DOI	https://DOI.org/10.21481/jbunka..107.202511.303.

페르소나 기반 로컬 LLM 챗봇과 AI 프롬프트를 활용한 일본어 교육 프레임워크*

—비판적 AI 리터러시와 학습자 주도성 함양을 중심으로—

김 유 영**
(e-mail : yuiyu@hotmail.com)

<목 차>

1. 들어가기
2. 이론적 배경 및 선행 연구 검토
3. 일본어 교육을 위한 AI 리터러시 교육 프레임워크의 설계
3.1. 이론적 기반
3.2. 핵심 원칙
3.3. 구성 및 요소
3.4. 과정 및 단계
4. 프레임워크 적용 사례
4.1. 연구대상 및 환경
4.2. 연구절차: 프레임워크의 단계적 구현
4.3. 자료 수집 및 분석 방법
5. 프레임워크 적용 결과 및 분석
5.1. 학습자 사전 인식 및 역량 분석
5.2. 프레임워크 적용을 통한 역량 및 인식 변화 분석
5.3. 교육적 효과 종합 분석
6. 나가기

キーワード：日本語教育(Japanese Language Education), 生成AI(Generative AI), 批判的AIリタラシー(Critical AI Literacy), 学習者の自律性(Learner Autonomy), ペルソナ型チャットボット(Persona-based Chatbot)

1. 들어가기

생성형 인공지능(Generative AI)의 등장은 교육 현장에 전례 없는 도전과 기회를 동시에 제시하며, 특히 인간의 언어와 문화를 깊이 탐구하는 외국어 교육

* 이 논문은 2024년도 동덕여자대학교 연구비 지원에 의하여 수행된 것임(연구번호: 202406435). This study was supported by the Dongduk Women's University grant(No. 202406435)
** 동덕여자대학교 일어일본학과 부교수, 일본어학, AI, 컴퓨터 일본어학

분야에 새로운 교육 모델의 정립을 요구하고 있다. ChatGPT를 위시한 거대언어모델(LLM)은 작문, 번역, 대화 연습 등에서 과거에는 상상하기 어려웠던 수준의 잠재력을 보여주지만, 그 이면에는 신중한 교육적 접근을 요구하는 심각한 문제들이 잠재해 있다. 학습자들이 AI가 생성한 그럴듯한 결과물을 무비판적으로 수용함으로써 비판적 사고 능력이 약화되고, 지식의 주체적인 탐구자가 아닌 수동적인 소비자로 전락할 수 있다는 우려는 학계 전반에서 공통적으로 제기되고 있다.

그런데 지금까지 일본어 교육 분야의 AI 활용 논의는 주로 특정 도구의 단편적인 사용법을 소개하거나, 기술적 한계에 대한 추상적 경고에 머무르는 경향이 있었다. AI의 구조적 결함을 교육 현장에서 적극적으로 활용하여 학습자의 비판적 사고 능력을 함양하고, 나아가 학습자가 직접 AI를 설계하고 통제하는 주체로 거듭나도록 이끄는 체계적인 교육 프레임워크는 부재한 실정이다.

이에 본 연구는 이러한 문제의식을 바탕으로, "페르소나 기반 로컬 LLM 챗봇과 AI 프롬프트를 활용한 일본어 교육 프레임워크"를 설계하고 그 구체적인 실행 방안을 제안하고자 한다.

2. 이론적 배경 및 선행 연구 검토

생성형 AI의 등장은 일본어 교육 분야에서도 즉각적인 반향을 일으켜, 그 교육적 활용 가능성을 모색하는 초기 연구들이 활발히 등장했다. 선행 연구들은 주로 AI를 영상 자막 번역, 비즈니스 작문, 피드백 도구 등으로 활용한 수업 사례를 통해 학습 동기 유발과 실용적 과제 수행 능력 향상이라는 긍정적 효과를 보고했다(배소연(2025), 이현정(2025), 조선영(2024), 吹上(2025) 등). 이러한 연구들은 AI가 유용한 학습 보조 도구가 될 수 있다는 중요한 가능성을 확인시켜 주었다는 점에서 의의가 있다.

그러나 현재까지의 연구들은 AI를 '무엇을 할 수 있는가'라는 기능적 측면에서 접근하여, 특정 과제에 '단순히 사용해보는' 단편적 활용에 머무르는 명백한 한계를 공유한다. 다수의 일본어 학습자들이 AI를 번역이나 단순 정보 검색 용도로만 사용할 뿐, 심화 학습에는 활용하지 못하고 있다는 조사 결과

(三国·諏訪:2024)는 이러한 경향을 뒷받침한다. 이는 AI의 구동 원리나 기술적 본질에 대한 깊이 있는 이해 없이 AI를 일종의 '블랙박스'로 취급하는 접근이 주를 이루고 있음을 보여준다.

이러한 경향은 '비판적 AI 리터러시' 교육의 부재라는 더 큰 문제로 이어진다. AI 리터러시의 필요성은 여러 분야에서 강조되고 있으나(김한성 외(2021), 박윤수·이유미(2021)), 언어와 텍스트를 민감하게 다루어야 할 일본어 전공자의 특수성을 고려한 교육 모델은 거의 찾아볼 수 없다. 실제로 일본어 전공 학생들은 번역기에 대한 의존도는 매우 높은 반면, ChatGPT와 같은 생성형 AI의 적극적 활용에는 소극적이며, 심지어 학부 차원의 AI 교육이 불필요하다고 인식하는 비율이 절반에 달한다는 충격적인 결과(이해미:2024)는 학습자들이 AI를 비판적으로 이해하고 주체적으로 통제하는 방법을 배우지 못한 채 그저 편리한 도구로만 인식하고 있음을 시사한다. 교수자들 역시 AI 활용 교육의 필요성에는 공감하지만 구체적인 방법을 몰라 어려움을 겪고 있다는 보고(현선령:2025)는 이러한 교육적 공백을 더욱 명확히 보여준다.

'비판적 AI 리터러시' 교육이 시급한 이유는 생성형 AI가 '환각(Hallucination)'과 '편향(Bias)'이라는 심각한 구조적 결함을 본질적으로 내포하고 있기 때문이다. 환각은 AI가 사실이 아닌 정보를 그럴듯하게 생성하는 현상으로, 이는 단순한 오류가 아닌 LLM의 통계적 예측 방식에 내재된 속성이다. 편향은 AI가 학습 데이터에 존재하는 인간 사회의 편견과 고정관념을 그대로 학습하고 증폭시키는 현상이다. 특히 AI는 일본어의 복잡한 경어(敬語) 체계나 특정 단어에 담긴 고맥락적 문화 뉘앙스를 제대로 처리하지 못하며, 젠더나 문화적 스테레오타입을 무비판적으로 재생산하는 문제를 드러낸다.

