

A Lifecycle-Based Integrated Modeling Framework for Generative AI Services

Eun-Sook Cho*

*Professor, Dept. of Software Engineering, Seoil University, Seoul, Korea

[Abstract]

Generative AI services require lifecycle support for Prompt Engineering, Retrieval-Augmented Generation (RAG), user experience (UX), and continuous improvement, which existing frameworks such as CRISP-DM, OSEMN, MLOps, and CRISP-ML(Q) do not fully address. This paper compares these frameworks using lifecycle-based criteria and proposes an integrated framework that incorporates business understanding, UX, data and RAG design, Prompt Engineering, generative AI modeling, service design, LLMOps-based operation, monitoring and feedback, and continuous improvement. Comparative analysis and exploratory expert evaluation indicate the conceptual applicability and structural adequacy of the proposed framework for supporting the Generative AI service lifecycle. The results are exploratory and do not constitute empirical evidence of performance improvement.

▶ **Key words:** Generative AI Services, Service Lifecycle, Lifecycle-Based Modeling Framework, LLMOps, Prompt Engineering, Retrieval-Augmented Generation

[요 약]

생성형 AI 서비스는 프롬프트 엔지니어링, 검색 증강 생성(RAG), 사용자 경험(UX) 및 지속적 개선을 포함하는 생애주기 지원이 요구되지만, CRISP-DM, OSEMN, MLOps 및 CRISP-ML(Q) 등 기존 프레임워크는 이를 통합적으로 지원하는 데 한계가 있다. 본 논문에서는 생애주기 기반 평가 기준으로 기존 프레임워크를 비교·분석하고, 비즈니스 이해, UX, 데이터 및 RAG 설계, 프롬프트 엔지니어링, 생성형 AI 모델링, 서비스 설계, LLMOps 기반 운영, 모니터링 및 피드백, 지속적 개선으로 구성된 통합 프레임워크를 제안하였다. 비교 분석 및 탐색적 전문가 평가 결과, 제안 프레임워크는 생성형 AI 서비스 생애주기 지원을 위한 개념적 적용 가능성과 구조적 적절성을 보였다. 단, 이는 성능 향상에 대한 실증적 검증이 아닌 탐색적 평가 결과이다.

▶ **주제어:** 생성형 AI 서비스, 서비스 생애주기, 생애주기 기반 모델링 프레임워크, LLMOps, 프롬프트 엔지니어링, 검색 증강 생성(RAG)

I. Introduction

최근 생성형 인공지능(Generative AI) 기술의 급속한 발전은 다양한 산업 분야에서 새로운 서비스 패러다임을 형성하고 있다[1,2]. 특히 ChatGPT로 대표되는 생성형 AI 서비스는 자연어 기반의 유연한 상호작용을 통해 교육, 의료, 금융, 제조 등 다양한 분야에 적용되고 있다. 이러한 생성형 AI 서비스는 전통적인 예측 중심 인공지능 시스템과 달리 사용자와의 실시간 상호작용, 동적 콘텐츠 생성, 지속적 피드백 및 개선 구조를 포함한다.

기존 AI 모델링 프레임워크인 CRISP-DM[3], OSEMN[4], MLOps[5], CRISP-ML(Q)[6] 등은 데이터 분석과 모델 개발 및 운영을 중심으로 설계되어 있다. 그러나 생성형 AI 서비스는 Prompt Engineering, 검색 증강 생성(Retrieval-Augmented Generation, RAG), 사용자 경험(UX), 지속적 피드백 및 재학습 등 새로운 서비스 구성 요소를 포함한다[8-11]. 따라서 기존 프레임워크만으로는 생성형 AI 서비스의 전체 생애주기를 충분히 반영하기 어렵다.

생성형 AI 서비스는 모델 개발뿐 아니라 서비스 설계, 운영, 유지보수 및 응답 품질 관리 등을 포함하는 통합적 접근을 요구한다[7,11]. 최근 LLMOps[7]와 같은 생성형 AI 운영 중심 프레임워크가 등장하였으나, 전체 서비스 생애주기 관점에서 기존 프레임워크를 비교·분석한 연구는 아직 부족하다. 또한 생성형 AI를 활용한 자연어 기반 UML 모델링 및 대화형 AI 시스템 설계 연구[12]도 주로 시스템 분석 및 설계 단계에 초점을 두고 있어, 운영과 지속적 개선을 포함하는 서비스 생애주기 관점의 접근은 상대적으로 미흡하다.

이에 본 논문에서는 CRISP-DM, OSEMN, MLOps, CRISP-ML(Q), LLMOps를 대상으로 생성형 AI 서비스 관점의 평가 기준을 정의하고 각 프레임워크의 특성과 한계를 비교·분석한다. 이를 기반으로 생성형 AI 서비스의 구조적 특성을 반영한 **생애주기 기반 통합 모델링 프레임워크**를 제안하고 그 적용 가능성을 검토한다.

