面保持开启后，可连续读取Web NFC并自动切换讲解。 | 二维码、馆藏列表。         |
| iPhone及不支持Web NFC的浏览器 | 被动NFC标签打开作品深链接；系统可能限制自动播放。        | 页面提供单一大号“播放讲解”按钮。 |
| 博物馆App或支付宝小程序         | 未来可调用原生NFC、蓝牙信标或馆方室内定位SDK。        | 复用同一馆藏ID和讲解接口。    |

网页不能在所有手机上无条件后台读取NFC。本项目明确写出授权、前台运行和自动播放限制，不把模拟按钮当作原生能力。

# 最核心能力：AI讲解重构Skill

它不是一个“更会写文案”的提示词，而是一条带来源、字段、禁用项、审核状态和发布门禁的内容生产流水线。

**1 Acquire 获取：**从馆方接口读取馆藏ID、标题、作者、年代、媒介、现有讲解、版权、更新时间和原始链接。

**2 Separate 分离：**拆分为可见事实、馆方事实、解释、不确定项；禁止把情绪、身份和象征猜测混入事实。

**3 Reconstruct 重构：**输出15秒定位、45秒观察、90秒理解、焦点坐标、短替代文本和朗读稿。

**4 Audit 核验：**检查来源完整、馆藏号一致、推测词、字段长度、事实与解释标记。

**5 Review 人审：**学生或馆方审核者逐条对照原图和来源，记录删改理由并签署状态。

**6 Publish 发布：**只有`approved=true`且所有门禁通过的版本，才能进入网站、音频与NFC讲解。

强制规则：AI原始稿不得直接朗读。来源缺失、馆藏号不一致、出现未标记推测或未完成人审时，系统必须拒绝发布。

## 可迁移输出

同一份结构化结果可以驱动网页、博物馆App、支付宝小程序、语音终端、二维码页面和馆方内容管理系统，不把研究成果锁死在一个界面里。

# 从“会编故事”到“可放心朗读”

## 案例：《神奈川冲浪里》

AI原始稿：“海啸即将吞没渔民，画面表达了绝望。”

### 问题标注

| 原词   | 风险                     | 修订              |
|------|------------------------|-----------------|
| 海啸   | 作品呈现巨浪，但不能仅由图像确认其科学类型。 | 改为“巨大的蓝色浪头”。    |
| 即将吞没 | 对下一秒情节和人物结局作出推测。       | 改为“巨浪卷向三艘狭长小船”。 |
| 渔民   | 人物身份需要来源支持。            | 改为“船上的人”。       |
| 绝望   | 把情绪解释写成事实。             | 删除；只描述俯低身体的姿态。  |