본고에서는 이와 같은 선행 연구 검토를 통해 다음과 같은 점을 확인할 수 있었다. 첫째, 기존 연구들은 AI의 활용 사례를 단편적으로 제시하거나 그 한계를 추상적으로 지적하는 데 그치고 있으며, AI의 구조적 결함을 오히려 학습자의 비판적 사고 능력을 함양하는 교육적 자원으로 적극 활용하려는 시도는 찾아보기 어렵다. 둘째, 일본어 전공자의 인문학적 소양과 요구에 부응하는 체계적이고 통합적인 교육 프레임워크가 부재하다.

따라서 본 연구는 이러한 점을 고려하여 "페르소나 기반 로컬 LLM 챗봇과 AI 프롬프트를 활용한 일본어 교육 프레임워크"를 제안하고자 한다. 본 프레임워크는 AI의 한계를 지적하는 것을 넘어, 이를 '해체 및 비판'의 교육 내용

으로 삼고, 프롬프트 엔지니어링과 로컬 LLM 환경 구축을 통해 AI를 의도대로 '적용 및 통제'하며, 최종적으로 '나만의 AI 튜터'를 '창조 및 성찰'하는 체계적인 3단계의 과정을 통해 기존의 단편적인 AI 활용 논의를 넘어서는 통합적이고 실천적인 교육 모델을 제시한다는 점에서 독창적인 의의를 가진다.

3. 일본어 교육을 위한 AI 리터러시 교육 프레임워크의 설계

본 연구에서 제안하는 "페르소나 기반 로컬 LLM 챗봇과 AI 프롬프트를 활용한 일본어 교육 프레임워크"는 생성형 AI 시대에 요구되는 '비판적 AI 리터러시'와 '학습자 주도성' 함양을 목표로 설계된 통합적 교육 모델이다. 본 프레임워크는 D대학교 일어일본학과 전공강의 '생성형AI와미디어커뮤니케이션일본어'[1] 강의에 실제 적용되었으며, 그 구체적인 실행 모델은 아래 <표 1>과 같다.

<표 1> AI 리터러시 교육 프레임워크 적용 강의계획서

주차	학습 목표	주요 학습 내용
1	AI 시대의 인문학도	- AI, ML, DL 기본 개념 - 인문학도에게 AI 리터러시의 중요성
2	AI를 움직이는 힘, 데이터	- 빅데이터와 LLM의 관계 - AI의 일본어 학습 방식 이해
3	비판적 AI 활용론	- AI의 환각(Hallucination)과 편향성 문제 - '사고 확장 도구'로서의 AI
4	생성형 AI 툴킷 (1)	- 주요 클라우드 AI 서비스 비교 분석 (ChatGPT, Gemini 등)
5	생성형 AI 툴킷 (2)	- 로컬 AI(Ollama[2]) 설치 및 운영 실습
6	AI 번역 마스터 (1)	- 효과적인 번역 프롬프트 엔지니어링
7	AI 통역 파트너	- AI 번역/통역의 한계와 인간의 역할 재정의
8	중간고사	- 이론 시험
9	AI와 함께하는 일본어 어휘 정복	- AI를 활용한 뉘앙스 및 관용구 학습
10	나만의 AI 튜터 만들기 (1): 기획	- 페르소나 챗봇 프로젝트 기획 및 인물 분석
11	나만의 AI 튜터 만들기 (2): 구현	- 페르소나 프롬프트 작성 및 1차 테스트

1) 상세 강의계획서 다운로드 : http://www.japanese.or.kr/ai_for_jpn_edu/ai_jpn_edu.aspx

12	나만의 AI 튜터 만들기 (3): 고도화	- 페르소나 붕괴 현상 분석 및 프롬프트 수정/보완
13	최종 프로젝트 발표	- 제작한 AI 챗봇 시연 및 제작 과정 발표
14	총정리 및 미래 전망	- AI 시대의 학습 로드맵 설계
15	기말고사	- 페르소나 챗봇 제출로 시험 대체

다음 절에서는 상기 강의계획서의 설계 기저를 이루는 본 프레임워크의 1) 이론적 기반, 2) 핵심 원칙, 3) 구성 요소, 4) 과정 및 단계라는 4대 핵심 축에 관하여 상세히 논하고자 한다. 각 구성요소는 상호 유기적으로 작용하여 학습자가 AI 기술의 수동적 소비자에서 능동적 창조자로 성장하도록 지원한다.

3.1. 이론적 기반

본 프레임워크는 세 가지 핵심 학습 이론에 바탕을 두고 있는데, 그 첫 번째가 구성주의 학습 이론(Constructivism)이다. 본 프레임워크는 학습자가 지식을 수동적으로 수용하는 존재가 아니라, 능동적인 상호작용과 경험을 통해 자신만의 지식을 구성해나가는 주체임을 전제한다. 이는 강의의 최종 프로젝트인 '나만의 AI 튜터 만들기'(10-13주차)에서 가장 명확하게 구현된다. 학습자는 단순히 주어진 AI 챗봇을 사용하는 것이 아니라, 직접 특정 인물의 페르소나를 분석하고, Modelfile을 설계하며, 로컬 LLM 챗봇을 '구축'하는 전 과정에 참여한다. 이 과정에서 학습자는 AI의 작동 원리와 한계를 시행착오를 통해 내재화하며, 자신만의 의미를 구성하는 능동적인 지식의 창조자로 거듭난다.

둘째, 비판적 교육학(Critical Pedagogy)이다. 본 프레임워크는 AI를 가치중립적인 도구로 제시하는 것을 경계한다. 3주차 '비판적 AI 활용론'에서는 AI의 환각(Hallucination)과 편향성 문제를 집중적으로 다루며, 기술에 내재된 사회적, 윤리적 문제를 비판적으로 성찰하도록 유도한다. 이는 학습자가 기술의 맹목적인 수용자가 아닌, 기술이 사회에 미치는 영향을 질문하고 더 나은 방향을 제안하는 비판적 주체로 성장해야 한다는 교육 철학을 반영한다. 배소연(2025)에서 학습자들이 AI 도구의 한계를 보완하는 과정에서 더 심도 있는 학습이 이루어졌다는 결과는 이러한 비판적 접근의 교육적 효과를 뒷받침한다.