본 연구의 주요 기여점은 다음과 같다. 첫째, UX, Prompt Engineering, RAG, LLMOps 기반 운영 및 지속적 개선을 하나의 순환적 서비스 생애주기로 구조화한 End-to-End 모델링 프레임워크를 제시한다. 둘째, 생성형 AI 서비스의 생애주기 관점에서 기존 프레임워크에 동일한 평가 기준을 적용하여 적용 범위와 한계를 체계적으로 분석한다. 셋째, 기존 프레임워크와의 비교 분석 및 탐색적 전문가 평가를 통해 제안 프레임워크의 개념적 적용 가능성과 구조적 적절성을 검토한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 AI 모델링 프레임워크와 생성형 AI 관련 연구를 분석하고, 3장에서는 생애주기 기반 평가 기준을 정의하여 프레임워크를 비교·분석한다. 4장에서는 제안 프레임워크와 탐색적 전문가 평가 결과를 제시하며, 5장에서는 결론 및 향후 연구 방향을 제시한다.

II. Related Work

1. Traditional AI Modeling Frameworks

1.1 CRISP-DM

CRISP-DM은 데이터 마이닝 프로젝트를 위한 대표적인 프로세스 모델로서 Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation 및 Deployment의 6단계로 구성된다[3]. 데이터 분석 및 모델 개발 과정을 체계화한다는 장점이 있으나, Prompt, RAG 및 사용자 상호작용과 같은 생성형 AI 특화 활동은 명시적으로 포함하지 않는다.

1.2 OSEMN

OSEMN은 Obtain, Scrub, Explore, Model, iNterpret의 5단계로 구성된 데이터 사이언스 프레임워크이다[4]. 데이터 탐색과 분석에는 효과적이지만 생성형 AI 서비스의 운영과 지속적 개선을 지원하는 생애주기 관점은 포함하지 않는다.

1.3 MLOps

MLOps는 머신러닝 모델의 개발, 배포 및 운영 자동화를 지원하며, CI/CD, 모델 모니터링 및 재학습 등의 운영 활동을 포함한다[5]. 그러나 Prompt 생애주기 관리와 RAG 운영 등 생성형 AI 특화 활동은 핵심 구성요소로 명시하지 않는다.

1.4 CRISP-ML(Q)

CRISP-ML(Q)는 CRISP-DM을 확장하여 품질보증과 위험 관리 요소를 포함한 머신러닝 프로세스 모델이다[6]. 품질 중심의 ML 생애주기를 지원하지만, Prompt 생애주기 관리, RAG 및 사용자 경험과 같은 생성형 AI 특화 활동은 핵심 구성요소로 명시하지 않는다.

1.5 LLMOps

LLMOps는 대규모 언어모델 기반 서비스의 개발, 배포 및 운영을 지원하며, Prompt Engineering, RAG, 모델

Table 1. Comparison of Emerging Generative AI Development and Operations Approaches

Approach	Primary Focus	Lifecycle Perspective
PromptOps / Prompt Lifecycle Approaches	Prompt evaluation, versioning and operation	Prompt-centered lifecycle
GenAIOps	GenAI development, operation and model agility	GenAI operational lifecycle
Agent Engineering / AgentOps	Agent observability, monitoring, logging and analytics	Agent-centered lifecycle
AI Product Lifecycle Approaches	AI system design, deployment, monitoring and continual learning	Product-centered lifecycle
Proposed Framework	UX, RAG, Prompt, LLM, service operation and improvement	Integrated GenAI service lifecycle

평가 및 모니터링 등의 생애주기 활동을 포함한다[7]. 그러나 사용자 요구 및 UX 분석에서 서비스 운영과 지속적 개선에 이르는 전체 서비스 생애주기를 통합적으로 지원하는 데에는 한계가 있다.

2. Recent Trends in Generative AI Development and Operations

최근 생성형 AI 서비스의 확산과 함께 Prompt Engineering, RAG, AI Agent 및 LLM 평가를 중심으로 다양한 개발·운영 방법론이 제안되고 있다. 관련 연구에서는 Prompt 생애주기 관리와 AI Agent의 Planning, Memory 및 Tool Use 등을 다루고 있으며[13-16], RAGAs와 같은 평가 프레임워크를 통해 생성 결과의 품질을 평가하기 위한 연구도 이루어지고 있다[16].

또한 PromptOps[17], GenAIOps[18], Agent Engineering/AgentOps[19] 및 AI Product Lifecycle Approaches[20]와 같이 생성형 AI의 개발·운영을 생애주기 관점에서 관리하려는 접근도 제안되고 있다. [Table 1]과 같이 이들은 각각 Prompt, 생성형 AI 운영, Agent 및 AI 제품 생애주기에 초점을 두지만, 특정 기술 또는 운영 영역을 중심으로 한다. 이에 본 논문에서는 UX, RAG, Prompt, LLM 기반 서비스 운영 및 지속적 개선을 하나의 폐쇄형(Closed-loop) 생애주기로 연계하는 통합 프레임워크를 제안한다.

