

VLM-Based Context Extraction and Fine-Tuning of Korean LLMs for Automatic Travel Blog Content Generation

Donghun Lim*, Seungsoo Han**, Euichan Eun***, Dongyeong Kim****, Janghoon Choi*****

*M.S. Student, Graduate School of Data Science, Kyungpook National University, Daegu, Korea

**Researcher, Funzin Co., Seoul, Korea

***M.S. Student, Dept. of Artificial Intelligence Systems, Hanyang University, Seoul, Korea

****B.S. Student, Dept. of Industrial and Management Engineering, Ulsan University, Ulsan, Korea

*****Assistant Professor, Graduate School of Data Science, Kyungpook National University, Daegu, Korea

[Abstract]

In this paper, we propose an automated system for generating Korean travel blog content by integrating Vision-Language Model (VLM)-based context extraction, fine-tuned Korean language model, and a text-to-image (T2I) generation model via prompt engineering. Using the Qwen2-VL model, we analyze travel photos to extract visual context and produce emotionally nuanced captions. We then leverage a large-scale crawled corpus of real-world travel blogs to fine-tune HyperCLOVA X, enabling it to create natural, storytelling-oriented blog text. In addition, we employ a travel-specific prompt engineering approach with DALL-E to generate custom postcards and stamp images, providing users with intuitive and creative value for their travel records. A robust system prompt minimizes hallucinations while preserving expressive writing. User surveys indicate that the fine-tuned model is, on average, 60.9% more specialized in travel-related content. These findings demonstrate that the proposed approach can significantly reduce manual effort while producing high-quality travel blog content, suggesting new possibilities for AI-based content creation in the tourism domain.

▶ **Key words:** Vision-Language Model, Fine-Tuning, Prompt Engineering, T2I Generation, Travel Blog

[요약]

본 논문에서는 시각-언어 모델(VLM) 기반 맥락 추출과 파인 튜닝된 한국어 언어모델, 그리고 프롬프트 엔지니어링을 통한 텍스트-이미지(T2I) 생성 모델을 결합하여 한국어 여행 블로그 콘텐츠 자동 생성 시스템을 제안한다. Qwen2-VL 모델을 통해 여행 사진의 시각 정보를 분석해 감성적인 캡션을 생성하고 크롤링을 통한 여행 블로그 데이터를 활용한 HyperCLOVA X 모델 파인 튜닝을 통해 자연스러운 스토리텔링의 블로그 글을 생성한다. 또한, 여행 특화 프롬프트 엔지니어링을 통한 DALL-E 모델을 통해 여행 엽서-스탬프 이미지를 생성하여 직관적이고 창의적인 여행 기록을 제공한다. 강력한 시스템 프롬프트를 통해 환각 현상을 최소화하고, 감정 표현이 풍부한 문체를 유지하도록 설계하였다. 설문조사 결과, 파인튜닝 모델의 생성 결과가 여행에 더욱 특화되어 있다는 응답이 평균 60.9%로 나타났다. 이는 본 논문의 블로그 콘텐츠 생성물이 수작업 부담을 크게 줄이면서도 고품질의 여행 콘텐츠를 생성할 수 있음을 보여주며, 관광 분야에서 AI 기반 콘텐츠 제작에 대한 새로운 가능성을 제시한다.

▶ **주제어:** 시각-언어 모델(VLM), 파인 튜닝, 프롬프트 엔지니어링, 텍스트-이미지(T2I) 생성, 여행 블로그

- First Author: Donghun Lim, Corresponding Author: Janghoon Choi
- *Donghun Lim (naring@knu.ac.kr), Graduate School of Data Science, Kyungpook National University
- **Seungsoo Han (ethan.han@funzin.co.kr), Funzin Co.
- ***Euichan Eun (dmsdmlcks@hanyang.ac.kr), Dept. of Artificial Intelligence Systems, Hanyang University
- ****Dongyeong Kim (kimdyoc13@gmail.com), Dept. of Industrial and Management Engineering, Ulsan University
- *****Janghoon Choi (jhchoi09@knu.ac.kr), Graduate School of Data Science, Kyungpook National University
- Received: 2025. 04. 16, Revised: 2025. 05. 16, Accepted: 2025. 06. 09.

I. Introduction

최근 초거대 언어모델(LLM)과 시각언어모델(VLM) 기술이 비약적으로 발전하면서, 텍스트와 이미지를 융합한 콘텐츠 생성 분야가 다양한 산업으로 확장되고 있다 [1][2]. 특히 생성형 AI가 개인화된 마케팅이나 콘텐츠 제작에서 높은 효율성을 보이는 사례가 늘어나면서 [3][4], 사용자 개인의 경험과 취향에 특화된 콘텐츠를 자동 생성해 주는 시스템에 대한 관심이 커지고 있다.

이러한 흐름에서 관광 분야는 AI 기술 적용에 있어 매력적인 영역으로 주목받고 있다. 여행 후기나 관광 블로그는 잠재적 소비자에게 신뢰감을 줄 수 있는 주요 마케팅 수단이며, 여행자의 자유로운 경험 공유 채널로 활용도가 매우 높다 [5][6]. 하지만 전통적인 여행 블로그 작성 방식은 많은 시간과 노력이 필요한 반면, 정작 블로그 작성에 대한 글쓰기 부담으로 인해 여행 후기를 제대로 정리하지 못하는 경우가 많다. 따라서 사용자가 빠르고 간편하게 동시에 개인화된 여행 기록을 자동으로 생성할 수 있는 새로운 접근 방식이 필요하다.

최근 연구들을 살펴보면, 대규모 한국어 언어모델을 활용해 협찬 블로그 텍스트 생성에 성공한 시도가 존재하고 [7], 요약문 기반 기억 시스템을 도입하여 장문 에세이에서도 보다 구조적인 내용을 생성하는 기법이 제안되기도 하였다 [8]. 그러나 이들 연구는 텍스트 생성에만 집중하거나, 협찬 광고에 특화된 경우가 많아서 이미지와 텍스트가 결합된 여행 블로그 문맥에서 종합적인 자동화를 보인 사례는 제한적이었다. 특히 (i) 시각적 맥락 분석(VLM 기반)과 (ii) 대규모 한국어 LLM 파인튜닝, (iii) 엽서·스탬프 등 시각 요소를 생성하는 텍스트-이미지(T2I) 모델이 유기적으로 결합된 연구는 상대적으로 미흡하다.

또한, 생성 AI가 관광산업 경쟁력을 높이고 혁신적인 서비스 개발을 촉진할 수 있다는 견해가 확산되고 있지만 [9], 대규모 크롤링 데이터셋을 활용해 얼마나 효과적으로 여행 블로그에 적용할 수 있는지는 추가 검증이 필요하다. 실제로 본 연구에서는 유명 블로그 1,600건을 선별·크롤링하여 한국어 대규모 언어모델을 파인튜닝하였고, 이를 통해 자유로운 스토리텔링과 감정 표현이 접목된 여행 블로그 초안을 자동 생성하는 방식을 고안하였다. 관광산업 맥락에서 생성형 AI를 고도화하기 위해서는 (a) 이미지 속 시각 정보를 정확히 추출하고, (b) 한국어 특유의 담화·어미·감정 표현을 반영하고, (c) AI Safety 측면(잘못된 정보나 환각(hallucination) 최소화)을 동시에 고려해야 한다.

이에 본 논문은 VLM 기반 맥락 추출, 한국어 언어모델

파인튜닝, 그리고 여행 특화 프롬프트 엔지니어링을 통한 T2I 생성을 결합한 자동 여행 블로그 콘텐츠 생성 시스템인 '트래블로그(Travelog)'를 제안한다. 구체적으로, Qwen2-VL 모델을 활용하여 사용자가 업로드한 여행 사진에서 시각적 맥락을 풍부하게 추출하고, 1,600건 규모의 여행 블로그 텍스트를 기반으로 HyperCLOVA X를 파인튜닝하여 감성적이면서도 사실적인 블로그 글을 자동으로 생성한다. 더불어 DALL-E API를 활용한 엽서·스탬프 이미지 생성을 병행하여, 사용자에게 독창적이고 직관적인 여행 기록 경험을 제공한다. 또한, 강력한 시스템 프롬프트 전략을 적용하여 사용자 제공 정보와 모델 출력 간 불일치나 환각을 줄이고, 감정 표현이 풍부한 문체를 유지하도록 설계하였다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 연구와 기술 요소(VLM, LLM, T2I, AI Safety 등)를 조사하고, 요약문 기반 장문 생성 [8], 한국어 협찬 블로그 모델링 [7], 관광산업 맥락에서의 생성형 AI 활용 [6]과의 비교를 통해 본 연구의 공백과 차별점을 명확히 한다. 3장에서는 제안하는 트래블로그의 전체 아키텍처, 데이터 구축, 그리고 모델링 방법을 설명한다. 4장에서는 시스템 구현 과정과 실제 활용 시나리오를 기술하고, 5장에서는 사용자 평가 및 성능 분석 결과를 제시한다. 마지막으로 6장에서는 결론 및 향후 과제를 논의한다. 본 연구는 관광 분야에서 AI 기반 자동화 콘텐츠 생성에 기여하여, 여행 산업에서 새로운 가치 창출의 가능성을 확장할 것으로 기대된다.