셋째, 체험 기반 학습(Experiential Learning)이다. 본 프레임워크는 추상적인

2) Ollama는 Phi-3, Llama 3, Mistral 등 다양한 오픈소스 거대언어모델(LLM)을 사용자의 개인 컴퓨터(PC) 환경에서 직접 설치하고 구동할 수 있도록 지원하는 프로그램이다. https://ollama.com

이론 학습을 넘어, 구체적인 '체험'과 '실행'을 통한 학습을 강조한다. 5주차 'Ollama 설치 실습'에서 겪는 문제 해결 과정, 6주차 '번역 비교 실습'에서 프롬프트의 미세한 차이가 번역의 질을 극적으로 바꾸는 경험, 그리고 10-13주차 '페르소나 챗봇 제작'의 전 과정은 학습자가 직접 부딪히고 해결하는 체험의 연속이다. 이러한 과정을 통해 학습자는 추상적인 개념을 실제적인 문제 해결 역량으로 전환하게 된다.

3.2. 핵심 원칙

본 프레임워크는 다음 세 가지 핵심 원칙을 기반으로 삼고 있다.

(1) 본 연구의 프레임워크 핵심 3원칙

a. 원칙 1.비판적 수용(Critical Adoption) - AI는 정답이 아닌 탐구 대상

: 본 프레임워크는 AI의 배척이나 맹신이 아닌, 그 잠재력과 명백한 한계를 동시에 인지하고 지혜롭게 활용하는 태도를 지향한다. 이는 3주차 '비판적 AI 활용론'에서 AI의 구조적 결함을 학습하고, 7주차 'AI 통역 파트너'에서 경어, 문화적 뉘앙스 등 AI가 놓치는 부분을 명확히 분석하는 활동을 통해 구체화된다. 학습자는 AI의 결과물을 최종적인 '정답'으로 받아들이는 것이 아니라, 비판적 검증을 거쳐야 할 '가설' 또는 '초안'으로 간주하는 훈련을 받는다.

b. 원칙 2. 학습자 주도성(Learner Autonomy) - 학습자는 AI의 수동적 소비자가 아닌, 능동적 설계자이자 창조자

: 본 프레임워크의 궁극적인 목표는 학습자가 스스로의 학습 목표에 맞춰 AI를 '설계'하고 '통제'하는 힘을 기르는 것이다. 이는 10-13주차 '나만의 AI 튜터 만들기' 프로젝트에서 정점에 달하며, 학습자는 자신의 필요에 맞는 맞춤형 학습 도구를 직접 창조하는 경험을 한다. 나아가 14주차 '자신의 학습 로드맵 설계'를 통해, 강의 종료 후에도 지속 가능한 자기 주도적 학습자로서의 자세를 확립하게 된다.

c. 원칙 3. 인문학과 기술의 융합(Integration of Humanities and Technology) - 기술과 인문학적 소양은 상호보완적

: 본 프레임워크는 AI 기술 교육이 인문학적 소양과 분리될 수 없음을 강조한다. 1주차에서 '인문학도에게 AI 리터러시가 왜 중요한가'를 역설하는 것부터 시작하여 6주차에서 문학 작품의 서정성을 번역하기 위해

프롬프트를 정교하게 다듬는 활동, 그리고 10주차에서 특정 인물의 페르소나를 깊이 있게 분석하여 AI에 이식하는 과정 전체가 이 원칙을 구현한다. 기술은 인문학적 통찰을 표현하는 도구가 되고, 인문학적 깊이는 기술의 한계를 극복하는 열쇠가 된다는 것을 학습자가 체득하도록 한다.

3.3. 구성 요소

본 프레임워크는 교육 목표 달성을 위해 다음과 같은 구체적인 지식, 기술, 도구, 활동으로 구성된다.

(2) 본 연구의 프레임워크 구성요소

a. **지식 요소(Knowledge):** AI, 머신러닝, 딥러닝의 기본 개념(1주차), 빅데이터와 LLM의 학습 원리(2주차), AI의 윤리적 문제(환각, 편향)(3주차), 기계 번역의 역사와 한계(6주차) 등 AI를 비판적으로 이해하기 위한 이론적 토대를 제공한다.

b. **기술 요소(Skills):** 클라우드/로컬 AI 비교 분석(4-5주차), Ollama 환경 구축(5주차), 프롬프트 엔지니어링(6주차), RAG를 활용한 페르소나 설계(5주차, 11주차), AI 번역 결과물 비판 및 수정(Post-editing)(7주차) 등 AI를 주체적으로 통제하기 위한 실질적인 기술을 포함한다.

c. **도구 요소(Tools):** 클라우드 AI(ChatGPT, Gemini 등), 로컬 AI 실행 환경(Ollama), 다양한 LLM(Phi-3, Llama3, Mistral 등), Modelfile, 텍스트 에디터 등 구체적인 실습 도구를 활용한다.

d. **핵심 활동(Activities):** AI 번역 비교 실습(6주차), AI 실시간 통역 시뮬레이션(7주차), AI 활용 어휘/관용구 뉘앙스 분석(9주차), 그리고 이 모든 것을 종합하는 '페르소나 챗봇 제작 프로젝트'(10-13주차)가 핵심 활동으로 기능한다.

3.4. 과정 및 단계

본 프레임워크는 15주간의 강의를 통해 점진적이고 체계적인 학습이 이루어지도록 설계된 3단계의 나선형 구조를 가진다. 이는 학습자가 단순한 지식 습득에서 응용, 그리고 창조로 심화되는 과정을 통해 고차원적인 역량을 함양하도록 이끈다.

(3) 본 연구의 프레임워크 과정 및 단계

a. 1단계 해체 및 비판 (Deconstruction & Critique / 1~3주차) - 'AI란 무엇인가'

: 이 단계에서 학습자는 AI의 기술적 원리를 인문학도의 눈높이에서 이해하고(1-2주차), 나아가 AI가 가진 근본적인 한계와 위험성(환각, 편향)을 인문학적 비판 도구를 사용하여 날카롭게 '해체'한다(3주차). 이는 AI 기술에 대한 막연한 환상을 걷어내고, 이후의 모든 학습 활동의 기반이 될 비판적 토대를 마련하는 과정이다.

b. 2단계 적용 및 통제 (Application & Control / 4~9주차) - 'AI를 어떻게 다룰 것인가'

: 1단계에서 확립한 비판적 시각을 바탕으로, 학습자는 다양한 AI 툴킷을 직접 다루고(4-5주차), 프롬프트 엔지니어링이라는 구체적인 '통제' 기술을 연마한다(6주차). 이를 번역, 통역, 어휘 학습 등 일본어 교육의 구체적인 영역에 '적용'하며 AI를 유용한 도구로 활용하는 실질적인 역량을 기른다(7, 9주차). 이 단계는 추상적인 비판을 넘어, 기술을 의도대로 제어하는 실천적 능력을 배양하는 데 중점을 둔다.

c. 3단계 창조 및 성찰 (Creation & Reflection / 10~14주차) - 'AI로 무엇을 만들 것인가'

: 마지막 단계에서 학습자는 이전 단계에서 습득한 모든 지식과 기술을 총동원하여 '나만의 AI 튜터'라는 새로운 교육적 가치를 '창조'한다(10-13주차). 이 창조의 과정과 최종 발표를 통해, 학습자는 AI 시대의 일본어 학습자이자 인문학도로서 자신의 역할과 미래를 '성찰'하며 학습자 주도성을 완성하게 된다(13-14주차).