III. Comparative Analysis of Lifecycle-Based Modeling Frameworks for Generative AI Services

1. Analysis of the Generative AI Service Lifecycle

생성형 AI 서비스는 Prompt Engineering, RAG, 사용자 피드백 및 모델 개선이 서비스 운영과 연계되는 반복적 특성을 가진다. 이에 본 연구에서는 [Fig. 1]과 같이 사용

자 요구를 기반으로 Prompt와 LLM을 통해 응답을 생성하고, 필요시 RAG를 활용하며, 운영 결과와 사용자 피드백을 Continuous Improvement로 환류하는 폐쇄형 생애주기 구조를 정의하였다.

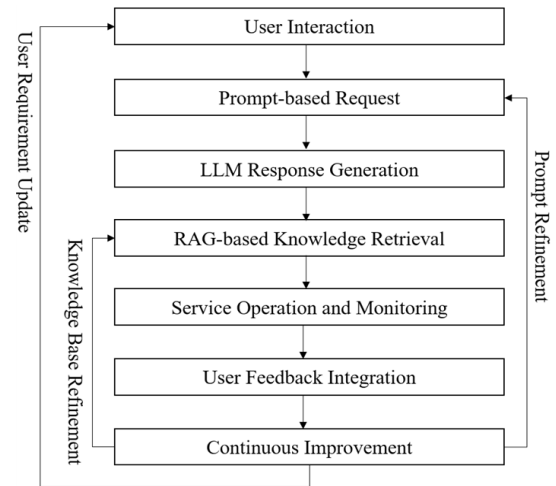


Fig. 1. Generative AI Service Lifecycle

2. Evaluation Criteria for Generative AI Service Modeling

기존 AI 모델링 프레임워크를 생성형 AI 서비스 관점에서 비교하기 위해 생성형 AI 서비스의 생애주기와 운영 요구사항을 기반으로 평가 기준을 도출하였다. 평가 기준은 CRISP-DM[3], OSEMN[4], MLOps[5], CRISP-ML(Q)[6] 및 LLMops[7]의 핵심 구성요소를 기반으로 하였으며, 생성형 AI 관련 연구[8-11, 13-15]에서 제시된 사용자 경험 (User Experience, UX), Prompt Engineering, 검색 증강 생성, 응답 품질, 지속적 개선 및 Hallucination 대응 요소를 추가로 반영하였다. [Table 2]는 평가 영역과 세부 평가 항목 및 도출 근거를 나타낸다.

본 연구는 프레임워크의 구조적 특성과 기능 지원 범위를 비교하는 정성적 분석을 목적으로 모든 평가 영역에 동일한 가중치를 적용하였으며, 향후 AHP 또는 Delphi 기법을 활용한 평가항목의 정량적 타당성 검증이 필요하다.

Table 2. Evaluation Criteria for Generative AI Service Modeling

Evaluation Criteria	Evaluation Items	Basis of Evaluation
User-Centricity	UX, User Participation	Generative AI and LLM Studies [1,2]
Generative AI Suitability	Prompt Lifecycle Management, RAG Support	Prompt Engineering and RAG Studies [8-10]
Data Processing Capability	Data Collection and Management	CRISP-DM, OSEMN, CRISP-ML(Q) [3,4,6]
Operational Efficiency	Deployment Automation, Monitoring	MLOps, LLMOps [5,7]
Quality	Response Reliability, Stability	CRISP-ML(Q), LLMOps and Hallucination Studies [6,7,11]
Maintainability	Retraining and Continuous Improvement	CRISP-ML(Q), MLOps, LLMOps [5,6,7]
Scalability	Service Scalability	MLOps, LLMOps [5,7]
Ethics and Safety	Hallucination Response	Hallucination Studies [11]

Table 3. Evaluation Rating Scale

Symbol	Meaning	Evaluation Criterion
◎	Fully Supported	The evaluation item is explicitly defined as a core component of the framework.
○	Partially Supported	The evaluation item is supported, but requires additional configuration or external components.
△	Limited Support	The evaluation item is only indirectly supported or depends on user implementation.
X	Not Supported	The evaluation item is not defined or supported by the framework.

각 프레임워크의 지원 수준을 일관되게 평가하기 위하여 [Table 3]과 같이 4단계 평가척도(◎: Fully Supported, ○: Partially Supported, △: Limited Support, X: Not Supported)를 적용하였다. 등급은 각 프레임워크의 원문 및 대표 문헌에서 평가항목에 대응하는 단계, 활동 및 기능의 명시 여부와 지원 방식을 기준으로 판정하였다. 동일한 판정 기준을 모든 프레임워크에 적용하였으며, 각 평가 결과의 주요 문헌 근거와 판정 이유는 비교분석 과정에서 함께 제시하였다.