### 发布稿的三层结构

**15秒定位：**巨浪占据左侧和上方，三艘小船位于浪谷，远处中央有很小的富士山。

**45秒观察：**从浪峰、泡沫、船只到富士山，按位置与尺度描述，不预言人物结局。

**90秒理解：**补充《富岳三十六景》系列和颜料信息；“尺度反差”明确标记为解释。

**发布证据：**馆藏ID 45434；The Met馆藏页；修订理由4条；人工审核通过；音频由发布稿生成。

# 麦田与柏树



文森特·梵高 · 1889 · 布面油画 · Public Domain

**15秒定位：**横向风景画。金黄色麦田位于下部，两株深绿色柏树在中央偏右，远处是蓝色山丘，上方布满卷曲的云。

**观察焦点：**柏树的垂直形态、麦田的短曲线笔触、天空的旋涡状云层。

**解释边界：**可以讨论垂直与横向层次、重复曲线带来的流动感；不把“宁静”等情绪当作画面事实。

# 神奈川冲浪里



葛饰北斋 · 约1830—1832 · 多色木版画 · Public Domain

**15秒定位：**巨浪占据左侧和上方，三艘小船位于浪谷，远处中央有一座很小的富士山。

**观察焦点：**向右弯曲的浪峰、爪尖状泡沫、俯低身体的人、被巨浪空隙框住的富士山。

**解释边界：**可讨论巨浪与富士山的尺度反差；不能把“海啸、吞没、绝望”当作确定事实。

# 持水罐的年轻女子



约翰内斯·维米尔 · 约1662 · 布面油画 · Public Domain

**15秒定位：**女子站在窗边，左手触碰窗框，右手握住桌上银色水罐。

**观察焦点：**左侧进入的光、白色头巾、浅黄色上衣、红色织物、水罐与墙上地图。

**解释边界：**可以描述冷光与暖红色的对比；不把人物确定为“女仆”，不猜测她正在准备早餐或心情平静。

# 舞蹈课



埃德加·德加 · 1874 · 布面油画 · Public Domain

**15秒定位：**多名芭蕾舞者分散在排练室，右侧站着一位手持长杖的教师，左前方舞者背对观众。

**观察焦点：**木地板斜线、前景纱裙、中景练习者、右侧教师和被画框截断的人物。

**解释边界：**可讨论偏离中心的观察视角；不能把排练说成整齐表演，也不能把教师的位置改成舞台中央。

# 苏格拉底之死



雅克-路易·大卫 · 1787 · 布面油画 · Public Domain

**15秒定位：**白衣老人坐在床中央，一手举起食指，另一手伸向身旁人递来的杯子，周围人物低头或掩面。

**观察焦点：**挺直明亮的身体、红衣递杯者、向上的手势、周围人物下垂和蜷缩的姿态。

**解释边界：**可讨论姿态对照；不能说“愤怒地拒绝毒药”，因为可见动作恰好是伸向杯子。

# 金匠铺中的金匠



佩特鲁斯·克里斯蒂 · 1449 · 橡木板油画 · Public Domain

**15秒定位：**金匠坐在柜台后使用小天平，右侧站着一对男女，柜台和墙架上陈列珠宝与器物。

**观察焦点：**小天平、钱币与戒指、圆形凸面镜、镜中街道和人影。

**解释边界：**可讨论交易、价值和精确称量；人物身份的历史解释存在变化，不直接认定金匠是圣人。

# A/B交叉实验：把“更好”变成可测量的问题

## 实验结构

同一参与者体验6件作品：3件只显示普通展签，3件显示分层描述；作品与条件交叉分配并随机排序。每件作品回答一道理解题，同时记录信心、认知负担和用时。

| 变量    | 测量方式             | 解释           |
|-------|------------------|--------------|
| 理解正确率 | 作品关键位置、动作或尺度选择题。 | 主要结果变量。      |
| 判断信心  | 0—100分滑块。        | 观察答案与信心是否匹配。 |
| 认知负担  | 1—5分，1轻松、5费力。    | 描述是否更易跟随。    |
| 完成时间  | 进入问题至提交的毫秒数。     | 辅助判断理解效率。    |
| 文本质量  | 错误、无依据解释、遗漏数量。   | 评价AI重构门禁。    |

## 样本与分析

- 先邀请10名同学进行认知试测，修正题目歧义和页面操作。
- 正式阶段建议40—80名高中生；平衡作品顺序，报告平均值、中位数和个体差异。
- 演示看板数据必须标注“演示”，正式采集前不得写成研究结论。
- 若邀请真实视障参与者，应另行知情同意并单独报告共同设计意见。



以上数字仅用于演示界面和分析方法，不是实际调查结果。

# 可访问网页、离线PWA与现场标签



## 已实现功能

- 六件作品数字展厅
- 三层视觉描述
- 本地女声与设备朗读回退
- AI修订记录
- 匿名A/B实验与结果导出
- 安卓Web NFC渐进增强
- NFC深链接和模拟触碰
- 标准、大字、特大字号
- 高对比与减少动态
- 首次联网后离线缓存

**隐私设计：**不需要账号，不采集姓名、学号或联系方式；实验数据只保存在当前设备，可由用户导出或删除。

# 八周实施日程与人员分工

|     |                               |
|-----|-------------------------------|
| 第1周 | 确定研究问题；学习无障碍指南与WCAG；选取开放版权作品。 |
| 第2周 | 整理馆藏元数据和普通展签；制定事实、观察、解释的标注规则。 |
| 第3周 | 使用AI生成原稿；逐条修订并保留差异；完成三层描述。    |
| 第4周 | 制作网页展厅、NFC深链接、PWA缓存和本地语音。     |
| 第5周 | 编写理解题、知情说明和匿名问卷；完成10人认知试测。    |
| 第6周 | 修正界面和题目；开展40—80人正式实验。         |
| 第7周 | 清洗匿名数据；比较正确率、信心、负担和用时。        |
| 第8周 | 撰写报告、制作答辩、整理来源与局限；发布可访问成果。    |