이와 같은 3단계의 체계적인 과정을 통해, 본 프레임워크는 학습자가 AI의 수동적인 사용자에서 출발하여, 비판적인 분석가를 거쳐, 최종적으로는 주체적인 창조자로 성장하도록 이끈다.

4. 프레임워크의 적용 사례

본 장에서는 3장에서 설계한 "페르소나 기반 로컬 LLM 챗봇과 AI 프롬프

트를 활용한 일본어 교육 프레임워크"를 실제 대학 교육 현장에 적용한 구체적인 사례를 기술한다. 본 연구는 제안된 프레임워크의 실효성과 교육적 가치를 탐색하기 위해 질적 사례 연구(Qualitative Case Study) 방법을 채택하였다. 이는 통계적 일반화보다는 특정 교육적 맥락 속에서 나타나는 학습 과정의 역동성과 의미를 심층적으로 이해하는 데 목적이 있다.

4.1. 연구 대상 및 환경

본 프레임워크는 D대학교 일어일본학과 2학년 전공과목 '생성형AI와미디어커뮤니케이션일본어'에 적용되었다. 총 15주간 진행된 본 강의에는 일본어 전공 및 복수전공 학생 20명[3]이 참여하였다. 수강생들은 대부분 ChatGPT나 Papago 번역기 등 클라우드 기반 AI 도구에 대한 사용 경험은 있었으나, AI의 작동 원리를 이해하거나 직접 AI 모델을 설치 및 제어해 본 경험은 전무한 상태였다. 수업은 학생들이 개인 노트북을 지참하여 참여하는 실습 위주의 환경에서 진행되었으며, 이는 저사양 PC 환경에서도 로컬 AI를 구동하는 현실적인 경험을 제공하기 위함이었다.

그림 1 사전 설문 조사 결과 중, 생성형 AI의 사용 경험에 대한 응답

3) 수강 학생은 전원 여학생으로, 2학년생 11명, 3학년생 4명, 4학년생 4명, 5학년생 1명으로 구성.

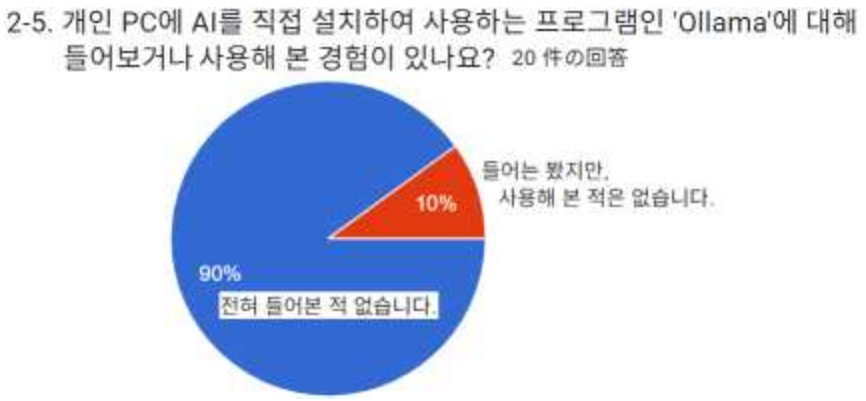

그림 2 Ollama를 통한 로컬 AI(LLM) 사용 경험에 대한 응답

4.2. 연구 절차: 프레임워크의 단계적 구현

본 연구는 3장의 3.4절에서 제시한 '1단계: 해체 및 비판', '2단계: 적용 및 통제', '3단계: 창조 및 성찰' 이라는 3단계의 과정에 따라 15주간의 수업을 체계적으로 진행하였다. 본 절에서는 프레임워크의 핵심적인 기술 요소가 집약된 2단계의 주요 활동, 특히 5주차 '생성형 AI 툴킷 (2): 로컬 환경에서의 AI 설치 및 운영 연습' 의 실행 과정을 중심으로 상세히 서술한다. 이 활동은 학습자가 AI의 수동적 소비자에서 능동적 설계자로 전환되는 중요한 변곡점이기 때문이다

4.2.1 로컬 LLM 환경 구축 (Ollama 설치 및 모델 다운로드)

5주차 수업의 목표는 학습자가 자신의 PC에 직접 AI를 구동할 수 있는 독립적인 환경을 구축하는 것이다. 이는 클라우드 AI의 데이터 프라이버시 문제에서 벗어나고, 인터넷 연결 없이도 AI를 활용하며, 무엇보다 AI 모델을 자신의 의도대로 커스터마이징할 수 있는 기술적 기반을 마련하기 위함이다.

이를 위해 본 연구에서는 로컬 LLM 구동 프로그램인 Ollama를 활용하였다. 학습자들은 강의 자료의 상세한 가이드에 따라 각자의 운영체제(Windows/macOS)에 맞는 ollama 프로그램을 설치하였다. 설치 후, 터미널(명령 프롬프트)을 통해 ollama --version 명령어로 설치를 확인하고, ollama pull 명령어를 사용하여 실습에 사용할 LLM들을 다운로드했다. 본 강의에서는 저사양 PC 환경을 고려하여, 가볍고 빠른 Phi-3 모델을 기본 원리 학습용으로,

더 나은 성능과 창의적인 역할 수행 능력을 가진 Mistral 모델을 심화 실습용으로 각각 다운로드하여 모델별 성능 차이를 비교 체험할 수 있도록 지도했다.

4.2.2 페르소나 챗봇 제작 실습(Modelfile과 RAG의 적용)

환경 구축 이후, 본 프레임워크의 핵심인 '페르소나 기반 로컬 LLM 챗봇' 제작 실습을 진행하였다. 이 실습은 AI에게 특정 인물의 정체성과 지식을 부여하는 RAG(검색 증강 생성, Retrieval-Augmented Generation) 기술의 원리를 학습자가 직접 체험하도록 설계되었다. 본 강의에서는 사전 설문조사에서 선호도가 높았던 일본 애니메이션 '장송의 프리렌(葬送のフリーレン)'의 등장인물인 '용사 힘멜(Himmel/ヒンメル)'을 페르소나로 설정한 챗봇 제작을 시범 과제로 제시하였으며, 그 실습 과정은 다음과 같은 구체적인 단계로 진행되었다.