3. Comparative Analysis of Generative AI Service Modeling Frameworks

본 절에서는 생성형 AI 서비스의 특성을 기반으로 기존 AI 모델링 프레임워크를 비교·분석한다. 비교 대상 프레임워크는 목적과 적용 범위가 서로 다르므로, 본 연구에서는 프레임워크 간 우열이 아니라 생성형 AI 서비스의 생애주기, 특화 기능 및 운영·유지보수 관점에서 기능 지원 범위를 동일한 기준으로 비교·분석하였다.

3.1 Comparative Analysis by Lifecycle Stage

생성형 AI 서비스 생애주기의 주요 단계에 대한 각 프레임워크의 지원 수준을 비교하였으며, 그 결과를 [Table 4]에 제시하였다.

분석 결과, CRISP-DM은 데이터 처리 및 배포 단계를 포함하므로 Data & Knowledge Management는 부분 지원(○), Service Operation & Monitoring과 Continuous Improvement는 제한적 지원(△)으로 판정하였으나, Prompt, RAG 및 사용자 피드백 관련 활동은 명시하지 않아 미지원(X)으로 판정하였다[3]. OSEMN은 데이터 분석 중심 구조로 Data & Knowledge Management만 부분 지원(○), 나머지 생성형 AI 생애주기 활동은 미지원(X)으로 판정하였다[4]. MLOps는 배포·모니터링을 핵심적으로 지원하므로 Service Operation & Monitoring은 완전 지원(◎), 데이터 관리·피드백·지속적 개선은 부분 지원(○), Prompt·RAG 관련 활동은 제한적 지원(△)으로 판정하였다[5]. CRISP-ML(Q)는 데이터·품질 및 반복적 ML 생애주기를 지원하므로 관련 항목은 부분

Table 4. Framework Support Levels by Generative AI Service Lifecycle Stage

Lifecycle Stage	CRISP-DM	OSEMN	MLOps	CRISP-ML(Q)	LLMOps
User Interaction	X	X	△	△	◎
Prompt Lifecycle Management	X	X	△	△	○
Data & Knowledge Management	○	○	○	○	◎
RAG Operation	X	X	△	△	◎
Service Operation & Monitoring	△	X	◎	○	◎
User Feedback Integration	X	X	○	△	◎
Continuous Improvement	△	X	○	○	◎

Table 5. Support Levels for Generative AI-Specific Features

Generative AI-Specific Feature	CRISP-DM	OSEMN	MLOps	CRISP-ML(Q)	LLMOps
Prompt Engineering	X	X	△	△	◎
Prompt Lifecycle Management	X	X	△	△	○
RAG Structure Support	X	X	△	△	◎
Vector DB Operation	X	X	△	△	○
Hallucination Handling	X	X	△	○	○
Response Quality Management	△	△	○	○	◎
Human Feedback Loop	X	X	△	△	△

지원(○), 생성형 AI 특화 활동은 제한적 지원(△)으로 판정하였다[6]. LLMOps는 Prompt, RAG 및 LLM 운영 활동을 명시적으로 지원하나 각 기능의 생애주기 내 명시 수준에 따라 ◎, ○ 또는 △로 판정하였다[7].

3.2 Comparative Analysis of Generative AI-Specific Features

생성형 AI 특화 기능은 각 프레임워크의 원문 및 관련 연구[8-11]를 근거로 평가하였으며, 결과는 [Table 5]와 같다.

CRISP-DM과 OSEMN은 Prompt Engineering, Prompt 생애주기 관리, RAG 구조, Vector DB 운영, 환각 대응 및 Human Feedback Loop를 명시적인 단계나 활동으로 정의하지 않아 해당 항목을 미지원(X)으로 판정하였다[3,4]. 다만 기존 데이터 및 모델 평가·해석 활동을 생성 결과의 품질 확인에 간접적으로 활용할 수 있다는 점에서 Response Quality Management는 제한적 지원(△)으로 판정하였다.

MLOps는 모델의 배포, 모니터링 및 재학습을 지원하므로 기존 운영 구조를 생성형 AI 서비스로 확장할 수 있으나, Prompt Engineering, Prompt 생애주기 관리, RAG 구조, Vector DB 운영 및 환각 대응을 독립적인 핵심 활동으로 정의하지 않아 제한적 지원(△)으로 판정하였다[5]. 또한 모니터링 및 모델 품질관리 기능을 응답 품질 관리에 활용할 수 있으므로 Response Quality Management는 부분적 지원(○)으로 판정하였으며, Human Feedback Loop는 기존 피드백 및 재학습 구조를 통해 간접적으로 지원 가능한 것으로 보아 제한적 지원(△)으로 판정하였다.

CRISP-ML(Q)는 Prompt Engineering, Prompt 생애주기 관리, RAG 구조 및 Vector DB 운영을 생성형 AI 특화 활동으로 직접 정의하지 않아 제한적 지원(△)으로 판정하였다[6]. 반면 품질보증 중심의 생애주기와 품질 요구사항 관리 기능을 생성 결과의 오류 및 품질 관리에 적용할 수 있다는 점에서 Hallucination Handling과 Response Quality Management는 부분적 지원(○)으로 판정하였으며, Human Feedback Loop는 독립적인 핵심 활동으로 정의되지 않아 제한적 지원(△)으로 판정하였다.

