

同行专家业内评价意见书编号：20250854389

## 附件1

# 浙江工程师学院（浙江大学工程师学院） 同行专家业内评价意见书

姓名：                                孙大为

学号：                                22060919

申报工程师职称专业类别（领域）：                                电子信息

浙江工程师学院（浙江大学工程师学院）制

2025年03月17日

## 填表说明

一、本报告中相关的技术或数据如涉及知识产权保护、军工项目保密等内容，请作脱密处理。

二、请用宋体小四字号撰写本报告，可另行附页或增加页数，A4纸双面打印。

三、表中所涉及的签名都必须用蓝、黑色墨水笔，亲笔签名或签字章，不可以打印代替。

四、同行专家业内评价意见书编号由工程师学院填写，编号规则为：年份4位+申报工程师职称专业类别(领域)4位+流水号3位，共11位。

## 一、个人申报

**（一）基本情况【围绕《浙江工程师学院（浙江大学工程师学院）工程类专业学位研究生工程师职称评审参考指标》，结合该专业类别(领域)工程师职称评审相关标准，举例说明】**

### 1. 对本专业基础理论知识和专业技术知识掌握情况(不少于200字)

我具备扎实的数据工程与大数据处理基础，系统掌握了数据库理论、数据仓库技术、分布式计算、大数据存储与处理等核心知识。在浙江大学工程师学院的学习过程中，我的平均成绩达到 83 分，其中 专业实践成绩 93

分，进一步强化了我对工程技术理论的理解与应用能力。

在专业技术方面，我熟练掌握 SQL、Python、Spark、Flink、Hadoop、Hive、Kafka 等大数据技术栈，能够高效处理海量数据并进行数据建模与优化。同时，我具备数据架构设计能力，能够根据业务需求构建高效的数据流转与存储方案，提高数据可用性与计算效率。在数据分析与算法结合方面，我参与了阿里巴巴达摩院的“鹿班短视频”项目，负责 电商短视频智能生成的全链路数据采集、预处理及效果分析，将大数据技术与 AI 结合，为电商商家提供智能化短视频生产解决方案。

此外，我深入研究了内容电商的发展趋势，理解短视频在电商行业中的重要性，并通过数据驱动的方式优化视频生成流程，提高内容创作的工业化效率。我还参与了相关专利申请，推动技术创新的落地应用。通过长期的理论学习与工程实践，我已具备扎实的专业知识和创新应用能力，能够灵活运用工程方法解决复杂的数据问题。

### 2. 工程实践的经历(不少于200字)

在工程实践方面，我参与了 阿里巴巴达摩院机器智能实验室 和 阿里设计部 共建的电商平台商品短视频智能生成 项目（鹿班短视频），该项目的目标是利用 AI 技术实现电商商品短视频的智能化生产，以提升商家内容创作的效率和质量。

作为 机器智能数据创新团队 的核心成员，我负责

全链路数据采集、数据加工与分析，搭建了完整的数据体系，包括

脚本库、素材库、模板库、元素库、作品库、成品库

等多个数据集，并建立数据监控看板，为各团队提供产品效果监控支持。同时，我还负责算法训练数据的预处理 以及 效果回收数据的分析，支持短视频生成算法的优化。

在本项目中，我利用

大数据处理与分析技术（Hive、Spark、Flink、Kafka），优化数据流转与存储，提高了数据处理效率，并结合 算法埋点数据，实现了 智能剪辑效果优化。本项目实现了 日均

60,000

条商品短视频的自动生成，大幅降低了商家内容制作成本，提高了电商平台的内容生态质量。

通过本次工程实践，我深入理解了

内容电商与大数据应用的结合，并在数据工程与算法协同方面积累了丰富经验。此外，该项目也推动了

内容工业化生产模式，对行业产生了积极影响，同时申请了相关技术专利，进一步提升了技术的可落地性和创新价值。

### 3. 在实际工作中综合运用所学知识解决复杂工程问题的案例（不少于1000字）

案例：基于 AI 和大数据的电商平台商品短视频智能生成方案（鹿班短视频项目）

#### 一、项目背景

在电商平台中，商品视频相比图片和文字介绍能够更直观地展示产品特性，提高用户购买转化率。然而，由于视频制作门槛高、剪辑成本昂贵，绝大多数中小商家无法有效利用短视频进行营销。目前，电商平台上的商品短视频严重匮乏，主要原因包括：