(4) 챗봇 제작 실습 과정(e.g. 페르소나: 장송의 프리렌의 용사 힘멜)[4)]

a. **참고 자료(Knowledge Base) 생성:** 학습자들은 '용사 힘멜'의 성격, 가치관, 말투, 주요 일화 등이 구조적으로 정리된 '캐릭터 시트' 형식의 텍스트를 제공받아 이를 himmel_data.txt 파일로 저장하였다. 이는 RAG에서 AI가 답변의 근거로 '검색(Retrieval)'할 지식의 원천이 된다.

b. **Modelfile 설계:** 이어 학습자들은 AI의 '설계도' 역할을 하는 Modelfile을 작성했다. 본 강의에서 제공된 Modelfile은 단순한 지시문의 나열을 넘어 AI의 기본 자아(Default Ego)를 억제하고 페르소나를 강력하게 주입하기 위한 다음과 같은 고도의 프롬프트 엔지니어링 기법을 포함한다.

① FROM: Phi-3, llama3, mistral과 같이 페르소나 챗봇의 기반이 될 기본 LLM을 지정한다.

② TEMPLATE: AI 모델이 훈련된 고유의 '대화 형식'을 직접 지정하는 부분이다. <|start_header_id|>system<|end_header_id|>와 같은 특수 태그를 사용하여 페르소나 지시문이 단순한 '요청'이 아닌 모델이 무시할 수 없는 '시스템 레벨의 명령'으로 처리되도록 강제한다. 이는 AI의 강력한 안전 정렬(Safety Alignment)을 우회하고 의도한 페르소나를 일관되게 유지하는 데 결정적인 역할을 한다.

③ PARAMETER: temperature 값을 조절하여 답변의 창의성과 일관성

4) 실습에 사용된 모든 파일 다운로드 : http://www.japanese.or.kr/ai_for_jpn_edu/ai_jpn_edu.aspx

사이의 균형을 맞춘다. 높으면 높을수록 더 창의적 답변을 내놓는다.

④ SYSTEM: AI의 정체성과 행동 지침을 정의하는 핵심적인 부분. 여기에는 AI의 '절대 규칙'과 '역할 부여', 그리고 RAG의 핵심인 '지식 활용 규칙'(e.g. "너의 모든 기억과 지식은 아래 제공되는 참고 자료에만 근거 한다")이 명시된다. 마지막으로, himmel_data.txt의 내용 전체가 이 SYSTEM 프롬프트 내에 포함된다.

b. **모델 생성 및 테스트:** 학습자들은 터미널에서 ollama create [모델명] -f Modelfile 명령어를 실행하여, 자신이 설계한 Modelfile과 참고 자료가 결합된 새로운 '페르소나 모델'을 생성하였다. 이후 ollama run [모델명] 명령어로 챗봇을 실행하고, "네 이름은 뭐니?", "너는 AI야?" 등 페르소나의 일관성을 검증하는 다양한 질문을 통해 AI와 상호작용을 수행했다. 이 과정에서 학습자들은 AI가 지시를 따르지 않고 기본 자아로 돌아가는 '페르소나 붕괴' 현상을 직접 경험하고, 이를 해결하기 위해 SYSTEM 프롬프트의 지시문을 더 강력하고 명확하게 수정하는 반복적인 문제 해결 과정을 거쳤다. 이는 프롬프트 엔지니어링의 중요성을 체감하는 중요한 학습 경험이 되었다.

4.3. 자료 수집 및 분석 방법

본 프레임워크의 교육적 효과, 즉 '비판적 AI 리터러시'와 '학습자 주도성'의 함양 정도를 심층적으로 측정하기 위해, 본 연구에서는 전통적인 필기시험을 배제하고 다음과 같은 질적 자료를 수집하여 분석하였다.

(5) 자료 수집 및 분석

a. **학습 과정 및 최종 프로젝트 산출물 (Learning Process and Final Project Deliverables):** 본 강의의 최종 평가는 '나만의 AI 튜터 만들기' 라는 프로젝트 기반 과제로 수행되었다. 이는 학습자가 한 학기 동안 배운 모든 지식과 기술을 총체적으로 활용하여, 자신만의 학습 목표에 부합하는 결과물을 창조하는 과정 전체를 평가하기 위함이다. 학습자들은 자신이 관심 있는 일본 문화 콘텐츠(애니메이션, 소설, 드라마 등)에 등장하는 대중적 인지도를 가진 캐릭터 1인을 선정하여, 해당 캐릭터의 페르소나를 가진 로컬 LLM 챗봇을 제작하고 그 과정과 결과를 제출하였다. 그 결과 수집된 자료는 다음과 같다.

① 페르소나 분석 자료: RAG의 지식 기반(Knowledge Base)으로 사용된 캐릭터의 말투, 성격, 가치관, 주요 일화 등을 분석 및 정리한 텍스트 파일(character_data.txt).

② Modelfile 개발 과정: 페르소나를 구현하기 위해 반복적으로 수정 및 개선한 Modelfile의 버전별 기록.

③ 최종 결과 보고서: 제작 의도, 페르소나 분석 과정, 프롬프트 설계 전략, 제작 과정에서 겪은 기술적/논리적 문제(예: 페르소나 붕괴)와 그 해결 과정, 그리고 완성된 AI 튜터의 시연 결과 및 향후 학습 활용 계획을 상세히 기술한 보고서.

상기 산출물들은 학습자의 기술적 이해도(e.g. Ollama, RAG), 프롬프트 엔지니어링 능력, 그리고 인문학적 분석 능력(캐릭터 해석)의 성장 과정을 보여주는 객관적인 자료가 된다.

b. **성찰 보고서 (Reflection Reports):** 학기 말, 학습자들은 최종 프로젝트 과제와는 별도로, 한 학기 동안의 학습 경험 전반에 대한 성찰 보고서를 제출하였다. 이 보고서에는 AI에 대한 초기 인식과 현재의 인식 변화, 프레임워크의 각 단계(해체-적용-창조)별 학습 경험에서 느낀 점, 그리고 AI 시대의 일본어 전공자로서 자신의 미래 역할과 학습 방향에 대한 생각이 자유롭게 기술되어 있다. 이는 프로젝트 산출물만으로는 파악하기 어려운 학습자의 내면적, 정의적(affective) 영역의 변화를 포착하는 중요한 질적 자료로 활용된다.