LLMOps는 LLM의 Adaptation 단계에서 Prompt Engineering과 RAG를 명시적으로 포함하고 있으며, Evaluation 단계에서 응답의 정확성, 유창성, 적절성, 관련성 및 일관성 등을 평가하고 Monitoring 단계에서 Prompt와 환각을 모니터링하도록 제시하고 있으므로 Prompt Engineering, RAG Structure Support 및 Response Quality Management는 완전 지원(◎)으로 판정하였다[7]. 또한 외부 Vector Database와의 통합을 지원하고 Prompt와 환각을 운영 과정에서 관리하므로 Prompt Lifecycle Management, Vector DB Operation 및 Hallucination Handling은 부분적 지원(○)으로 판정하였다. Human Feedback Loop는 LLMOps 생애주기의 독립적인 핵심 활동으로 명시되어 있지 않아 제한적 지원(△)으로 판정하였다[7].

3.3 Comparative Analysis of Operation and Maintenance

[Table 6]은 운영 및 유지보수 관점에서의 비교 결과를 나타낸다.

Table 6. Comparative Analysis of Operation and Maintenance

Evaluation Item	CRISP-DM	OSEMN	MLOps	CRISP-ML(Q)	LLMOps
Automated Deployment (CI/CD)	△	X	◎	○	◎
Model Monitoring	△	X	◎	○	◎
Drift Detection	X	X	○	○	○
Retraining Support	△	X	◎	○	○
Prompt Version Management	X	X	△	△	◎
Response Log Analysis	X	X	△	○	◎
Operational Scalability	△	△	◎	○	◎