## 最小现场部署清单

| 物料       | 数量     | 说明                     |
|----------|--------|------------------------|
| NFC标签    | 每件作品1枚 | 写入HTTPS深链接和馆藏号，可另贴二维码。 |
| 安卓测试手机   | 1—2部   | 验证Web NFC授权和连续读取。      |
| 耳机或骨传导设备 | 按需要    | 避免公共空间音频干扰。            |
| 内容审核表    | 每件作品1份 | 记录来源、问题、修订和审核人。        |

# 对“AI向善”最重要的约束

## 共同设计

无障碍产品应尽早邀请真实盲人、低视力用户或专业机构参与；不能用蒙眼体验替代他们的经验。

## 事实透明

视觉观察、馆方事实、艺术解释和不确定项必须分开，用户能看到来源和修订理由。

## 不拟人化权威

语音清晰温和即可，不模仿真实主持人、演员或馆员，不借熟悉声音制造虚假权威感。

## 版权合规

只使用明确开放的藏品图像；馆方内部讲解需取得接口权限和内容使用授权。

## 最少数据

实验不收集身份信息；正式研究仍需知情同意、退出机制和数据保留期限。

## 技术诚实

不支持Web NFC时明确说明；扫码、手动选择与模拟触碰是回退或演示，不冒充后台自动感知。

## 当前局限

- 六件作品来自同一开放馆藏平台，尚不能代表不同类型博物馆。
- 讲解由学生审核，不等于博物馆策展人或无障碍专家审核。
- 当前产品尚未接入真实馆方内部导览接口。
- 演示数据不支持统计推断；正式结论需真实样本。

**下一步优先级：**先与本地博物馆或无障碍机构共同评审一件作品，再扩大作品数量，而不是先追求更多AI功能。

# 三分钟答辩稿

大家好，我的研究项目叫“触见名画”。我关注的不是AI能不能替代讲解员，而是一个更具体的问题：如果一个人无法直接看清画面，普通展签只告诉他作者和年代，是否足够？

我选取了大都会艺术博物馆六件公共领域名画，把讲解分成三层。第一层用15秒说明画面类型和主要对象的位置；第二层用45秒按顺序描述动作、色彩和细节；第三层用90秒补充经过来源核验的背景与解释。

项目最核心的不是朗读，而是一套AI讲解重构Skill。AI原稿必须经过六步：获取馆方资料、分离事实与解释、分层重构、自动核验、人工审核、批准发布。比如AI把《神奈川冲浪里》写成“海啸将吞没绝望的渔民”，我删除了“海啸、吞没、渔民、绝望”等无法由画面或来源确认的词，改成可以核验的位置和动作描述。

在使用方式上，我设计了NFC触碰导览。安卓手机在用户授权并保持网页开启后，可以靠近标签自动识别馆藏号并播放；其他手机通过NFC深链接或二维码进入同一作品。系统不会假装所有浏览器都支持后台自动识别。

研究部分采用被试内A/B实验，让同一位参与者分别体验普通展签和分层描述，比较理解正确率、信心、认知负担与用时。正式实验前，网页里的数据只标记为演示。

这个项目的局限也很明确：在真实视障用户参与前，我不能声称产品已经验证无障碍效果。因此下一步不是增加更多炫目的功能，而是邀请视障用户、无障碍机构和博物馆共同评审一件作品，让内容真正有用、可信并且尊重使用者。

一句话总结：触见名画不是让AI替人“看”，而是把名画讲解变成有来源、有层次、可核验、能被听见的信息。

# 匿名问卷与知情说明

**知情说明：**本实验用于研究不同作品说明方式对理解的影响。参加完全自愿，可随时退出；不收集姓名、学号、手机号或联系方式。数据仅用于学生研究性学习，导出前将再次检查是否含身份信息。

## 体验前

1. 你参观美术馆或艺术展览的频率：从不 / 每年1次 / 每年2—4次 / 每年5次以上。
2. 你平时是否使用语音朗读、屏幕阅读器或大字模式：从不 / 偶尔 / 经常。
3. 你对西方绘画或日本版画的熟悉程度：1完全不了解—5非常熟悉。