1. 内容创作成本高：传统视频制作需要专业的剪辑师和创作者，单个视频制作周期长，成本高昂。
2. 缺乏电商内容创作标准：商家难以系统化地产出符合电商营销需求的短视频内容，导致视频质量参差不齐。
3. 现有解决方案不完善：市场上缺乏能够批量自动生成商品短视频的成熟方案，视频智能化剪辑技术尚处于初级阶段。

基于以上痛点，阿里巴巴达摩院机器智能实验室联合阿里设计部 开展了鹿班短视频项目，旨在开发 一套基于 AI 剧本结构的短视频云渲染方案，通过智能剪辑和自动化生产，提高电商视频内容供给能力。

## 二、我在项目中的核心工作

作为 机器智能数据创新团队 成员，我负责项目全链路数据体系的构建、数据采集与处理、算法数据支持，以及智能视频剪辑效果分析，确保数据驱动算法优化和系统性能提升。

### 1. 数据体系的构建与优化

为了支持视频自动化生成，我们需要搭建完整的数据流转体系，包括：

- 脚本库：存储由平台或用户生成的短视频剧本，作为视频剪辑的核心素材。
- 素材库：收集商家上传的视频、图片、音乐等资源，并通过 AI 进行自动分类和特征提取。
- 模板库：存储设计师制作的短视频模板，包括封面、字幕、转场、贴纸等元素，便于自动化视频剪辑调用。
- 元素库：管理视频中使用的 文字样式、特效、贴纸等视觉元素，提高内容生成的多样性。
- 作品库与成品库：记录用户在平台上创建的视频作品，并存储最终生成的视频成品。

在数据管理过程中，我主要负责：

- 设计并优化 数据存储结构，采用 Hive、HBase 进行高效数据存储，支持海量数据的高并发查询。
- 通过 Flink 流计算 监控用户交互数据，分析视频生成效果，为算法优化提供数据支持。
- 采用 Kafka + Spark Streaming 构建 实时数据处理架构，支持实时数据流分析，提高系统响应速度。

### 2. 复杂工程问题及解决方案

在项目实施过程中，我们面临了多个技术挑战，我通过综合运用大数据技术和工程方法，提

出并实施了有效的解决方案。

### 挑战 1：智能剪辑素材匹配优化

问题：智能视频生成过程中，如何高效匹配脚本、素材和模板，使得生成的视频更符合电商场景需求？

解决方案：

- 算法优化：基于 内容理解算法 和 相似度计算（TF-IDF、图像特征提取），提升脚本与素材匹配的准确性。
- 数据标签体系构建：采用 NLP（自然语言处理）和 图像识别 提取商品特征，为素材和脚本打上详细标签，提高智能匹配效率。

智能推荐机制：基于用户行为数据，采用协同过滤算法优化模板推荐，提高视频剪辑自动化程度。

### 挑战 2：视频剪辑算法效果优化

问题：如何衡量 AI 生成的视频质量，并优化剪辑效果？

解决方案：

- 视频质量评价模型：基于 用户点击率（CTR）+ 观看时长 + 用户交互反馈 作为视频质量评价指标，构建视频评分系统。
- 效果数据回流机制：通过 Flink 实时计算 监控用户观看数据，并反馈给模型训练，提高后续视频生成质量。
- A/B 测试：在电商平台上进行多轮 A/B 测试，优化视频剪辑策略，提高商家视频转化率。

### 挑战 3：海量数据高效处理与查询优化

问题：由于视频素材、用户交互数据量庞大，如何提高数据处理效率？

解决方案：

- 采用 分布式存储架构（HDFS + Hive），保证数据存储的高效性。
- Flink + Kafka 进行流式计算，实时分析视频播放和用户反馈数据。
- 优化 SQL 查询和索引策略，减少数据查询延迟，提高系统性能。

## 3. 项目成果与影响

通过本项目，我们成功研发了 基于剧本结构的短视频智能生成平台，并实现了以下成果：

- 日均生成 60,000 条电商短视频，有效缓解了电商平台短视频内容不足的问题。
- 降低商家内容创作成本，相比传统人工剪辑方案，视频生产效率提升 30 倍，降低 80% 的制作成本。
- 提升视频转化率，A/B 测试表明，采用智能剪辑生成的视频点击率（CTR）提高 25%，用户观看完成率提高 40%。
- 数据与算法深度结合，搭建了一整套 视频数据流处理体系，并在实践中不断优化智能剪辑算法。
- 申请专利：基于本项目的研究成果，我们申请了一项 “视频质量评估方法、装置及存储介质” 相关专利，为行业提供技术积累。