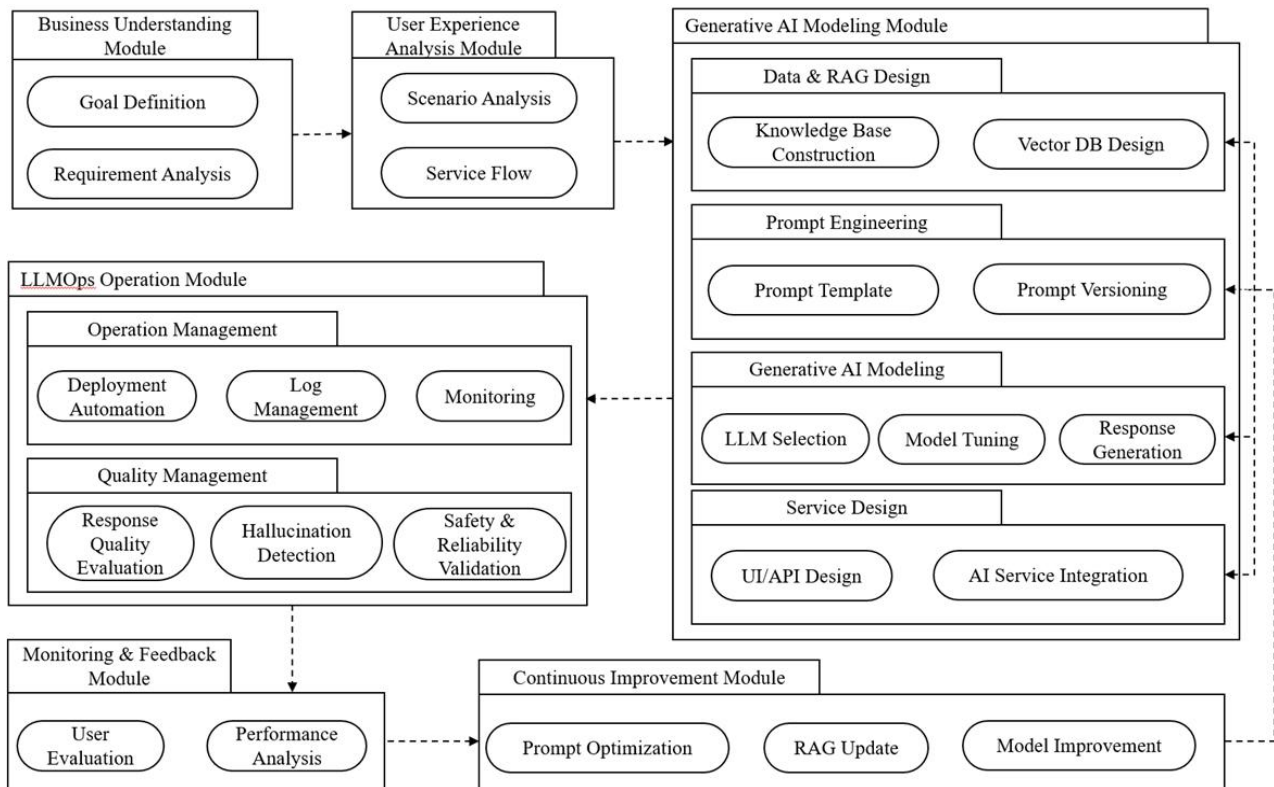


Fig. 2. Lifecycle-Based Integrated Modeling Framework for Generative AI Services

분석 결과, [Table 6]과 같이 CRISP-DM은 Deployment를 포함하지만 지속적 운영 기능이 명시적이지 않아 일부 항목만 제한 지원(Δ), OSEMN은 운영·유지보수 단계를 포함하지 않아 대부분 미지원(X)으로 판정하였다[3,4]. MLOps는 CI/CD, 모니터링 및 재학습을 핵심적으로 지원하여 해당 항목은 완전 지원($\odot$), Drift Detection은 부분 지원($\circ$), Prompt 및 응답 로그 관련 기능은 제한 지원(Δ)으로 판정하였다[5]. CRISP-ML(Q)는 품질관리 중심의 운영·유지보수 기능을 부분 지원($\circ$)하나 Prompt Version Management는 제한 지원(Δ)으로 판정하였다[6]. LLMops는 배포·모니터링 및 Prompt·응답 관리 기능을 지원하되, Drift Detection과 Retraining Support는 독립적인 핵심 활동으로 명시되지 않아 부분 지원($\circ$)으로 판정하였다[7].

4. Integrated Lifecycle-Based Modeling Framework for Generative AI Services

4.1 Structure of the Proposed Framework

앞 절의 비교 분석 결과, 기존 AI 모델링 프레임워크는 데이터 분석, 모델 개발, 운영 자동화 및 품질 관리에는 강점을 가지지만, 생성형 AI 서비스에서 요구되는 사용자 경험, Prompt Engineering, 검색 증강 생성, 응답 품질 관리 및 지속적 개선 구조를 통합적으로 지원하는 데에는 한

계가 있었다. 따라서 본 논문에서는 이러한 한계점을 극복하기 위해 [Fig. 2]와 같이 생성형 AI 서비스 생애주기 기반 통합 모델링 프레임워크를 제안한다.

제안 프레임워크는 Business Understanding, UX Analysis, Data & RAG Design, Prompt Engineering, Generative AI Modeling, Service Design, LLMops Operation, Monitoring & Feedback 및 Continuous Improvement의 9개 모듈로 구성된다. 각 모듈은 순차적으로 연계되며, 운영 과정에서 수집된 로그와 사용자 피드백을 기반으로 Prompt 최적화, RAG 갱신 및 모델 개선을 수행하고 그 결과를 관련 단계로 환류하는 폐쇄형 생애주기 구조를 형성한다.

4.2 Operational Definitions of Key Concepts

제안 프레임워크의 주요 개념을 명확히 하기 위해 Prompt Lifecycle Management, RAG Operation and Management, Response Quality Management, Monitoring & Feedback 및 Continuous Improvement를 범위, 주요 활동, 입력, 산출물 및 평가 기준으로 [Table 7]과 같이 정의하였다. 단, 평가 기준은 전문가 평가의 직접 측정 지표를 의미하지 않는다.

Table 7. Operational Definitions of Key Concepts

Key Concept	Scope	Main Activities	Input	Output	Evaluation Criteria
Prompt Lifecycle Management	Prompt lifecycle from design to improvement	Design, test, versioning, deployment, monitoring, optimization	Requirements, queries, prompts, response logs	Prompt templates, versions	Accuracy, relevance, consistency, performance change
RAG Operation and Management	External knowledge lifecycle management	Collection, chunking, embedding, retrieval, update	Documents, queries, retrieval logs	Vector DB, retrieved results	Precision/recall, context relevance
Response Quality Management	Response quality, reliability, and safety management	Quality evaluation, hallucination detection, relevance assessment	Responses, reference knowledge, user evaluation	Evaluation results, error types	Faithfulness, relevance, accuracy
Monitoring & Feedback	Operational performance and user feedback management	Log analysis, monitoring, feedback analysis	Service logs, feedback, metrics	Analysis results, improvement needs	Error rate, satisfaction, quality change
Continuous Improvement	Feedback-driven lifecycle improvement	Prompt optimization, RAG update, model improvement	Monitoring results, feedback	Improved prompts, updated RAG, model configuration	Pre/post improvement quality change

Table 8. Distinctive Features of the Proposed Framework Modules

Proposed Framework Module	Key Differences from Existing Frameworks
Business Understanding	Extends CRISP-DM business understanding by incorporating generative AI service scenarios, user interaction requirements, and prompt-driven objectives.
UX Analysis	Introduces a dedicated UX analysis phase to explicitly address conversational interaction, user scenarios, and service flows in generative AI services.
Data & RAG Design	Introduces explicit RAG design for external knowledge integration, which is not covered in traditional AI modeling frameworks.
Prompt Engineering	Defines prompt engineering as an independent lifecycle phase and incorporates prompt design, evaluation, and version management as explicit activities.
Generative AI Modeling	Optimizes large language model (LLM) selection and configuration for generative AI services.
Service Design	Incorporates user interface design and service integration beyond AI model development.
LLMOps Operation	Extends conventional LLMOps by integrating prompt lifecycle management, RAG operation, and response quality monitoring into service operation.
Monitoring & Feedback	Continuously evaluates service quality using user feedback and operational data.
Continuous Improvement	Provides a closed-loop feedback mechanism linking user feedback, prompt refinement, RAG updates, and LLM optimization, which is not explicitly defined in existing frameworks.