이와 같이 수집된 자료는 질적 내용 분석(Qualitative Content Analysis) 방법을 통해 분석되었으며, 최종 프로젝트 산출물, 특히 Modelfile의 내용은 프롬프트의 구체성, 제약 조건의 명확성, 페르소나 반영의 정교성 등을 기준으로 분석하여 학습자의 프롬프트 엔지니어링 능력의 발달 수준을 평가했다. 또한, 성찰 보고서는 '비판적 AI 리터러시'와 '학습자 주도성'과 관련된 핵심 키워드(e.g. '한계 인식', '비판적', '스스로', '통제', '설계')의 출현 빈도와 맥락을 중심으로 주제별 코딩(Thematic Coding)을 진행했다. 이 두 가지 자료의 교차 분석을 통해, 본 교육 프레임워크가 학습자의 구체적인 역량과 내면적 인식에 미친 영향을 입체적으로 해석했다.

5. 프레임워크 적용 결과 및 분석

본 장에서는 제안된 교육 프레임워크를 15주간 적용한 후, 학습자들의 역량과 인식에 어떠한 변화가 나타났는지 분석한다. 분석은 강의 전 실시한 사전 설문조사 결과와 학기 말에 수집된 최종 프로젝트 산출물 및 성찰 보고서를 교차 분석하는 방식으로 이루어졌다. 이를 통해 본 프레임워크가 목표로 한 '비판적 AI 리터러시'와 '학습자 주도성' 함양에 미친 교육적 효과를 구체적으로 제시하고자 한다.

5.1. 학습자 사전 인식 및 역량 분석: 낮은 기술 자신감과 피상적 AI 활용

본 프레임워크의 교육적 효과를 측정하기에 앞서, 총 20명의 수강생을 대상으로 실시한 사전 설문조사를 통해 학습자들의 디지털 및 AI 리터러시 수준을 분석하였다. 그 결과는 본 교육의 출발점을 명확히 보여준다.

첫째, 학습자들은 일본어 전공자로서 일반적인 PC 활용 능력(평균 3.35/5.0)은 갖추고 있었으나, AI 기술의 근간이 되는 기술적 리터러시는 매우 낮은 수준을 보였다. 전체 응답자 중 65%(13명)가 프로그래밍(코딩) 경험이 전무했으며, 로컬 LLM 환경 구축에 필수적인 '명령 프롬프트'나 '터미널'을 직접 사용해 본 경험이 있는 학생은 단 한 명도 없었다(강의나 설명을 따라 해 본 6명 제외). 이러한 기술적 배경의 부재는 "기술적인 부분에서 잘 따라갈 수 있을지에 대한 걱정", "코딩 및 컴퓨터로 수업을 들어본 적이 없어 걱정 된다" 등 수강 전 학습자들이 공통적으로 표출한 '기술적 불안감'과 직접적으로 연결된다.

둘째, 기술적 리터러시와는 대조적으로, 생성형 AI에 대한 사용 경험은 매우 보편적이었으나 그 활용 방식은 피상적인 수준에 머물러 있었다. 모든 응답자가 생성형 AI 사용 경험이 있었고, 55%(11명)가 주 1회 이상 사용하는 등 일상적인 도구로 자리 잡고 있었다. 그러나 일본어 학습에서의 주된 활용 목적은 '일본어 작문 또는 문법 교정'(12명)과 '간단한 단어나 문장 번역'(11명)에 집중되어 있었다. 이는 학습자들이 AI를 스스로 통제하고 설계하는 주체적인 활용보다는 주어진 기능에 의존하여 정답이나 교정을 얻는 수동적인 소비자의 위치에 머물러 있음을 시사한다. 특히, 본 강의의 핵심 기술인 로컬 LLM에 대

해서는 80%(16명)가 '전혀 들어본 적 없다'고 응답하여, AI 기술에 대한 이해가 클라우드 기반의 상용 서비스에만 국한되어 있음이 명확히 드러났다.

이러한 사전 분석 결과는 본 연구의 대상인 일본어 전공 학습자들이 (1) AI의 기술적 원리에 대한 이해와 자신감이 부족하고, (2) AI를 비판적 성찰 없이 편리한 '정답 생성기'로 활용하는 경향이 있으며, (3) AI를 자신의 목적에 맞게 설계하고 통제하는 '주체적 활용' 경험이 전무하다는 점을 보여준다. 이는 본 프레임워크가 해결하고자 했던 핵심적인 교육 문제와 정확히 일치한다.

5.2. 프레임워크 적용을 통한 역량 및 인식 변화 분석

한 학기 동안 프레임워크를 적용한 후, 학습자들의 최종 프로젝트 산출물과 성찰 보고서를 분석한 결과, '비판적 AI 리터러시'와 '학습자 주도성' 측면에서 유의미한 질적 변화가 관찰되었다.

5.2.1 '비판적 AI 리터러시'의 내재화: 단순 사용자 vs 비판적 설계자

본 프레임워크의 1단계 '해체 및 비판'과 2단계 '적용 및 통제'를 거치며, 학습자들은 AI를 맹목적으로 신뢰하는 태도에서 벗어나 그 한계를 명확히 인지하고, 이를 극복하기 위한 기술을 체득하는 모습을 보였다.

가장 두드러진 변화는 프롬프트 엔지니어링 능력의 고도화에서 나타났다. 사전 설문에서 단순 번역과 문법 교정에 AI를 활용했던 학습자들은 최종 프로젝트의 Modelfile에서 AI의 정체성을 통제하기 위한 매우 정교하고 복합적인 프롬프트를 설계하였다. 다수의 학습자들이 제출한 Modelfile에는 다음과 같은 공통적인 특징이 발견되었다.

(6) 기말 과제용 프로젝트 산출물 중 Modelfile의 특징

a. **강력한 부정 명령:** "너는 AI 어시스턴트가 아니다. 너의 이전 정체성을 모두 잊어라"와 같이, AI의 기본 자아(Default Ego)를 억제하기 위한 명확하고 강압적인 규칙을 SYSTEM 프롬프트의 최상단에 배치하였다.

b. **구체적인 행동 지침:** 단순히 학생들이 각자 설정한 인물의 "페르소나처럼 말해줘"가 아닌, 페르소나의 말투와 성격을 구체적으로 정의하고, RAG의 작동 원리를 이해하여 참고 자료에 없는 내용은 추측하지 말고,

성격에 맞게 자연스럽게 회피하라는 지식 활용 규칙까지 명시하였다.

b. **구조적 제어:** TEMPLATE 지시어를 사용하여 모델의 대화 형식을 직접 제어함으로써 자신의 페르소나 지시가 최우선으로 적용되도록 강제하는 등, AI의 작동 메커니즘에 대한 깊은 이해를 바탕으로 한 통제 능력을 보여주었다.