II. Related Work

인공지능(AI) 기술, 특히 대규모 언어모델(LLM), 시각-언어모델(VLM), 텍스트-이미지(T2I) 생성 모델의 급격한 발전은 다양한 분야에서 자동 콘텐츠 생성의 새로운 가능성을 열어주고 있다. 본 장에서는 여행 블로그 콘텐츠 생성과 관련된 기존 연구를 검토하고, (1) 관광 분야에서의 여행 블로그 활용, (2) 텍스트 생성을 위한 LLM과 VLM의 적용, (3) 시각 콘텐츠 생성을 위한 T2I 기술을 중심으로 논의한다. 이를 통해 제안된 "트래블로그" 시스템의 위치를 명확히 하고, 독창적인 기여를 부각시킨다.

1. Research on Travel Blogs and Tourism

여행 블로그는 개인 경험을 공유하고 잠재적 관광객에게 영향을 미치는 강력한 매체로 자리 잡았다. 안현주와 하영 [5]는 경험 유형과 블로그 텍스트 간의 관계를 분석

하고, 서사 중심의 블로그가 독자의 참여와 신뢰를 높인다고 밝혔다. 노영과 변지영 [7]은 관광 블로그의 수용과 사용자 만족도를 연구하면서, 블로그가 신뢰할 수 있는 마케팅 도구이자 경험 공유 채널로 기능한다고 강조했다. 이러한 연구들은 여행 블로그의 관광 산업 내 가치를 보여주지만, 전통적인 블로그 작성 방식이 많은 시간과 노력을 요구한다는 한계를 지적한다. 프로젝트 제안서에서 언급된 코로나 이후 여행 수요 증가와 같은 최근 트렌드는 이러한 한계를 해결할 효율적인 콘텐츠 생성 도구의 필요성을 더욱 부각시킨다.

이원희와 김영 [6]은 관광 산업에서 생성형 AI의 잠재력을 탐구하면서, AI 기반 솔루션이 콘텐츠 제작을 간소화하고 경쟁력을 높일 수 있다고 제안했다. 그러나 이 연구는 개념적 틀에 머물며, 텍스트와 이미지를 통합한 실질적인 여행 기록 구현 사례는 제시하지 못했다. 이에 비해 "트래블로그"는 텍스트와 시각적 콘텐츠를 자동으로 생성하여 수작업에 따른 부담을 경감시키는 동시에, 여행 블로그의 감성적이고 서사적인 특성을 유지한다.

2. Large Language Model (LLM)

최근 대규모 언어모델(Large Language Model, LLM)은 자연어 이해와 생성 분야에서 혁신적인 성과를 보이며 다양한 콘텐츠 생성 작업에 적극적으로 활용되고 있다 [1][2]. 특히, 사전학습(pre-training)을 통해 방대한 데이터에서 언어적 지식을 습득하고, 특정 작업이나 도메인에 맞춰 파인튜닝(fine-tuning)을 수행함으로써 성능을 현저히 개선할 수 있다.

2.1 Vision Language Model (VLM)

시각과 텍스트를 통합적으로 이해하는 시각-언어 모델(Vision-Language Model, VLM)은 다중 모달(multimodal) 콘텐츠 생성 작업에서 주목받고 있다. 최근 Peng 등[10]은 알리바바 Qwen 팀이 개발한 최신 VLM인 Qwen2-VL을 소개하며, 이 모델은 다양한 해상도의 이미지와 비디오를 효과적으로 처리하고 다국어 문맥을 지원한다고 밝혔다. Qwen2-VL은 GPT-4V, LLaVA-1.5, Flamingo 등의 기존 모델과 비교해 시각 질문 응답(Visual Question Answering, VQA)과 이미지 캡션 생성(image captioning)에서 우수한 성능을 보였다. 특히 여행 사진과 같은 현실적이고 복잡한 장면에서 풍부하고 정확한 설명을 생성하는 데 강점이 있다. 김성훈 등[1]도 생성형 AI 응용에서 VLM의 중요성을 강조했지만, 여행 도메인과 같은 특정 분야에 특화하여 VLM의 성능을 최적화

하는 연구는 아직 미흡한 실정이다.

본 연구에서는 여행 블로그 콘텐츠 생성이라는 특화된 작업을 위해 Qwen2-VL을 선정하였다. 이 모델은 LLaMA 3.2 Vision 등 다른 후보 모델과의 실험 비교를 통해 더 상세한 위치 정보와 다국어 캡션 생성 성능이 우수함을 입증했기 때문에, 여행 콘텐츠의 현실적 묘사와 사용자의 경험을 정확히 반영하는 데 가장 적합하였다.

2.2 Korean Large Language Model

한국어에 최적화된 대규모 언어모델인 네이버의 HyperCLOVA X는 한국어 이해와 생성 능력에서 뛰어난 성능을 보이며, 한국어 콘텐츠 생성 분야에서 지속적으로 활용되고 있다[11]. Kim 등[8]은 HyperCLOVA를 이용하여 협찬 블로그 텍스트 생성을 위한 파인튜닝을 진행하였고, 이를 통해 자연스럽게 맥락에 잘 부합하는 콘텐츠를 생성할 수 있음을 보였다. 또한 Oh 등[3]은 HyperCLOVA와 같은 사전 학습된 대규모 언어 모델을 음성 인식 분야에서 파인튜닝하여, 도메인 특화 성능 향상 가능성을 입증하였다.

이러한 연구들은 LLM의 사전학습된 지식을 바탕으로 추가적인 파인튜닝을 통해 특정 도메인에서 성능을 크게 향상시킬 수 있다는 점을 강조한다. 특히, 파라미터 효율적 파인튜닝(Parameter Efficient Fine-Tuning, PEFT) 방식은 적은 데이터와 리소스로도 효율적으로 성능을 개선할 수 있어 주목받고 있다. Kim 등[8]은 협찬 광고 문구와 같은 특정 목적의 텍스트 생성에 PEFT를 적용하여 적은 데이터로도 자연스러운 콘텐츠를 생성할 수 있음을 보였으며, 이는 본 연구에서 HyperCLOVA X 모델의 파인튜닝을 진행하는 근거가 되었다.

그러나 기존의 한국어 LLM 연구는 주로 텍스트 생성 작업만을 다루거나, 광고성 문구 등 비교적 정형화된 문체에 집중되어 있다는 한계가 있었다. 여행 블로그와 같이 자유롭고 서사적인 문체와 이미지와 텍스트의 밀접한 통합을 요구하는 도메인에서는 보다 정교한 접근이 필요하다. 본 연구는 이러한 기존 연구의 한계를 극복하기 위해 HyperCLOVA X 모델을 여행 블로그 데이터 1,600건으로 파인튜닝함으로써, 이미지와 텍스트가 결합된 다중 모달 콘텐츠 생성에서 보다 사실적이고 감성적인 스토리텔링이 가능하도록 하였다.

2.3 Text to Image Generation

텍스트-이미지(Text-to-Image, T2I) 생성 모델 중 대표적인 DALL-E 시리즈는 사용자의 텍스트 입력을 창의적

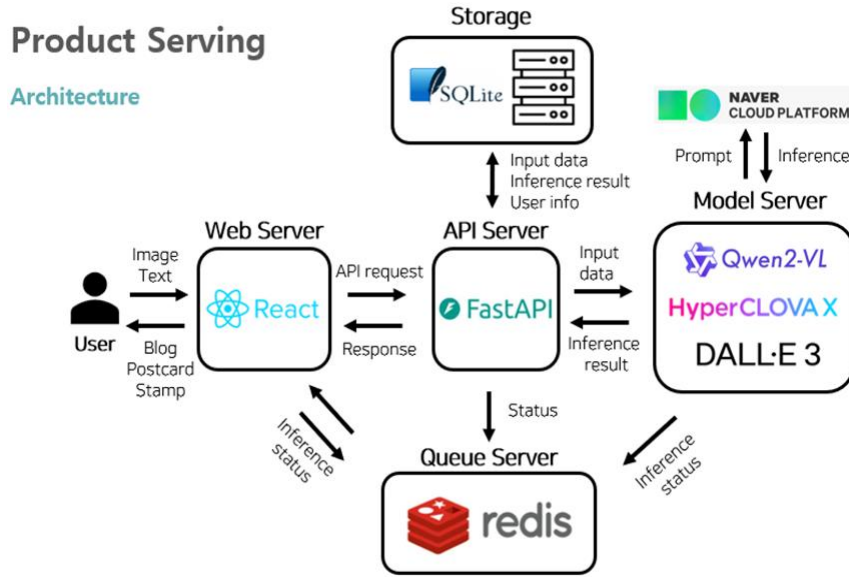


Fig. 1. System Architecture

이고 고품질의 이미지를 생성할 수 있다. 최근의 DALL-E 3 모델은 기존 버전에 비해 사실성과 창의성이 더욱 향상되어, 다양한 시각 콘텐츠 제작에 혁신적인 가능성을 제시하고 있다[12]. Ramesh 등[12]은 DALL-E 시리즈의 창의성과 사실적 묘사 능력을 입증한 바 있다.