## 每件作品后

4. 根据当前说明回答一道空间、动作或细节理解题。
5. 你对答案的信心：0—100分。
6. 理解当前说明的负担：1轻松—5费力。
7. 当前说明中是否有难懂、含糊或像猜测的句子？请指出。

## 体验后

8. 哪种说明更帮助你形成画面结构：普通展签 / 分层描述 / 无明显差异。
9. 三层说明中最有用的是：15秒定位 / 45秒观察 / 90秒理解。
10. 你是否愿意在博物馆使用触碰自动讲解：愿意 / 视情况 / 不愿意。
11. 你最担心的问题：事实错误 / 隐私 / 操作复杂 / 语音不自然 / 其他。
12. 请写下一条最重要的改进建议。

# 内容审核表与数据字典

## 单件作品发布审核表

| 检查项   | 通过标准                    | 记录      |
|-------|-------------------------|---------|
| 身份字段  | 馆藏号、标题、作者、年代、媒介与馆方页面一致。 | 通过 / 退回 |
| 图像来源  | 公共领域或已取得授权，保留原始链接。      | 通过 / 退回 |
| 15秒定位 | 包含类型、主要对象和位置，不含心理猜测。    | 通过 / 退回 |
| 45秒观察 | 顺序清楚，动作、色彩、光线可由图像核对。    | 通过 / 退回 |
| 90秒理解 | 馆方事实有来源，形式分析标记为解释。      | 通过 / 退回 |
| 不确定项  | 争议身份和未知内容明确说明，不强行确定。    | 通过 / 退回 |
| 人工审核  | 审核人、日期、版本和修订理由完整。       | 通过 / 退回 |

## 匿名数据字典

| 字段               | 类型    | 含义                          |
|------------------|-------|-----------------------------|
| artworkId        | 文本    | 项目内部作品标识。                   |
| condition        | 枚举    | standard普通展签 / layered分层描述。 |
| choice / correct | 整数    | 参与者选择与正确选项索引。               |
| confidence       | 0—100 | 判断信心。                       |
| load             | 1—5   | 主观认知负担。                     |
| elapsedMs        | 整数    | 单题完成时间，毫秒。                  |

明确排除：姓名、学号、手机号、社交账号、精确位置和设备广告标识。

# 权威来源与引用说明

1. UNESCO. “Guidelines on Museum Accessibility in China.” 2026-01-26.  
<https://www.unesco.org/en/articles/guidelines-museum-accessibility-china>
2. W3C. “Web Content Accessibility Guidelines (WCAG) 2.2.” <https://www.w3.org/TR/WCAG22/>
3. Smithsonian Institution. “Accessibility.” <https://www.si.edu/visit/accessibility>
4. The Metropolitan Museum of Art. Open Access API. <https://metmuseum.github.io/>
5. The Met Collection, Object 436535. “Wheat Field with Cypressess.”  
<https://www.metmuseum.org/art/collection/search/436535>
6. The Met Collection, Object 45434. “Under the Wave off Kanagawa.”  
<https://www.metmuseum.org/art/collection/search/45434>
7. The Met Collection, Object 437881. “Young Woman with a Water Pitcher.”  
<https://www.metmuseum.org/art/collection/search/437881>
8. The Met Collection, Object 438817. “The Dance Class.” <https://www.metmuseum.org/art/collection/search/438817>
9. The Met Collection, Object 436105. “The Death of Socrates.”  
<https://www.metmuseum.org/art/collection/search/436105>
10. The Met Collection, Object 459052. “A Goldsmith in his Shop.”  
<https://www.metmuseum.org/art/collection/search/459052>

## 图片与内容说明

本报告中的作品图像来自The Met开放获取馆藏，并依据馆藏页的Public Domain状态用于学生研究与非商业演示。馆藏元数据以原始页面为准。项目由天问研究所出品，使用已锁定的天字品牌标记，作者为Nankong；不代表任何博物馆官方发布或背书。

## 版本与交付

**网页版本：**V1.0.1

**内容版本：**六件作品审核稿

**技术版本：**静态PWA、Web NFC渐进增强、本地语音

**研究状态：**方法与演示系统完成，正式样本尚未采集

**版权：**Copyright © 2026 Nankong. 保留所有权利

真正的无障碍不是替别人想象需求，而是把技术做成可以被检验、被质疑、被共同修改的工具。