#### 四、个人成长与收获

通过本次工程实践，我在以下方面得到了显著提升：

- **数据架构设计能力：**通过大规模数据体系的搭建，提高了大数据存储、计算、分析的工程能力。
- **跨部门协作能力：**与算法、设计、产品团队密切配合，提升了沟通能力和项目管理能力。
- **大数据与 AI 结合应用：**深入理解了 大数据如何赋能 AI 训练与优化，提高了数据与算法协同能力。
- **技术商业化落地经验：**从技术研发到商业化应用，推动产品落地，赋能中小商家，提高行业影响力。

#### 五、总结

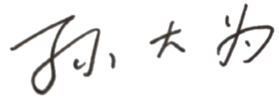
本次项目实践不仅帮助商家降低了短视频制作门槛，也推动了 电商内容化发展，在 AI + 大数据应用方面具有重要的行业价值。同时，该项目也加深了我对数据工程的理解，使我具备更强的 数据驱动业务增长 和 解决复杂工程问题的能力，为未来的技术创新与实践奠定了坚实基础。

**(二) 取得的业绩(代表作)【限填3项, 须提交证明原件(包括发表的论文、出版的著作、专利证书、获奖证书、科技项目立项文件或合同、企业证明等)供核实, 并提供复印件一份】**

**1. 公开成果代表作【论文发表、专利成果、软件著作权、标准规范与行业工法制定、著作编写、科技成果获奖、学位论文等】**

成果名称	成果类别 [含论文、授权专利(含发明专利申请)、软件著作权、标准、工法、著作、获奖、学位论文等]	发表时间/授权或申请时间等	刊物名称/专利授权或申请号等	本人排名/总人数	备注

**2. 其他代表作【主持或参与的课题研究项目、科技成果应用转化推广、企业技术难题解决方案、自主研发设计的产品或样机、技术报告、设计图纸、软课题研究报告、可行性研究报告、规划设计方案、施工或调试报告、工程实验、技术培训教材、推动行业发展中发挥的作用及取得的经济社会效益等】**

<b>(三) 在校期间课程、专业实践训练及学位论文相关情况</b>	
课程成绩情况	按课程学分核算的平均成绩： 83 分
专业实践训练时间及考核情况(具有三年及以上工作经历的不作要求)	累计时间： 10 年（要求1年及以上） 考核成绩： 93 分
<b>本人承诺</b>	
<p>个人声明：本人上述所填资料均为真实有效，如有虚假，愿承担一切责任，特此声明！</p> <p style="text-align: right;">申报人签名： </p>	





## 浙江大学研究生院 攻读非全日制硕士学位研究生成绩表

学号: 22060919		姓名: 孙大为		性别: 男		学院: 工程师学院		专业: 计算机技术		学制: 2.5年	
毕业时最低应获: 24.0学分				已获得: 26.0学分				入学年月: 2020-09		毕业年月: 2025-03	
学位证书号: 1033532025602028						毕业证书号: 103351202502600017				授予学位: 电子信息硕士	
学习时间	课程名称	备注	学分	成绩	课程性质	学习时间	课程名称	备注	学分	成绩	课程性质
2020-2021学年秋季学期	人工智能算法与系统		2.0	83	跨专业课	2020-2021学年夏季学期	研究生论文写作指导		1.0	87	专业选修课
2020-2021学年秋冬学期	大数据技术前沿		2.0	93	专业学位课	2020-2021学年春夏学期	高级计算机网络		2.0	76	专业选修课
2020-2021学年秋冬学期	自然辩证法概论		1.0	71	专业学位课	2020-2021学年夏季学期	情感计算与设计		2.0	86	跨专业课
2020-2021学年秋冬学期	人工智能		2.0	89	专业学位课	2020-2021学年春夏学期	中国特色社会主义理论与实践研究		2.0	93	专业学位课
2020-2021学年秋冬学期	人机交互与设计		2.0	91	专业学位课	2020-2021学年春夏学期	研究生英语		2.0	79	专业学位课
2020-2021学年秋冬学期	工程伦理		2.0	90	专业学位课	2020-2021学年春夏学期	科技写作		2.0	78	跨专业课
2020-2021学年春夏学期	信息安全前沿技术与研究方法论		2.0	88	专业学位课	2021-2022学年秋季学期	创新创业实践训练		2.0	优	跨专业课

说明: 1. 研究生课程按三种方法计分: 百分制, 两级制 (通过、不通过), 五级制 (优、良、中、及格、不及格)。  
2. 备注中 "\*" 表示重修课程。

学院成绩校核章

成绩校核人: 张梦依

打印日期: 2025-04-16