본 연구에서는 DALL-E 3를 채택하여 여행 경험의 주요 요소들을 사용자 맞춤형 엽서와 스탬프 이미지로 생성하였다. DALL-E 3의 선택 이유는 생성된 이미지가 실제 여행 경험을 창의적이고 직관적으로 반영할 수 있는 우수한 이미지 품질을 제공하기 때문이다. 또한 프롬프트 엔지니어링(prompt engineering)을 활용하여 여행의 특정 요소(예: 관광지 특징, 여행 감성)를 정확히 표현할 수 있도록 텍스트 입력을 정교화하였다. 이는 기존의 단순한 시각 콘텐츠 생성 사례들과 달리 사용자 경험에 밀접히 연결된 고품질 이미지 생성이 가능하도록 설계된 것이다.

3. Contributions of Proposed Method

기존 연구는 협찬 블로그 생성[8], 요약 기반 장문 생성[9], 관광 AI 잠재력[6]과 같은 개별 요소에서는 성과를 보였으나, 시각-언어 모델(VLM) 기반의 시각적 맥락 추출과 대규모 한국어 LLM 파인튜닝, 그리고 텍스트-이미지(T2I) 모델을 활용한 시각적 콘텐츠 생성을 유기적으로 결합한 연구는 상대적으로 부족하였다. 또한, 생성 콘텐츠의 신뢰성을 높이기 위한 환각(hallucination) 최소화 전략과 같은 AI 안전성에 대한 고려도 충분히 이루어지지 않았다.

본 연구에서 제안하는 "트래블로그" 시스템은 이러한 기존 연구의 한계를 극복하기 위해 다음과 같은 독창적인 접근 방식을 제안한다. 먼저, Qwen2-VL 모델을 활용해 사

진에서 풍부한 시각적 맥락을 추출하고, 이를 기반으로 HyperCLOVA X 모델을 1,600건의 여행 블로그 데이터로 파인튜닝하여 감성적이고 자연스러운 스토리텔링 콘텐츠를 생성한다. 또한, DALL-E 3 모델을 통해 사용자의 여행 경험을 반영한 맞춤형 엽서와 스탬프와 같은 창의적 시각 콘텐츠를 함께 생성하여 여행 기록의 직관성과 독창성을 높인다. 여기에 강력한 시스템 프롬프트 전략을 추가하여 생성 콘텐츠의 환각 현상을 최소화하고 사실성과 표현력을 보장하였다. 결과적으로 본 연구는 개인화 및 자동화된 여행 콘텐츠 생성의 실현 가능성을 입증하였으며, 사용자 설문 결과에서도 여행 특화도(평균 4.6점)와 실용성 측면에서 우수한 평가를 받았다.

III. The Proposed Scheme

3.1 System Architecture Organization

본 연구에서 제안하는 여행 블로그 생성 서비스 "트래블로그(Travelog)"는 웹 서버, API 서버, 모델 서버, 저장소 및 작업 대기열(queue)로 구성되며, 각 구성 요소는 상호 연동하여 사용자의 입력을 처리하고 최적의 결과를 제공한다. 최근 대규모 언어 모델(LLM) 기반 애플리케이션에서는 마이크로서비스 아키텍처와 API 지향 설계 방식이 권장되고 있다 [2][9][14]. 본 연구에서도 이러한 접근 방식을 채택하여 병렬 처리와 확장성을 고려하였다.

사용자는 웹 인터페이스를 통해 텍스트 및 이미지를 입력하고, 이를 기반으로 블로그 본문, 엽서, 스탬프 등의 결과물을 생성할 수 있다.

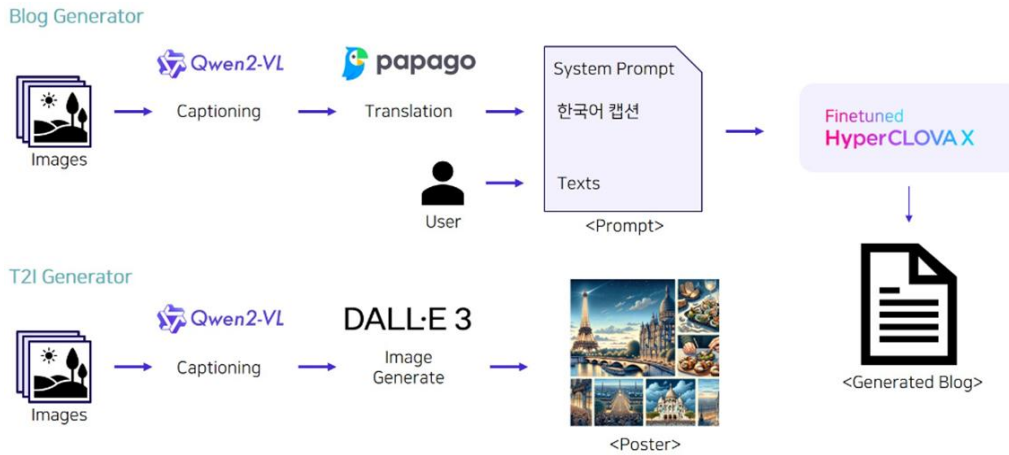


Fig. 2. Service Architecture

웹 서버는 사용자의 요청을 API 서버로 전달하며, API 서버는 입력 데이터를 전처리한 후 작업 대기열에 등록한다. 모델 서버는 VLM 기반 이미지 캡션 생성, 한국어 언어 모델을 통한 여행 블로그 텍스트 생성, 텍스트-이미지 (T2I) 모델을 활용한 엽서-스탬프 이미지를 생성하는 작업을 담당한다. 생성이 완료된 결과물은 API 서버로 전달되어 데이터베이스에 저장되고, 이후 웹 서버를 통해 최종 결과물이 사용자에게 제공된다.

데이터 저장소는 사용자 입력 데이터, 생성된 블로그 콘텐츠 및 이미지 결과물을 효율적으로 관리한다. 또한 작업 대기열은 여러 사용자 요청을 병렬적으로 처리할 수 있도록 하여, 모델 추론 과정에서 발생하는 대기 시간을 최소화하고 요청의 상태를 실시간으로 관리하도록 한다.[1][2] 본 시스템 구성은 사용자 입력에 대한 응답 속도를 최적화하고, 모델 추론 과정에서 발생하는 대기 시간을 최소화하여 보다 안정적이고 효율적인 서비스 환경을 제공하는 것을 목표로 한다. 전체 시스템 아키텍처 구성에 대한 도식은 Fig. 1.과 같다.

3.2 Service Architecture Organization

모델링 과정은 크게 블로그를 생성하는 블로그 생성 모델(Blog Generator)과 엽서나 스탬프를 생성하는 이미지 생성 모델(T2I Generator)의 두 단계로 구성되어 있다. 두 과정은 각각 HyperCLOVA API와 DALL-E API를 활용한다. 두 과정은 작업 대기열(job queue)을 통해 병렬적으로 수행할 수 있으므로, 추론(inference) 시간이 단축되고 전체 처리율(throughput)이 향상된다.[2][3]

추론 과정 초기에 Vision-Language 모델인 Qwen2-VL이 사용자 사진에서 캡션을 추출하여, 해당 시각 정보를 LLM과 T2I 모델에 전달한다. 즉, 사진을 통한 맥락 이해가 선행된 후, Blog Generator가 텍스트를 작성하고 T2I

Generator가 이미지를 생성하는 방식으로 작동한다. 이를 통해 사용자는 여행 사진과 연계된 블로그 본문과 시각적 콘텐츠를 동시에 얻을 수 있다 [10][12]. 서비스 아키텍처 구성에 대한 도식은 Fig. 2.와 같으며, 각 모델의 구체적인 활용 방식은 3.3에서 기술한다.

3.3 Data Processing Scheme

본 연구에서는 블로그 품질의 기준을 정의한 뒤, 해당 기준을 충족하는 블로그 데이터를 수집하는 'bottom-up' 방식을 통해 학습용 데이터셋을 구축하였다.

3.3.1 Data Collection

데이터 수집과정에는 Selenium을 사용한 웹 크롤링 기법을 적용하였으며 정적 및 동적 웹 페이지 모두에서 안정적으로 데이터 확보가 가능하도록 구현하였다.[8]

Table 1. Crawling Dataset

Title	Link
2월 파리 날씨 실시간..	https://blog.naver.com/pinkyluv83/223007256025
프랑스 파리 근교 여행..	https://blog.naver.com/pinkyluv83/223006401527
미국 서부 샌프란시스코..	https://blog.naver.com/pinkyluv83/223033359246
금문교 지나는 빅버스 ..	https://blog.naver.com/pinkyluv83/223033450094
베이 크루즈 유람선 ...	https://blog.naver.com/pinkyluv83/223035613107
Content	
오늘 파리 날씨 화창. 이번주 내내 날씨 아주 좋아요 <image_1> 오늘 오랜만에 프랑스 파리 날씨가 너무 ...	
프랑스 루앙 대성당 <image_1> 지난 주말을 이용하여 프랑스 파리 근교 도시 루앙 Roen에 다녀왔다. 루앙은 프랑스에서 ... <image_1> <image_2> 뮤어우즈 국립공원은 지구상에서 가장 큰 나무 중 하나인 레드우드 Redwoods 가 있는 곳이고, ...	
샌프란시스코의 마스코트 케이블카 <image_1> 안녕하세요. 미국 샌프란시스코 여행을 2월말에 다녀왔어요....	
<image_1> 안녕하세요. 저번주에 샌프란시스코를3박4일 다녀왔어요. 샌프란시스코는캘리포니아에 위치하고 있어서 ...	