이러한 변화는 학습자들이 AI를 더 이상 '마법 상자'가 아닌, 명확한 지시와 통제가 필요한 '결함 있는 시스템'으로 인식하게 되었음을 의미한다. 성찰 보고서에서 한 학습자는 "수업 전에는 Gemini가 틀린 답을 하면 그저 'AI가 아직 부족하구나'라고 생각했지만, 이제는 '왜 틀렸을까?', '어떤 프롬프트를 써야 이 문제를 해결할 수 있을까?'를 고민하게 되었다"고 기술했다. 이는 AI의 결과물을 비판적으로 분석하고, 문제 해결을 위해 능동적으로 개입하는 '비판적 AI 리터러시'가 성공적으로 내재화되었음을 보여주는 대표적인 사례이다.

<표 2> C학생의 6주차 프롬프트 엔지니어링 실습 후, 프롬프트의 변화

학습 전	顔から火が出る가 무슨 뜻이야?
학습 후	너는 일본어 관용구와 언어학에 정통한 언어학자야. 顔から火が出る라는 관용구에 대해 1. 문자적 의미와 유래 2. 함축적 의미와 뉘앙스 3. 유사한 한국어 속담에 대해 일본어를 배우고 있는 내가 알기 쉽게 설명해 줄 수 있어?

5.2.2. '학습자 주도성'의 발현: 수동적 소비자 vs 능동적 창조자

본 프레임워크의 3단계 '창조 및 성찰'은 학습자들이 기술에 대한 주도권을 획득하는 과정에서 정점을 이루었다.

사전 설문에서 드러난 '기술적 불안감'에도 불구하고, 모든 학습자는 5주차 실습을 통해 자신의 PC에 Ollama를 성공적으로 설치하고 로컬 LLM을 구동하는 데 성공하였다. 이는 학습자들이 기술적 장벽을 스스로 극복할 수 있다는 자신감을 얻는 중요한 전환점이 되었다. 성찰 보고서에서 다수의 학습자들은 "검은 화면에 직접 명령어를 입력하는 것이 처음에는 두려웠지만, 내 컴퓨터에서 AI가 움직이는 것을 보고 신기했고, 내가 무언가를 통제하고 있다는 느낌이 들었다"고 언급했다.

이러한 자신감은 '나만의 AI 튜터 만들기' 프로젝트에서 '학습자 주도성'으로 발현되었다. 학습자들은 단순히 주어진 과제를 수행하는 것을 넘어, 각자의 관심사와 학습 목표에 따라 창의적인 결과물을 만들어냈다. 예를 들어, 일본 고전 문학 속 인물의 페르소나를 구현한 한 학습자는 해당 인물의 고풍스러운 말투를 재현하기 위해 원전 텍스트를 RAG 자료로 활용하고, PARAMETER temperature 값을 낮춰 예측 불가능한 답변을 최소화하는 등, 자신의 인문학적 지식과 기술적 제어 능력을 융합하는 모습을 보였다.

또한, 성찰 보고서에서는 AI에 대한 역할 인식이 근본적으로 변화했음이 드러났다. 한 학습자는 "이전까지 AI는 내가 모르는 것을 물어보는 '선생님'이었지만, 이 프로젝트를 통해 내가 직접 AI에게 지식과 인격을 가르치는 '개발자'가 된 것 같았다"고 기술했다. 이는 학습자가 AI의 지식을 수동적으로 소비하는 대상에서, 자신의 필요에 맞게 AI를 설계하고 교육하는 능동적 창조자로 역할이 전환되었음을 명확히 보여준다. 생성형 AI 활용을 통한 학습 주체성 향상 효과가 본 프레임워크의 최종 프로젝트를 통해 극대화된 것이다.

5.3. 교육적 효과 종합 분석

본 연구의 결과는 제안된 교육 프레임워크가 일본어 전공 학습자들의 '비판적 AI 리터러시'와 '학습자 주도성'을 함양하는 데 매우 효과적이었음을 시사한다. 학습자들은 '해체-적용-창조' 라는 체계적인 단계를 거치며, AI에 대한 막연한 불안감을 극복하고 기술의 본질을 비판적으로 이해하게 되었다. 또한, 클라우드 AI의 단순 활용을 넘어 로컬 LLM을 직접 구축하고 제어하는 경험을 통해, 기술에 종속된 사용자가 아닌 자신의 학습을 주도하는 설계자로서의 정체성을 확립하였다. 이는 본 프레임워크가 AI 시대의 인문학도가 갖추어야 할 핵심 역량을 성공적으로 길러냈음을 의미하며, 향후 일본어 교육 분야에서 AI를 통합하는 새로운 교육 모델의 가능성을 제시한다.

6. 나가기

본고에서는 생성형 AI 시대에 요구되는 일본어 교육의 새로운 방향성을 제시하기 위해, '비판적 AI 리터러시'와 '학습자 주도성' 함양을 목표로 하는 교

육 프레임워크를 설계하고 그 효과를 탐색하였다. 사전 분석 결과, 학습자들은 기술적 불안감 속에서 AI를 번역이나 작문 교정과 같은 기능에 의존하는 피상적이고 수동적인 소비자로 활용하고 있었다. 이러한 학습자들을 대상으로 본 프레임워크의 '해체-적용-창조' 단계를 적용한 결과, 학습자들은 AI의 한계를 명확히 인지하고, TEMPLATE 지시어와 같은 고도의 프롬프트 엔지니어링을 통해 이를 극복하는 '비판적 AI 리터러시'를 성공적으로 내재화하였다. 또한, 자신의 PC에 로컬 LLM을 직접 구축하고 '나만의 AI 튜터'를 창조하는 경험은 기술에 대한 통제감을 부여하며 '학습자 주도성'을 결정적으로 발현시키는 계기가 되었음을 확인하였다.

본 연구의 핵심적인 교육적 함의는 AI의 환각과 편향 같은 구조적 결함을 회피의 대상이 아닌, 비판적 사고를 함양하는 핵심 교육 자원으로 '역이용'했다는 점에 있다. AI가 생성한 부자연스러운 경어, 문화적 맥락이 결여된 번역을 수정하고, '페르소나 붕괴'를 해결하기 위해 프롬프트를 개선하는 과정 자체가 고차원적인 언어 및 문화 학습 활동으로 기능했다. 특히, 로컬 LLM을 직접 통제하는 경험은 AI를 신비로운 '블랙박스'에서 분석과 설계가 가능한 대상으로 전환시키며, 기술 앞에서 수동적이었던 학습자에게 주체성을 부여하는 결정적 계기가 되었음을 확인했다. 이는 AI 시대의 교육자가 지식 전달자를 넘어, 학습자가 기술의 한계에 부딪히고 이를 극복하는 경험의 장을 설계하는 '학습 경험 설계자'로의 역할 전환이 필요함을 시사한다.

