

基于生成式 AI 的文创产品设计

--软件项目计划书

参与开发者：郑仕博 陈奕嘉 苏泳豪

Contents

- 1 引言 2
 - 1.1 编写目的 2
 - 1.2 背景 2
 - 1.3 定义 2
 - 1.4 项目内容 3
- 2 项目概述 3
 - 2.1 项目目标 3
 - 2.2 产品目标与范围 3
 - 2.3 应交付成果 3
 - 2.3.1 需完成的软件 3
 - 2.3.2 需提交用户的文档 3
 - 2.3.3 须提交内部的文档 4
 - 2.3.4 应当提供的服务 4
 - 2.4 项目开发环境 4
- 3 项目团队组织 4
 - 3.1 人员分工 4
- 4 实施计划 4
 - 4.1 风险评估及对策 4
 - 4.2 总体进度计划 4
 - 4.3 项目里程碑计划 5
- 5 预算 5
 - 5.1 人员成本 5
 - 5.2 设备成本 5
 - 5.3 其它经费预算 5

1 引言

1.1 编写目的

文创，顾名思义，文化创意，以文化为元素、融合多元文化、整理相关学科、利用不同载体二构建的再造与创新的文化现象，也是旅游和文化结合的产物。《如果国宝会说话》纪录片于 2017 年 12 月 25 日在国家博物馆首映，不仅没有让纪录片变得高冷，反而以最现代、最流行、最鲜活灵动的形式，自然而然地流淌进观众的心中。本项目的核心目标，是赋能每一位个体，凭借其丰富的想象力，轻松创造出独一无二的文创作品，让静默的文物在创意的驱动下“活”起来，讲述它们自己的故事。图 1.1 展示了新华网所创作的作品中，青花瓷所传达的话语。



图 1.1

1.2 背景

习近平总书记强调，要推进文化和旅游融合发展，把文化旅游业培育成为支柱产业。文化是旅游的灵魂，旅游是文化的载体，二者有着天然的联系。文化元素丰富旅游产品供给，旅游元素释放文化产业经济价值。推动文化和旅游融合发展，既有助于推动文化繁荣发展，也有利于推进文化旅游业转型升级，更好满足人民群众对美好生活的新期待。^[1]

1.3 定义

本项目构建的实际上是基于 OCR 技术与扩散模型的个性化文创产品生成系统，其中：

OCR 技术（Optical Character Recognition）是指对包含文本内容的图像或视频进行处理和识别，并提取其中所包含的文字及排版信息的过程。

扩散模型是一种生成模型，主要用于图像生成和其他计算机视觉任务。

利用这两种技术，我们可以满足对个性化文创产品生成的需求。

1.4 项目内容

本项目将致力于改进现有的 AnyText^[2] 模型，旨在提升其汉字渲染效果。在此基础上，本项目期望通过与其他模型的融合，实现多样化的功能，包括汉字渲染、文字删除、图片修改以及文字到图片的生成。鉴于中文数据集的稀缺性，本项目计划充相关的中文训练数据。在文字渲染模型中，本项目将运用 OCR（光学字符识别）技术，并考虑采用更先进的 OCR 技术来进一步提升性能。同时，我们将利用现有的预训练扩散模型，并结合收集的图片数据进行进一步训练，旨在赋予其文化创意生成的能力。

2 项目概述

2.1 项目目标

本项目将探索将模型应用于文创生成领域的可能性，以满足广大用户的需求。

2.2 产品目标与范围

预期部署一个网站，主要功能为通过用户文字输入来生成图片，若用户输入的文字与文创产品相关，则可以以较好的效果生成相关图片，满足用户个性化生成文创产品的需求。产品预期面向人群为文创创作者以及文创产品爱好者。

2.3 应交付成果

2.3.1 需完成的软件

模型微调的代码实现和权重文件
前后端相关代码的实现

2.3.2 需提交用户的文档

《免责声明》
软件著作权说明书

2.3.3 须提交内部的文档

开发周报

《需求分析》

《系统设计概述》

《报告 PPT》

2.3.4 应当提供的服务

对模型的漏洞进行实时调整。

对用户的反馈进行调查并完善用户体验。

2.4 项目开发环境

Windows & linux

3 项目团队组织

3.1 人员分工

郑仕博：统筹资源，决定方向，代码撰写与调整，训练模型

陈奕嘉：后端开发，文书撰写，为代码调整方向提供可行性思路

苏泳豪：前端开发，超参数调整，为代码调整方向提供可行性思路

4 实施计划

4.1 风险评估及对策

用户使用不当风险及对策：若用户输入模型未训练的文字，模型不能保证生成的准确性。

若用户输入违法信息，如期望生成色情、犯罪、以及反党反社会图片时，模型应及时辨别并立即终止生成并对用户予以警告，若同一用户不断进行此类行为，保留对该用户追责的权利。

在扩散模型的训练过程中，有可能会遭遇过拟合的问题，这可能导致模型泛化能力不足。此外，经过训练的模型可能在生成文物种类方面存在局限性，所能创造的文物样式相对有限。

4.2 总体进度计划

1-3 周 项目计划书完成，模型初步调试

4-7 周 模型深度调整，前后端开发，相关文档撰写
8-9 周 模型部署以及运营测试
10-12 周 扩充数据集，对模型进行加强训练
13-16 周 修复可能漏洞，完成交付准备

4.3 项目里程碑计划

第 4 周 完成全部文档的撰写
第 8 周 完成前后端的开发
第 12 周 确认最终模型
第 14 周 正式开放项目供用户使用

5 预算

5.1 人员成本

除组内三人外无其他人员成本。
如网站运营扩大，将会增加运维人员。

5.2 设备成本

租卡费用（主要是 4090，可能会用虚拟显卡）：2000 元

5.3 其它经费预算

软件著作权：500 元

参考文献

- [1](2024-11-04 16:25)习近平总书记强调，把文化旅游业培育成为支柱产业. https://www.sohu.com/a/823541100_234564
- [2] Junnan Li, Dongxu Li, Silvio Savarese, and Steven C. H. Hoi. BLIP-2: bootstrapping language-image pre-training with frozen image encoders and large language models. arXiv preprint,abs/2301.12597, 2023.