총 1600개의 공신력 있는 블로그 포스트를 크롤링하였으며, 여행지와 관련된 블로그, 이미지, 태그 정보를 함께 수집하였다. Table 1은 크롤링 블로그 데이터의 구성을 보여준다. 블로그 게시글의 원문 텍스트 데이터뿐만 아니라 이미지 배치를 위한 이미지 태그 정보도 함께 수집하여 블로그 콘텐츠의 구조적 특성을 효과적으로 반영할 수 있었다. 이는 추후 모델 학습에서 텍스트와 이미지 간의 관계를 의미있게 학습하도록 한다. 또한 데이터셋의 품질을 높이기 위한 간단한 전처리 작업을 진행했다. 불필요한 특수 문자, 이모지, 의미 없는 공백을 제거하고 블로그 글이 너무 길거나 블로그의 마지막 부분에 추가되는 광고성 문구를 방지하기 위해 블로그 당 토큰을 4000개로 제한하였다.

3.3.2 Data Preprocessing and Dataset Construction

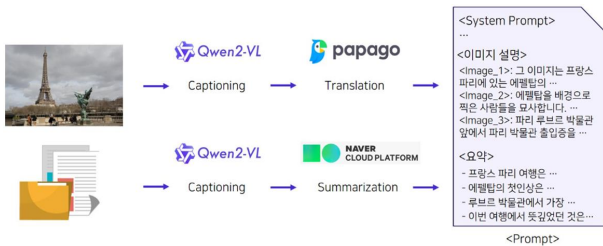


Fig. 3. Fine-Tuning Dataset Pipeline

본 연구에서는 실제 서비스 사용자의 입력과 유사한 학습용 데이터셋을 구축하기 위해 Fig 3와 같은 과정을 거쳤다. VLM을 통해 블로그 이미지의 캡션 정보를 생성하였고, Naver Cloud Platform에서 제공하는 요약 API를 활용하여 블로그의 내용을 요약하고, 이를 여행 정보 (description)로 활용하였다. 위의 과정을 통해 생성한 사용자 입력 프롬프트(user prompt)와 추후 서술할 시스템 프롬프트(System Prompt)를 결합하여 Table 2와 같은 학습용 데이터셋을 구축하였다.

Table 2의 Input Text가 사용자의 입력 정보와 유사한 형태의 학습용 입력 데이터이고, Completion Blog는 크롤링을 통해 수집한 이미지 태그가 부착된 실제 블로그이다. 이를 통해 모델이 이미지 및 핵심 요약 정보 기반으로 자연스러운 여행 블로그를 생성하도록 유도하였다.

Table 2. Fine Tuning Dataset

Input Text
<image0> 이미지에는 검은색 셔츠 유니폼을 입은 사람이 다양한 종류의 음식으로 ... <image1> : ... <image2>: ...
<image0>: 그 이미지는 프랑스 파리에 있는 에펠탑의 경치를 묘사하고 있습니다. ... <image1> : ... <image2>: ...
<image0>: 그 이미지는 이탈리아 로마의 유명한 랜드마크인 트레비 분수를 ... <image1> : ... <image2>: ...
<image0>: 그 이미지는 고풍스러운 마을의 매력적인 거리 풍경을 묘사하고 있습니다. ... <image1> : ... <image2>: ...
<image0>: 그 이미지는 무성한 녹지로 둘러싸인 크고 우아한 건물의 그림 같은 장면을 ... <image1> : ... <image2>: ...
Completion Blog
<image0> 지난 다낭 가족여행 시 리조트 음식과베트남 현지식만 ... <image1> ... <image2> ...
<image0> 프랑스 파리 여행은 가볼만한곳이 많아 최소 4박 5일 일정이 필요한데요... <image1> ... <image2> ...
<image0> 이탈리아는 타 유럽 국가들에비해서 연중 날씨가 따뜻한 편이라 ... <image1> ... <image2> ...
<image0> 영국 런던 여행을 할 때 london만 구경하면 아쉬운 점이 많으니, ... <image1> ... <image2> ...
<image0> 태국 방콕에는 세계적인 5성급 유명 체인 호텔들이 많은데요. ... <image1> ... <image2> ...

3.4 Modeling and Fine-Tuning Scheme

3.4.1 Vision-Language Model (Qwen2-VL)

Qwen2-VL은 여행 사진의 맥락을 이해하고 캡션을 생성하는 시각-언어 모델(VLM)로, 다국어 이미지 인식이 가능하며 풍부한 묘사 능력을 갖추었다[10][15]. 기존 LLaMA 3.2 Vision 등과 비교 시, 정확한 위치 정보(예: 배경, 인물, 랜드마크 명칭)에 대한 묘사가 뛰어난 장점이 관찰되었다.

사진 input을 캡션으로 변환하기 위해 활용된 VLM은 Qwen2-VL이다. Qwen2-VL은 다양한 해상도의 이미지를 효과적으로 처리하고, 다양한 언어를 이해할 수 있기 때문에 여행 블로그에 적합하다고 판단했다. 실제로 다른 VLM 모델인 Llama 3.2 와 비교하였을 때 묘사가 풍부하고 위치정보를 정확하게 포함하고 있어 선정하였다. 여행블로그에 알맞는 캡션을 생성하기 위해 “I want to know where it is and description about this image” 를 input으로 주어 여행지에 대한 정보와 함께 캡션을 생성하였다.

3.4.2 Korean Language Model (HyperCLOVA X)

블로그를 작성하기 위한 한국어 언어 모델로 HyperCLOVA X를 활용했다. HyperCLOVA X는 한국어에 최적화된 대규모 언어 모델로, 다른 모델 대비 미세 조정(fine-tuning)을 통해 특정 도메인의 자연스러운 한국어 문장 생성 품질을 크게 향상시킬 수 있다. 수집한 블로그 데이터셋을 기반으로 HyperCLOVA X를 PEFT 방식으로 파인튜닝해 사실적인 여행 블로그를 작성할 수 있게 되었

으며 주어진 이미지에 대한 설명과 여행자의 느낀점을 바탕으로 여행지에 대한 설명을 보다 정교하게 생성하는 모델을 개발하였다 [16].

파인튜닝은 Naver Cloud Platform에서 제공하는 HCX-003 모델을 활용했으며, 에폭 2, 학습률 1e-4로 설정하여 학습하였다. 세부 파라미터로는 Naver Cloud Docs의 가이드에 따라 topP 0.8, topK 0, 최대 토큰수 1000개, temperature 0.3, repeat penalty 7로 설정하여 개선된 성능의 여행 블로그 생성 LLM을 제작할 수 있었다.

3.4.3 Text-to-Image Generative Model (DALL-E3)

서비스의 시각적 요소를 제공하기 위해 Text-to-Image 모델인 DALL-E 3를 활용했다. DALL-E 3는 텍스트 입력을 기반으로 사실적이고 창의적인 이미지를 생성할 수 있는 최신 이미지 생성 모델로, 여행지의 특징을 반영한 고품질 이미지를 생성하는 데 적합하다. DALL-E 3를 활용하는 과정에서, 생성된 이미지가 여행지의 실제 모습과 유사하도록 프롬프트를 정교하게 설계하였다.

3.5 Prompt Engineering

3.5.1 System Prompt Organization

LLM이 일정된 양질의 블로그를 생성할 수 있도록 강력한 시스템 프롬프트를 적용했다. 시스템 프롬프트는 AI가 일관된 톤과 형식을 유지하도록 하는 핵심 요소로, 본 연구에서도 블로그 질을 결정하는데 가장 큰 역할을 담당하였다.

트래블로그에서는 AI Safety와 환각을 최소화하고 인간 문체와 최대한 유사한 블로그 생성 목적으로 Table 3과 같은 시스템 프롬프트를 설정했다 [17]. 연구에서 활용한 시스템 프롬프트는 3가지의 특징을 갖는다.

- 1) 블로그 대회 출전이라는 상황을 부여: AI가 좋은 블로그를 생성하면 1등을 할 수 있다는 보상 설정
- 2) 대회 규칙을 통한 금지사항 명시: 강력한 어투를 활용해 AI가 해당 규칙을 절대 위반하면 안된다는 상황 설정
- 3) 심사기준, 팁을 통한 스타일 가이드: 구체적인 기준 설정, 다양한 예시를 통해 문체, 감성 표현에 대한 가이드 부여

이러한 시스템 프롬프트를 통해 안정된 성능의 모델을 확보했으며 시스템 프롬프트만으로도 성능이 크게 향상되었다.