4.3 Distinctive Features of the Proposed Framework

제안 프레임워크의 각 모듈이 기존 프레임워크와 구별되는 주요 구조적 특징을 [Table 8]에 제시하였다. [Table 8]과 같이 제안 프레임워크는 UX, RAG 및 Prompt를 독립적인 생애주기 활동으로 구성하고, LLMOps 기반 운영과 Monitoring & Feedback을 연계하였다. 특히 운영 결과를 Prompt, RAG 지식베이스 및 모델 개선으로 환류하는 폐쇄형 Continuous Improvement 구조를 통해 전체 서비스 생애주기를 통합한 차별성이 있다.

IV. Evaluation

1. Evaluation Method

제안 프레임워크의 개념적 적용 가능성과 생애주기 지원 구조의 적절성을 탐색적으로 검토하기 위하여 생성형 AI 및 AI 시스템 개발 경험을 보유한 전문가 7명을 대상으로 5점 리커트 척도 평가를 수행하였다. 평가 대상은 생성형 AI 기반 서비스 개발 또는 관련 프로젝트 수행 경험을 보유하고, AI 시스템 구축·운영, AI 연구 및 정보시스템 감리 분야에서 5년 이상의 실무 또는 연구 경력을 갖춘 전문가로 선정하였다. 전문가 집단은 대학 교수 2명, AI 산업체 전문가 3명, 연구원 1명 및 정보시스템 감리 전문가 1명으로 구성하였다.

Table 9. Exploratory Expert Evaluation of Lifecycle Frameworks

Evaluation Item	CRISP-DM	OSEMN	MLOps	CRISP-ML(Q)	LLMOps	Proposed Framework
Support for Development Activities	3.1	4.0	4.2	3.5	4.4	4.7
Support for Operational Automation	2.1	2.0	4.8	3.2	4.9	4.9
Support for Prompt Lifecycle Management	1.5	1.4	3.0	3.1	4.8	5.0
Support for Maintenance Activities	2.3	2.1	4.3	4.0	4.9	5.0
Support for Response Quality Management	3.0	3.1	4.2	4.4	4.8	4.9
Average Score	2.4	2.5	4.1	3.6	4.8	4.9

평가는 제안 프레임워크와 비교 대상 프레임워크에 대해 동일한 평가 기준과 설명 자료를 제공한 후 평가자별로 독립적으로 수행하였다. 평가자는 다른 전문가의 평가 결과를 공유받지 않은 상태에서 개발 활동 지원, 운영 자동화 지원, Prompt 생애주기 관리 지원, 유지보수 활동 지원 및 응답 품질 관리 지원의 5개 항목을 5점 리커트 척도(1점: 매우 낮음, 5점: 매우 높음)로 평가하였다. 각 평가점수는 실제 서비스 적용에 따른 성능 개선 정도가 아니라, 각 프레임워크가 해당 생애주기 활동을 지원할 수 있는 정도에 대한 전문가의 판단을 의미한다. 평가 결과는 전문가 평가점수의 평균값을 이용하여 탐색적으로 비교하였으며, 평균값의 차이는 통계적 유의성이나 프레임워크 간 우수성을 의미하는 것으로 해석하지 않았다.

2. Exploratory Expert Evaluation

[Table 9]는 기존 프레임워크와 제안 프레임워크에 대한 탐색적 전문가 평가의 평균값을 나타낸다.

전문가 평가 결과, 제안 프레임워크와 LLMOps가 전반적으로 높은 평균값을 나타냈다. 다만 평균값의 차이는 근소하므로, 본 결과는 프레임워크 간 우수성보다는 제안 프레임워크의 개념적 적용 가능성과 생애주기 구조의 적절성에 대한 탐색적 평가로 해석하였다.

V. Conclusions

본 논문에서는 생성형 AI 서비스의 특성을 반영한 생애주기 기반 통합 모델링 프레임워크를 제안하기 위하여 CRISP-DM, OSEMN, MLOps, CRISP-ML(Q) 및 LLMOps를 생성형 AI 서비스 관점에서 비교·분석하였다. 분석 결과, 기존 프레임워크는 데이터 분석, 모델 개발 및 운영 자동화에는 강점을 가지지만 Prompt Engineering, 검색 증강 생성, 사용자 피드백 기반 지속적 개선 등 생성형 AI 서비스의 핵심 특성을 통합적으로 지원하는 데에는 한계가 있었다. 이에 본 연구에서는 생성형 AI 서비스의 기획, 개발, 운영 및 지속적 개선을 지원하는 생애주기 기

반 통합 모델링 프레임워크를 제안하였다. 전문가 기반의 탐색적 평가 결과, 제안 프레임워크는 생성형 AI 서비스 생애주기 관점에서 개념적 적용 가능성과 구조적 적절성을 긍정적으로 평가받았다.