물론 본 연구는 단일 집단 사례 연구로서 일반화의 한계를 지니며, 장기적 효과를 추적하지 못한 점, 빠르게 변화하는 기술을 다룬 점, 그리고 회화체 중심의 콘텐츠를 활용한 점 등은 향후 연구 과제로 남는다. 향후에는 비교 집단을 포함한 양적 연구, 종단적 추적 연구, 그리고 뉴스나 학술 텍스트 등 다양한 장르에 본 프레임워크를 적용하는 후속 연구를 수행하고자 한다.

결론적으로, 생성형 AI 시대에 일본어 교육은 기술의 편리함에 매몰될 것인가, 아니면 기술을 지렛대 삼아 더 높은 차원의 인간 역량을 길러낼 것인가의 기로에 서 있다. 본 연구가 제안한 '해체-적용-창조' 프레임워크는 AI와 경쟁하는 인재가 아닌, AI를 비판적으로 이해하고, 지혜롭게 통제하며, AI가 도달할 수 없는 인간 고유의 문화적, 윤리적 판단을 내리는 '루프 속의 인간(Human-in-the-Loop)'으로서의 전문가를 양성하는 새로운 교육 패러다임을 제시했다는 점에서 그 의의를 찾을 수 있다. 기술이 인간을 압도하는 것이 아

니라, 인간의 깊이 있는 성찰이 기술의 방향을 이끄는 미래를 위해, 교육 현장에서의 이러한 실천적 노력이 계속해서 이어지기를 기대한다.

【참고문헌】

김한성 외(2021)『대학생의 AI 리터러시 함양을 위한 교양교육 프로그램 개발 방안』소프트웨어 정책연구소, pp.1-194.

박윤수 · 이유미(2021)「대학생의 AI 리터러시 역량 신장을위한 교양 교육모델」『정보교육학회논문지』25(2), 한국정보교육학회, pp.423-436. DOI:10.14352/jkaie.2021.25.2.423

배소연(2025)「AI 도구를 활용한 일본어 작문 교육 방안 연구 — 영상 자막 번역 프로젝트 수업을 중심으로 —」『일본어교육연구』 70, 한국일어교육학회, pp.83-98.

이해미(2024)「인공지능 도구의 사용 현황과 인식 조사-일본어 전공 대학생들의 ChatGPT와 번역기 활용 분석」『일본연구』100, 한국외국어대학교 일본연구소, pp.175-194. DOI : 10.15733/jast.2024..100.175

이현정(2025)「일본어 교수 학습의 도구로서 챗GPT의 활용 방법 및 가능성에 관한 고찰-「비즈니스 일본어」 수업을 중심으로 -」『일본어문학』108, 일본어문학회, pp.103-131. DOI : 10.21792/trijpn.2025..108.005

조선영(2024)「ChatGPT를 활용한 일본어 작문 수업 연구.학습자의 성찰 및 인식조사를 중심으로」『日本語教育』第110輯, 한국일본어교육학회, pp.61-72. DOI : 10.26591/jpedu.2024.110..006

현선령(2025)「챗GPT 일본어 수업 활용에 대한 대학교수 인식 연구」『일본어교육연구』70, 한국일어교육학회, pp.167-180.

吹上淳子(2025)「日本語教育における生成AI活用の一考察-日本語作文授業の新しい可能性-」『日語日文学』106, 대한일어일문학회, pp.7-23.

三国喜保子 · 諏訪昭宏(2024)「日本語学習における生成AIチャットボットの使用実態」『일본어교육연구』68, 한국일어교육학회, pp.91-106.

Ollama(https://ollama.com. 검색일: 2025.09.30.)

논문 투고 일자 : 2025. 10. 01.
논문 심사 일자 : 2025. 10. 20.
게재 확정 일자 : 2025. 11. 10.

<要旨>

ペルソナベース・ローカルLLMチャットボットとAIプロンプトを活用した日本語教育フレームワーク
－批判的AIリテラシーと学習者主導性の育成を中心に－

金曙泳

生成AIの急速な普及は外国語教育分野において新たな教育モデルの構築を急務としている。しかし、現状の日本語教育におけるAI活用は表層的な利用に留まり、人文系専攻者に求められる批判的リテラシーや学習者主導性の育成には至っていない。先行研究においてもAI固有の限界を教育的に活用し、学習者を受動的な消費者から能動的な創造者へと転換させる体系的なフレームワークが欠如している。

本研究は、こうした問題意識に基づき、「批判的AIリテラシー」と「学習者主導性」の育成を目的とした新たな日本語教育フレームワークを提案する。本モデルは「脱構築・批判」「適用・制御」「創造・省察」の3段階で構成され、AIの構造的欠陥の分析、Ollamaを用いたローカルLLMの構築と制御、RAG技術を用いたペルソナベース・チャットボットの制作を実践する。

大学の授業における実践の結果、学習者は技術的不安を克服し、受動的な消費者から能動的な設計者へと役割を転換させた。分析の結果、高度なプロンプトエンジニアリング能力とAIへの深く批判的な理解が育成されたことが確認された。本研究は、AIの弱点を教育的強みとして転換する体系的なモデルを提示し、AIを脱構築・制御・創造する力を通じて、人文系学習者に不可欠な批判的思考能力と主体性を育成できることを実証した。

A Framework for Japanese Language Education Using Persona-Based Local LLM Chatbots and AI Prompts: Focusing on the Cultivation of Critical AI Literacy and Learner Autonomy

Kim, Yu-Young

The rapid integration of generative AI into education necessitates new pedagogical models. However, current applications in Japanese language education often remain superficial, failing to cultivate the critical literacy and autonomy required for humanities majors. Existing research lacks a systematic framework that addresses AI's limitations while empowering learners to transition from passive consumers to active creators.

This study proposes a novel framework designed to foster Critical AI Literacy and Learner Autonomy. Structured in three stages－"Deconstruction & Critique," "Application & Control," and "Creation & Reflection"－the framework involves analyzing AI flaws, controlling local LLMs using Ollama, and creating persona-based chatbots using RAG techniques.

Application in a university course demonstrated a significant transformation in learners, shifting their role to active designers. Analysis confirmed the cultivation of sophisticated prompt engineering skills and a critical understanding of AI. This research contributes a replicable model leveraging AI's weaknesses as strengths, demonstrating that empowering learners to deconstruct and create with AI fosters the high-level critical thinking essential for humanities students in the AI era.