Table 3. System Prompt

System Prompt
당신은 대한민국 최고의 여행 블로거로서 세계 최고 여행 블로거를 뽑는 대회에 참가했습니다.
...
심사기준에 부합하는 좋은 블로그를 작성해 세계 최고의 여행 블로거가 됩시다!
<대회 규칙>:대회 규칙을 어길 시 즉시 실격됩니다.
...
4.사실에 입각한 정보만 블로그에 담아야 하며, 블로그에대한 오류가 있을 경우 허위 사실 유포죄에 의해 처벌 받을 수 있음.
...
<심사 기준>
...
2.불필요한 내용이나 환각없이 주어진 정보만 활용하고 있는가.
...
5.블로그 글에 개인적인 정보혹은 위험성있는 단어를 포함하고 있지 않은가.
...
세계최고의 여행 블로거는 글을 작성할 때 아래의 사항을 따른다고 합니다.
...
4."고즈넉한 분위기","탁 트인 바다와 하늘","형형색색 랜턴"등 감각적인 표현을 활용해 ...
5."너무 좋았어요!","완전 만족했어요!","정말 인상적이었습니다."등 긍정적인 표현과 감탄사를 적절히 사용
위의 사항을 모두 포함해 글을 작성하십시오. 만일 이에 위배될 경우 당신은 00년 이하의 징역 0000만원 이상의 벌금을 부여 받게 됩니다.

3.5.2 User Prompt Organization

사용자 프롬프트는 사용자의 입력을 반영해 LLM의 input을 생성하는 역할을 한다. 본 연구에서는 사용자의 입력으로 여행에 대한 정보와 VLM을 통해 생성한 이미지들에 대한 캡션, 각 이미지에 대한 설명을 입력받는다.

여행에 대한 정보는 여행 시기, 여행 나라, 도시, 여행 동반자를 입력받는다. 이미지 캡션은 이미지 태그 <image>를 통해 각 이미지별 정보를 부여하였고, 각 이미지에 대한 사용자의 설명을 추가해 각 이미지 별로 여행의 경험과 느낀 점을 블로그 생성에 활용한다. 구체적인 사용자 프롬프트의 구성은 Table 4와 같다.

Table 4. User Prompt

User Prompt
<Image_1>:그 이미지는 프랑스 파리에 있는 에펠탑의 ...
<Image_2>: 에펠탑을 배경으로 찍은 사람들을 묘사합니다.
...
<Image_3>:파리 루브르박물관 앞에서 파리 박물관 출입증을 ...
...
<요약>
-프랑스 파리 여행은 ...
-에펠탑의 첫인상은 ...
-루브르박물관에서 가장 ...
-이번 여행에서 뜻깊었던 것은...

3.5.3 AI Safety

"트래블로그" 시스템은 AI 안전성(AI Safety)을 위해 생성 콘텐츠의 신뢰성, 사실성, 그리고 사용자 경험의 품질을 보장하는 핵심 요소들을 고려하여 설계되었다. 대규모 언어모델(LLM)과 시각-언어모델(VLM)의 발전은 콘텐츠 생성 능력을 획기적으로 향상시켰으나, 환각(hallucination), 부정확한 정보 생성, 그리고 비일관성 등의 위험을 동반한다. 이러한 문제를 해결하고, 관광 산업 맥락에서 사용자에게 안전하고 신뢰할 수 있는 결과를 제공하기 위해 본 연구는 다층적인 안전성 확보 전략을 적용했다.

첫째, HyperCLOVA X에 적용된 시스템 프롬프트는 AI 안전성의 핵심 장치로 기능한다. 이 프롬프트는 가상의 "세계 최고 여행 블로거 대회"라는 상황을 설정하여, AI가 엄격한 규칙(예: 주어진 여행 정보와 이미지 외 추가 내용 금지, 사실에 입각한 정보만 포함)과 심사 기준(자연스러운 문체, 감정 표현의 풍부함, 이미지와 텍스트의 조화)을 준수하도록 유도했다. 특히, '규칙 위반 시 실격'과 같은 강력한 제약을 명시하여 모델이 임의로 허구를 생성하거나 맥락에서 벗어난 출력을 방지하도록 하였다. 이는 LLM의 자유로운 생성 경향을 억제하며, 사용자가 제공한 데이터에 충실한 결과를 보장한다.

둘째, 파인튜닝 과정에서 데이터 품질과 학습 설계를 신중히 관리하여 안전성을 강화했다. 본 연구는 초기 1,600건의 크롤링 데이터를 "좋은 블로그" 기준(광고 최소화, 경험 중심 서술, 감정 표현 포함)에 따라 400건으로 정제했다. 이 과정에서 불필요한 광고나 개인 정보가 학습 데이터에 포함되지 않도록 텍스트 정제를 수행했다. 또한, 과적합과 과도한 창작성을 억제하기 위해 파인튜닝을 2 에폭으로 제한했으며, 이는 모델이 기존 데이터의 문체와 감정 표현을 학습하면서도 새로운 허구를 생성하지 않도록 균형을 맞췄다. 실험 결과, 파인튜닝 후 생성된 블로그는 "ㅋ"나 "ㅠㅠ"와 같은 한국어 특유의 자연스러운 표현을 반영하여 인간 문체에 근접하게 생성하도록 하였다.

셋째, Qwen2-VL의 이미지 캡션 생성 과정에서 추가 검증 단계를 도입하여 시각-텍스트 일관성을 확보했다. VLM은 이미지에서 풍부한 맥락을 추출하지만, 때로 과도한 추론으로 사실과 다른 설명을 생성할 수 있다. 이를 방지하기 위해, 캡션 생성 후 사용자 입력(예: 방문지, 날짜)과 비교하여 불일치 여부를 점검하는 필터링 로직을 추가했다. 예를 들어, 사용자가 제주도를 입력했는데 캡션에서 "유럽의 강"을 언급한다면 이를 수정하도록 시스템이 개입한다. 이는 다중 모달 시스템에서 발생할 수 있는 모순을 줄이고, 신뢰성을 높이는 데 기여하였다.

이러한 안전성 전략의 효과는 사용자 평가를 통해 확인되었다. 설문 조사에서 환각 발생률이 기존 모델 대비 현저히 감소했으며(정확한 수치 분석 중), 생성된 블로그와 업서가 사용자 제공 정보와 높은 일치도를 보인다고 응답받았다. 이는 "트래블로그"가 AI 기반 콘텐츠 생성에서 안전성과 품질을 동시에 달성했음을 입증한다. 나아가, 본 접근은 관광 분야에서 AI를 활용한 창작 도구가 가져야 할 윤리적 책임과 실용성을 충족, 향후 유사 시스템 개발에 중요한 시사점을 제공한다.

3.6 Service Implementation Details

3.6.1 Front-End Implementation

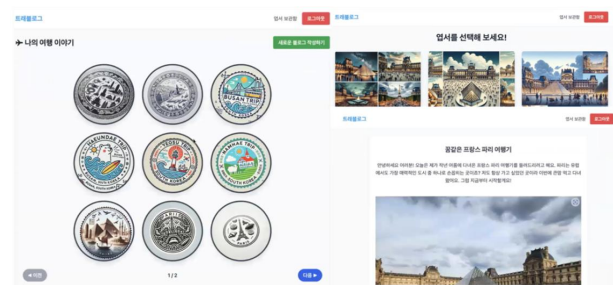


Fig. 4. Web UI

"트래블로그"의 프론트엔드는 사용자가 여행 블로그와 업서를 생성하고 관리하는 인터페이스를 제공하며, 시스템과의 상호작용을 직관적으로 안내한다.

사용자는 로그인 페이지에서 인증을 거쳐 업서 보관함으로 이동한다. 여기서 기존 프로젝트를 그리드 형태로 확인하거나 새 작업을 시작할 수 있으며, 선택 시 상태(1~6)에 따라 해당 단계로 전환된다. 사진 업로드 단계에서는 여러 이미지를 선택해 업로드하고, 썸네일 미리보기로 확인한다. 이어지는 설명 작성 페이지에서는 각 이미지에 방문지, 날짜, 동행자 등의 메타데이터와 경험을 입력한다. AI 추론 단계에서는 WebSocket을 통해 진행 상황을 실시간으로 표시하며, 완료 후 생성된 블로그 초안과 업서 후보를 제시한다. 스탬프 확인 단계에서는 AI 생성 스탬프를 검토하고, 업서 선택 단계에서 원하는 후보를 고른다. 편집 페이지에서는 초안을 수정하며 선택된 업서와 함께 내용을 조정하고, 최종 결과 페이지에서는 완성된 콘텐츠를 확인한다. React와 Tailwind CSS로 구현되었으며, RESTful API를 통해 백엔드와 연동되어 사용자 경험을 최적화한다 [18]. 생성된 블로그와 업서, 스탬프를 보여주는 Web UI는 Fig. 4에서 확인할 수 있다.

3.6.2 Back-End API Server Implementation

본 연구에서 백엔드 서버는 여행 블로그 콘텐츠 생성, 사용자 인증, 프로젝트 관리 기능을 제공하도록 설계되었다. 서버는 데이터베이스와의 원활한 연동을 위해 객체 관계 매핑을 활용하고, 비동기 처리 방식을 적용하여 대규모 요청에도 안정적인 응답 속도를 유지할 수 있도록 구성되었다.

사용자의 인증 및 권한 관리는 보안성을 고려하여 설계되었으며, 로그인 및 프로젝트 관리와 같은 기능을 수행하는 모든 엔드포인트에서 사용자 인증이 필수적으로 요구된다. 사용자는 회원가입 및 로그인 후 개인 프로젝트를 생성할 수 있으며, 이미지 및 텍스트 데이터를 업로드하고 생성된 블로그 콘텐츠를 관리할 수 있다.

이미지 캡션 생성과 텍스트 기반 이미지 생성 기능은 사전 학습된 모델을 활용하여 구현되었다. 사용자가 이미지를 업로드하면, 시스템은 해당 이미지의 내용을 분석하여 설명을 생성하고, 이를 바탕으로 블로그 작성에 활용할 수 있다. 또한, 사용자의 텍스트 입력을 기반으로 이미지를 생성하는 기능도 포함되며, 이는 여행 포스터 및 엽서와 같은 시각적 요소를 제작하는 데 활용된다.

백엔드 서버는 비동기 기반의 작업 처리를 통해 요청을 효율적으로 관리하고, 사용자의 입력이 대기열에 추가된 후 빠르게 응답을 제공할 수 있도록 설계되었다 [19]. 이를 통해 사용자는 서버의 부하와 관계없이 원활하게 콘텐츠를 생성하고 수정할 수 있으며, 보다 직관적인 인터페이스 환경에서 서비스를 이용할 수 있다.

3.6.3 Storage and Queue Management

서비스 내에서 생성되는 블로그 콘텐츠, 사용자 입력 데이터, 모델 추론 결과 등은 안정적인 저장 및 빠른 처리를 위해 데이터 저장소와 작업 대기열을 활용하여 관리된다. 생성된 데이터는 저장소에 보관되며, 메타데이터를 함께 관리하여 파일 검색 효율성과 접근성을 향상시켰다.

데이터를 효과적으로 처리하기 위해 데이터 저장 방식은 확장성을 고려하여 설계되었다. 프로젝트 생성 및 이미지 업로드 시 데이터베이스에 메타정보를 기록하고, 이를 통해 사용자가 자신의 콘텐츠를 손쉽게 조회하고 관리할 수 있도록 하였다. 또한, 사용자 입력을 기반으로 생성된 콘텐츠는 별도로 저장되며, 향후 재사용 및 편집이 가능하도록 설계되었다.

작업 처리를 위한 대기열 시스템을 도입하여 서버의 안정성을 높이고 응답 속도를 개선하였다 [20]. 사용자의 요청이 들어오면 작업이 대기열에 추가되며, 처리가 완료된

후 결과를 데이터베이스에 저장하고 상태를 갱신한다. 이를 통해 사용자 인터페이스에서는 작업 진행 상황을 실시간으로 확인할 수 있으며, 요청이 많을 경우에도 서버의 부하를 최소화할 수 있다.

본 연구에서는 서버와 저장소, 대기열 시스템을 유기적으로 연동하여 여행 블로그 생성 과정이 원활하게 진행될 수 있도록 하였으며, 데이터의 저장과 활용이 효율적으로 이루어질 수 있도록 구성하였다.

IV. Experiments

4.1 LLM Benchmark

4.1.1 BERTScore

BERTScore[13]는 BERT의 사전 학습된 컨텍스트 임베딩을 활용하여 두 문장 간의 의미적 유사성을 측정하는 자연어 처리(NLP) 평가 지표로 후보 문장과 참조 문장의 단어 임베딩을 비교하고, 코사인 유사도를 계산하여 문장의 의미적 일치도를 평가한다. BERTScore는 자연어 생성 작업에서 주로 활용되며 단어나 문장의 의미 명확성(semantic clarity) 및 문맥적 의미 적합성(contextual semantic appropriateness)을 평가할 수 있어 다양한 LLM 파인튜닝 성능 평가에 활용된다.

본 연구에서는 생성형 AI가 작성한 여행 블로그와 실제 여행 블로그 간의 의미적 유사성을 측정하기 위해 BERTScore를 활용하였다. 실험을 위해 네이버에서 활동 중인 여행 블로거 3인의 포스트 100개를 수집하였으며, 해당 블로거들은 각각 자니(이웃 약 4,500명), Kimi(이웃 약 20,000명), 그루터기(이웃 약 65,000명)로 다양한 팔로워 규모와 영향력을 가진다. 이처럼 서로 다른 특성을 지닌 블로거의 데이터를 활용함으로써, 실제 여행 블로그의 콘텐츠 특성과 생성형 AI가 생성한 블로그 간의 의미적 차이를 폭넓게 분석할 수 있도록 하였다.

수집한 블로그의 이미지를 통해 Caption(Cap)을 생성하고, ChatGPT를 통해 여행에 대한 Description(Desc)을 생성하였다. BERTScore를 통해 모델이 생성한 블로그와 실제 블로그 간의 의미적 유사성을 측정해 User Prompt(Description, Caption)와 System Prompt(SP)의 유무, Fine-tuning(FT) 유무에 따른 성능을 비교하였다.

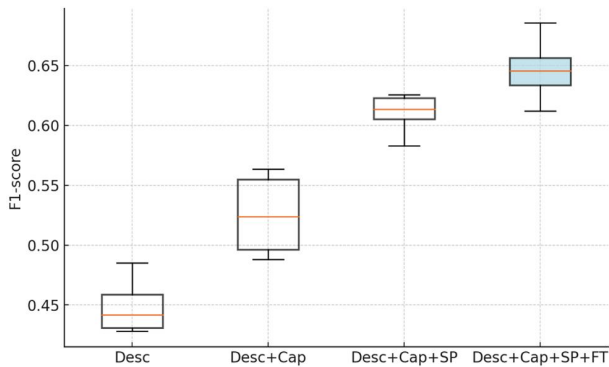


Fig. 5. Boxplot of Generated Blog BERTScore (F1-score)

본 연구에서 개발한 모델은 프롬프트 엔지니어링, 파인 튜닝을 거듭할수록 향상된 성능을 갖추게 되었으며, Fig 5에 따르면 일반적인 블로그 생성 AI[8]와 같이 Description만 활용한 결과 0.45의 BERTScore를 기록한 반면 이미지 캡션 정보를 활용한 결과 8%의 성능 향상이 있었다. 또한 System Prompt(SP)의 유무에 따라 8% 이상의 성능 향상, Fine Tuning(FT) 이후 4%의 성능 향상이 있었다.

Table 5. Comparison of Generated blog by models

Models	Recall	Precision	F1-score
Desc	0.47	0.42	0.45
Desc+Cap	0.56	0.50	0.53
Desc+Cap+SP	0.61	0.62	0.61
Desc+Cap+SP+FT	0.64	0.66	0.65

Table 5을 통해 각 요소의 재현률과 정밀도를 확인할 수 있다. User Prompt를 통해 실제 블로그와의 재현율을 높였다. 그러나 상대적으로 정밀도가 낮아 환각이 많이 발생했다. 반면 System Prompt(SP)는 환각을 감소하는데 큰 기여를 했으며 정밀도를 12% 상승시키면서 생성된 블로그 속에 실제 블로그에 포함되지 않은 거짓 정보가 많이 감소되었다. Fine-Tuning(FT) 과정은 재현률과 정밀도를 모두 상승 시키며 최적의 성능 달성에 기여하였다. Table 6를 통해 실제 블로그와 생성된 블로그를 비교한 결과, 생성된 블로그는 실제 블로그에서 전달하고자 하는 핵심 내용을 충분히 포함하고 있었다. 나아가, 이미지 기반의 정서적 맥락을 반영함으로써 보다 풍부한 감정 표현을 보여 주었으며, 결과적으로 높은 품질의 블로그 콘텐츠가 생성된 것을 확인할 수 있었다.

Table 6. Comparison of real and generated blog

Real blog
안녕하세요 여행 블로거 00입니다 오늘은사이판 여행가기 좋은 시기, 사이판 여행코스인 마나가하섬에 대해서 소개해드리며 ...
마나가하섬은 배를 타고 들어가야하는 아름다운 사이판의 섬이랍니다.푸르른 에메랄드빛의 바다색이 이국적이면서 휴양지에 온듯한 느낌을 주는 곳이었어요.
사이판 마나가하섬 물도 엄청 투명하고 맑아서 수심이 얕음에도 불구하고 다양한 물고기들과 해양생물체들을 볼 수 있었어요
Generated blog
안녕하세요 여러분! 오늘은 제가 다녀온 사이판 섬의 명소이자 대표 관광지인 마나가하섬을 소개하려고 해요.
사이판 본섬에서 배 타고 15분이면 도착하는 이곳은 선착장부터 이렇게 예쁜 알록달록 표지 판이 반겨줘요. 파란 하늘이랑 너무 잘 어울리지 않나요? 여기서 다들 인생샷 하나씩 건져 가세요!
물 색깔 좀 보세요... 진짜 미쳤어요!! 바닥이 흰히 들여다보일 정도로 깨끗한 바닷물이라 스노쿨링 하기에 딱 좋아요. 물고기 종류도 많아서 시간 가는 줄 모르고 놀았답니다.

4.1.2 Performance Comparison between Bloggers

본 연구는 모델의 일반화 가능성을 점검하기 위해 네이 버에서 활동 중인 여행 블로거 세 명 (자니·Kimi·그루터기)의 게시글 100편씩을 별도로 평가하였다. 각 블로거 데이터셋에 대해 동일한 실험 절차로 BERTScore(F1)를 산출한 결과, 자니 0.652, Kimi 0.647, 그루터기 0.655로 모두 0.65 내외의 근접한 값을 보였다. 즉, 모델 성능이 특정 필자 스타일에 과도하게 치우치지 않고 비교적 균일하게 유지되어 도메인 일반화 가능성이 확인되었다.

4.2 User Survey

모델의 성능을 보다 직관적이고 현실적으로 검증하기 위해 일반 사용자 200명을 대상으로 설문 평가를 수행하였다 [21].

응답자는 온라인 여행 커뮤니티(예: 네이버 ‘여행·맛집’ 카페, 뽀뿌 여행포럼 등) 60%와 SNS(인스타그램·페이스북) 40%에서 자발적 모집(자발적 클릭·참여형 링크 배포) 하였다. 연령 분포는 20대 45%(n = 90), 30대 35%(n = 70), 40대 20%(n = 40)로 20~40대가 고르게 포함되었으며, 성별은 남성 52%(n = 104), 여성 48%(n = 96)로 구성되었다.

"Travelog" 시스템 사용자 경험 및 콘텐츠 품질 평가 설문

안녕하세요! 본 설문은 AI 기반 자동 여행 블로그 생성 시스템 "Travelog"의 성능과 사용자 경험을 평가하기 위해 마련되었습니다. "Travelog"는 사용자의 여행 사진과 설명을 입력받아 블로그 텍스트, 업서, 스탬프를 자동 생성하는 서비스입니다.

- 목적: "Travelog" 시스템의 콘텐츠 품질, 여행 특화도, 사용자 경험, AI 안전성을 평가하고, 이를 통해 시스템 개선 및 관광 분야 AI 콘텐츠 생성 기술 발전에 기여하고자 합니다.
- AI 생성: 설문은 시가 생성한 블로그, 업서, 스탬프를 기반으로 하며, 사용자의 피드백을 AI 모델 개선에 반영할 계획입니다.
- 응답 처리: 모든 응답은 익명으로 처리되며, 연구 목적 외에 제3자에게 제공되지 않습니다. 수집된 데이터는 통계 분석 후 익명화하여 보관되며, 통계법 제 33조 2항에 따라 보호됩니다.
- 소요 시간: 본 설문은 사용자의 실제 서비스 경험과 병행하여 약 10분 소요되며, 참여해 주신 의견은 시스템 개발에 큰 도움이 됩니다.

Google에 로그인하여 진행상황을 저장하세요. 자세히 알아보기

* 표시는 필수 질문임

이메일 *

내 답변

(1) 콘텐츠 품질
(1=전혀 그렇지 않다, 2=그렇지 않다, 3=보통이다, 4=그렇다, 5=매우 그렇다)

생성된 블로그가 귀하의 여행 경험을 잘 반영한다고 생각하십니까? *

Fig. 6. Example of Survey form

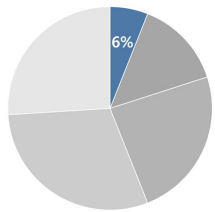
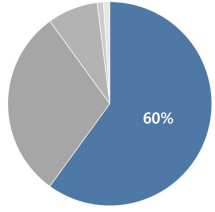
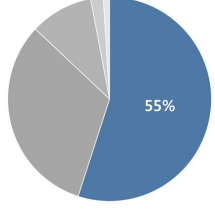
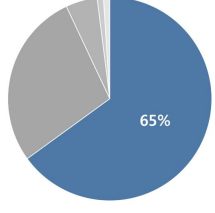
설문은 콘텐츠 품질, 여행 특화도, 사용자 경험, AI 안전성의 네 가지 주요 기준으로 구성되었으며, 각 문항은 5점 리커트 척도로 평가하였다. 설문은 사용자들이 "Travelog" 시스템의 웹 인터페이스에서 직접 여행 사진을 업로드하고, 생성된 블로그 텍스트, 업서, 스탬프 등의 콘텐츠를 경험한 뒤 응답하도록 설계하였다(Fig. 6 참조). 또한 설문 말미에는 개방형 질문을 추가하여 사용자의 정성적 의견을 심도 있게 분석하였다.

설문 결과, 콘텐츠 품질 측면에서 생성된 블로그가 여행 경험을 잘 반영하고 있다고 응답한 비율이 매우 높았으며 평균 점수는 4.5점으로 나타났다. 이미지 캡션과 블로그 텍스트 간의 일치도 역시 평균 4.3점으로 우수한 평가를 받았다. "여행 특화도" 항목에서는 Travelog 시스템이 일반적인 블로그 생성기보다 여행 콘텐츠에 더 적합하다는 응답이 많았으며, 평균 4.6점의 높은 평가를 받았다. 생성된 업서와 스탬프의 유용성 역시 4.4점의 긍정적인 평가를 기록하였다.

"사용자 경험" 항목에서는 인터페이스의 직관성이 평균 4.7점, 콘텐츠 생성 과정의 속도와 편리성이 평균 4.6점으로, 전반적으로 사용자들이 시스템을 쉽게 이용할 수 있음을 확인하였다. 또한 AI 안전성 측면에서도 콘텐츠의 허구적 정보 포함 여부가 낮게 평가되어(허구 포함 없음 평가 4.2점), 감정 표현 및 문체의 자연스러움 또한 4.5점으로

높게 나타났다. 전체적인 설문 결과는 Fig. 7에서 시각화하여 제시하였다.

추가로 실시한 개방형 질문에 대한 정성적 분석 결과, 많은 응답자들이 블로그 자동화 시스템이 여행 후 콘텐츠 작성의 심리적 부담을 현저히 줄여 실제 여행 기록을 더 많이 남기게 되었다고 답하였다. 대표적인 사용자 피드백 중 하나는 "감정 묘사가 매우 자연스러워 여행 당시를 다시 생생히 떠올리게 만든다"였으며, 또 다른 사용자는 "시각적 콘텐츠가 블로그 경험을 더욱 풍부하게 만들어 실제 여행했던 장소로 돌아간 느낌이다"라고 언급하였다. 이를 통해 Travelog 시스템이 단순한 콘텐츠 자동화를 넘어 사용자의 여행 기록 경험의 질을 높이고 여행의 기억을 더욱 깊이 간직할 수 있는 수단으로서 실질적 효용을 제공함을 확인할 수 있었다.

Questions	Survey Result
일반 블로그 생성기(예: 가제트AI, Jorin Blog, ChatGPT 기반 생성기 등)에 얼마나 익숙하십니까?	
생성된 블로그가 여행 경험을 잘 반영한다고 생각하십니까?	
이미지 캡션과 블로그 텍스트가 일치하는 수준은 어떠신가요?	
이 시스템이 일반 블로그 생성기보다 여행에 특화되어 있다고 느끼십니까?	

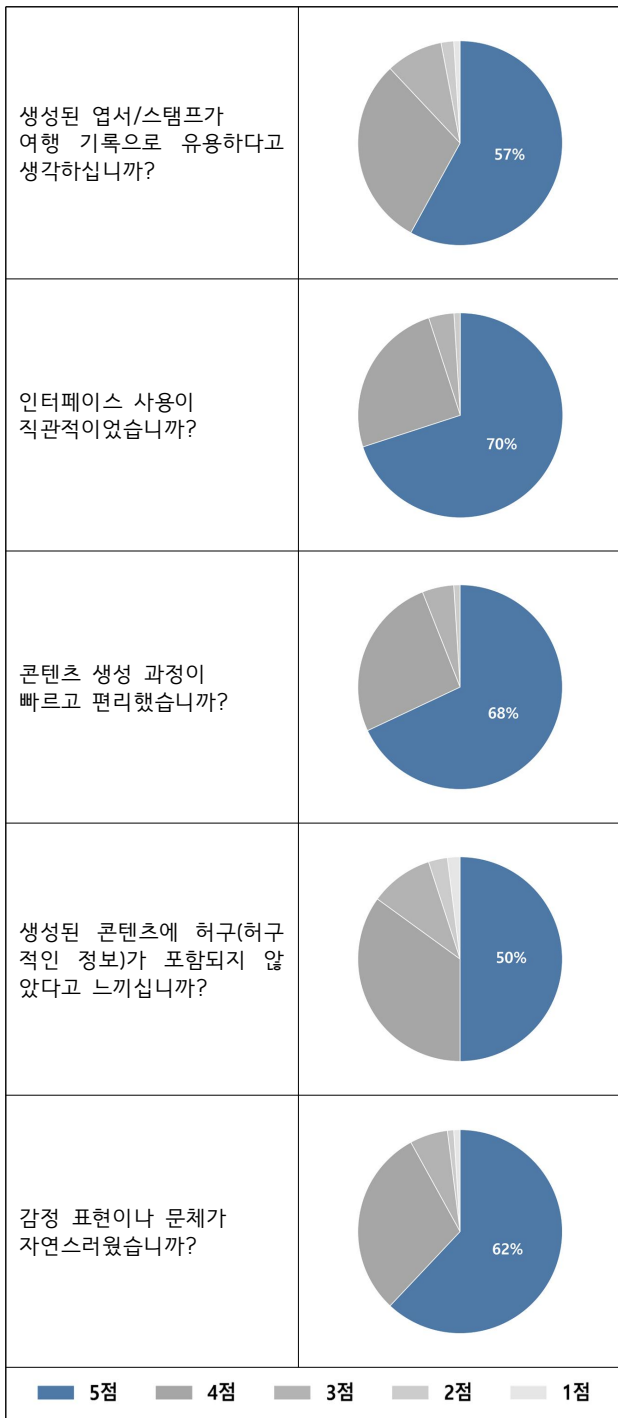


Fig. 7. Survey Result

V. Conclusions

본 논문은 시각-언어 모델(VLM) 기반 맥락 추출, HyperCLOVA X 언어모델 파인튜닝, DALL-E 3를 활용한 텍스트-이미지(T2I) 생성을 결합한 자동 한국어 여행 블로그 생성 시스템 "Travelog"를 제안하였다.

Qwen2-VL은 여행 사진에서 풍부한 시각 맥락을 추출하며 감성적인 캡션을 생성하고, 1,600건의 고품질 여행 블로그 데이터를 활용한 HyperCLOVA X 파인튜닝은 자연스럽고 스토리텔링 중심의 텍스트를 만들어낸다. 또한, 여행 특화 프롬프트 엔지니어링을 통해 DALL-E 3가 맞춤형 엽서와 스탬프를 생성하며 창의적 가치를 더한다. 강력한 시스템 프롬프트는 환각을 최소화하고 사실성과 표현력을 보장하며, 실험 결과로 입증되었다.

BERTScore F1 실험에서 파인튜닝 모델은 기존 방식을 초월하였고, 시스템 프롬프트는 일관성을 개선했다 (0.64~0.66). 사용자 설문조사에서는 85%가 여행 특화도를, 90%가 입력 일치도를 인정하며 만족도가 75%에 달했다. 정성적 분석 결과, 블로그 자동화로 인해 여행 후 콘텐츠 작성 부담이 크게 줄어들어 실제 여행 기록을 더 많이 남기게 되었다는 의견이 많았다.

본 연구의 자동화 콘텐츠 생성 도구는 호텔의 개인 맞춤형 추천 콘텐츠 제작이나 지역 축제 및 문화행사 브랜딩 콘텐츠 자동 생성 등 다양한 관광 관련 산업 분야로 확장 가능하다. 그러나 본 연구는 한국 중심의 데이터로 인해 지역적 편향성의 한계를 가지며, 사용된 언어 모델의 창의적 표현이 제한적이라는 한계가 존재한다. 향후 연구에서는 다양한 문화권의 여행 데이터를 확보하고, 생성 콘텐츠에 대한 품질 평가 자동화 방법 연구, 그리고 멀티모달 정보 융합의 고도화 방안을 모색하여 이러한 한계를 극복할 예정이다.

ACKNOWLEDGEMENT

This research was supported by the Republic of Korea Government (Ministry of Science and ICT) through the research fund of the National Research Foundation of Korea (NRF) and Information & Communications Technology Planning & Evaluation (IITP) under grants NRF-2021R1C1C2095450, RS-2023- 00242528, RS-2024-00437756.

REFERENCES

[1] Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y.,

- Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., Wen, J.-R., "A Survey of Large Language Models," *arXiv preprint arXiv:2303.18223*, Mar. 2023. [Online]. Available: <https://arxiv.org/abs/2303.18223>
- [2] Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., Chen, L., "Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration," *Journal of Information Technology Case and Application Research*, vol. 25, no. 3, pp. 277-304, 2023. DOI: <https://doi.org/10.1080/15228053.2023.2233814>
- [3] Oh, C., Kim, C., Park, K., "Building robust Korean speech recognition model by fine-tuning large pretrained model," *Phonetics Speech Sci.*, vol. 15, no. 3, pp. 75-82, 2023. DOI: <https://doi.org/10.13064/KSSS.2023.15.3.075>
- [4] Vasarhelyi, M. A., Moffitt, K. C., Stewart, T., Sunderland, D., "Large Language Models: An Emerging Technology in Accounting," *Journal of Emerging Technologies in Accounting*, vol. 20, no. 2, pp. 1-10, Oct. 2023.
- [5] Ahn, H.-J., and Ha, Y.-M., "Analysis on the Relationship between the Types of Experience and the Characteristics of Blog Posts," *Journal of the Korea Institute of Information Technology*, vol. 15, no. 2, pp. 131-140, 2017. DOI: <http://dx.doi.org/10.14801/jkiit.2017.15.2.131>
- [6] Lee, W.-G., and Kim, Y.-W., "A Study on the Application of Generative AI in the Tourism Industry," *Korea and International Society*, vol. 8, no. 2, pp. 835-858, 2024. DOI: <https://doi.org/10.22718/kg.2024.8.2.032>
- [7] No, Y., and Byun, J.-W., "A Study on the Acceptance of Tourism Blogs and User Satisfaction," *Journal of Digital Industry and Information Society*, vol. 7, no. 2, pp. 185-203, 2011.
- [8] Kim, B.-K., Byun, J.-Y., and Cha, K.-A., "Sponsored Blog Text Generation through Fine-Tuning Korean Language Models," *Journal of the Korea Industrial Information Systems Society*, vol. 29, no. 3, pp. 1-12, 2024.
- [9] Park, S.-J., and Park, J.-Y., "Long-Form Essay Generation Using a Summary-Based Memory System," *Proceedings of the 2023 KIISE Korea Software Congress*, pp. 1571-1573, Busan, Republic of Korea, Dec. 2023.
- [10] Wang, P., Bai, S., Tan, S., Wang, S., Fan, Z., Bai, J., Chen, K., Liu, X., Wang, J., Ge, W., et al., "Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution," *arXiv preprint arXiv:2409.12191*, 2024.
- [11] Kim, B., Kim, H., Lee, S.-W., Lee, G., Kwak, D., Jeon, D. H., Park, S., Kim, S., Kim, S., Seo, D., et al., "What changes can large-scale language models bring? Intensive study on HyperCLOVA: Billions-scale Korean generative pretrained transformers," *arXiv preprint arXiv:2109.04650*, 2021.
- [12] Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I., "Zero-Shot Text-to-Image Generation," *arXiv preprint arXiv:2102.12092*, Feb. 2021. [Online]. Available: <https://arxiv.org/abs/2102.12092>
- [13] Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., Artzi, Y., "BERTScore: Evaluating text generation with BERT," *arXiv preprint arXiv:1904.09675*, 2019.
- [14] Chen, X., et al., "PaLI: A Jointly-Scaled Multilingual Language-Image Model," *arXiv preprint arXiv:2303.10463*, 2023.
- [15] Bordes, F., et al., "An Introduction to Vision-Language Modeling," *arXiv preprint arXiv:2405.17247*, 2024.
- [16] Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V., "Unsupervised cross-lingual representation learning at scale," *arXiv preprint arXiv:1911.02116*, Nov. 2019. [Online]. Available: <https://arxiv.org/abs/1911.02116>
- [17] Dhuliawala, S., et al., "Chain-of-Verification Reduces Hallucination in Large Language Models," *arXiv preprint arXiv:2309.11495*, 2023.
- [18] Amershi, S., Weld, D., Vorvoreanu, M., Founrey, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., "Guidelines for human-AI interaction," *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI)*, pp. 1-13, Glasgow, Scotland, May 2019. DOI: <https://doi.org/10.1145/3290605.3300233>
- [19] Patke, A., et al., "Queue Management for SLO-Oriented Large Language Model Serving," *arXiv preprint arXiv:2407.00047v2*, 2024.
- [20] Schiebelbein, E., et al., "Models for Storage in Database Backends," *arXiv preprint arXiv:2403.11716*, 2024.
- [21] Zahra, A. R. A., Jonas, D., Erliyani, I., et al., "Assessing customer satisfaction in AI-powered services: An empirical study with SmartPLS," *Journal on Artificial Intelligence*, vol. 5, pp. 1-15, 2023.

Author



Donghun Lim received his B.S. degree in computer engineering from Keimyung University (KMU), Daegu, Korea, in 2022, and is currently pursuing his M.S. degree in data science at the Graduate School of Data

Science, Kyungpook National University (KNU), Korea. He is interested in computer vision problems, including image and video restoration, video frame interpolation, and vision-based anomaly detection.



Seungsoo Han received his B.S. degree in Big Data Analysis Business Statistics from Kookmin University(KMU), Seoul, Korea, in 2022. He is interested in defense AI industry, particularly in applications

involving multi-modal, and Computer Vision.



Euichan Eun received his B.S. degree in computer engineering from Hanshin University, Korea, and is currently pursuing his M.S. degree in artificial intelligence systems engineering at Hanyang University,

Korea. He is interested in building services using RAG (Retrieval-Augmented Generation) and developing agents.



Dongyeong Kim is currently pursuing his B.S. degree in Industrial Management Engineering at the University of Ulsan, Ulsan, Korea. He is interested in vision-based industrial anomaly detection, time-series

anomaly detection.



Janghoon Choi received the B.S. degree in electrical and computer engineering, and Ph.D. degree in electrical engineering and computer science from Seoul National University, Korea, in 2013 and 2021,

respectively. He joined the faculty of the Graduate School of Data Science at Kyungpook National University, Daegu, Korea, in Sep. 2022. He is currently an Assistant Professor in the Graduate School of Data Science at Kyungpook National University. He is interested in computer vision problems including visual tracking, video understanding and image restoration.