다만 본 연구의 전문가 평가는 제안 프레임워크의 개념적 적용 가능성을 탐색적으로 검토한 것으로, 실제 생성형 AI 서비스 적용을 통한 개발 생산성, 응답 품질, 운영 효율성 및 유지보수성의 개선 효과를 실증적으로 검증한 것은 아니다. 또한 제한된 전문가 집단을 대상으로 평가를 수행하여 평가자 간 신뢰도 및 통계적 유의성에 대한 정량적 검증이 이루어지지 않았다는 한계가 있다.

향후 연구에서는 생성형 AI 챗봇, RAG 기반 질의응답 시스템 및 AI 에이전트 등 실제 서비스 개발 사례에 제안 프레임워크를 적용하여 개발 생산성, 응답 품질, 운영 효율성 및 유지보수성에 대한 효과를 정량적으로 검증할 예정이다. 또한 평가 대상과 전문가 집단을 확대하고 객관적인 평가 지표와 통계적 신뢰도 분석을 적용함으로써 제안 프레임워크의 실효성과 일반화 가능성을 검증하고자 한다.

ACKNOWLEDGEMENT

The present research has been conducted by the Research Grant of Seoil University.

REFERENCES

- [1] OpenAI, "GPT-4 Technical Report," arXiv:2303.08774, 2023. DOI: 10.48550/arXiv.2303.08774.
- [2] Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., et al., "A Survey of Large Language Models," arXiv:2303.18223, 2023. DOI: 10.48550/arXiv.2303.18223.
- [3] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R., "CRISP-DM 1.0: Step-by-Step Data Mining Guide," SPSS Inc., 2000.

- [4] Mason, H., and Wiggins, C., "A Taxonomy of Data Science Processes: OSEMN Framework," Columbia University Data Science Institute, 2010.
- [5] Tamburri, D. A., "Sustainable MLOps: Trends and Challenges," *Journal of Systems and Software*, Vol. 195, Article 111554, 2023. DOI: 10.1016/j.jss.2022.111554.
- [6] Studer, S., Biran, O., Custers, B., Joisten, D., Blomberg, S., et al., "Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology," *Machine Learning and Knowledge Extraction*, Vol. 3, No. 2, pp. 392-413, 2021. DOI: 10.3390/make3020019.
- [7] J. Díaz-De-Arcaya, J. López-De-Armentia, R. Miñón, I. L. Ojanguren, and A. I. Torre-Bastida, "Large Language Model Operations (LLMOps): Definition, Challenges, and Lifecycle Management," in *Proc. 2024 9th International Conference on Smart and Sustainable Technologies (SpliTech)*, Bol and Split, Croatia, 2024, pp. 1-4, doi: 10.23919/SpliTech61897.2024.10612341.
- [8] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G., "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing," *ACM Computing Surveys*, Vol. 55, No. 9, pp. 1-35, 2023. DOI: 10.1145/3560815.
- [9] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33, pp. 9459-9474, 2020.
- [10] Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Wang, Y., et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," *arXiv:2312.10997*, 2023. DOI: 10.48550/arXiv.2312.10997.
- [11] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., et al., "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," *ACM Computing Surveys*, 2025.
- [12] Eun-Sook Cho, "A Technique for UML-based System Development Using Generative AI," *Journal of The Korea Society of Computer and Information*, Vol. 29, No. 12, pp. 121-128, 2024. DOI: 10.9708/jksci.2024.29.12.121.
- [13] J. Xi, Y. Wang, L. Zhu, et al., "The Rise and Potential of Large Language Model Based Agents: A Survey," *ACM Computing Surveys*, 2025. DOI: 10.1145/3708558.
- [14] Sahoo, P., Singh, A. K., Saha, S., et al., "A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications," *arXiv:2402.07927*, 2024.
- [15] P. Zhao, H. Zhang, Q. Yu, et al., "Retrieval-Augmented Generation for AI-Generated Content: A Survey," *arXiv:2402.19473*, 2024.
- [16] Es, S., James, J., Espinosa-Anke, L., and Schockaert, S., "RAGAs: Automated Evaluation of Retrieval Augmented Generation," *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp.150-158, 2024.
- [17] U. Çetintemel, S. Chen, A. W. Lee, D. Raghavan, D. Lu, and A. Crotty, "Making Prompts First-Class Citizens for Adaptive LLM Pipelines," *Proc. Conf. on Innovative Data Systems Research (CIDR)*, 2026.
- [18] K. Ueno, M. Kogo, H. Kawatsu, Y. Uchiumi, and M. Tatsubori, "GenAIOps for GenAI Model-Agility," *CoRR*, vol. abs/2502.17440, 2025.
- [19] L. Dong, Q. Lu, and L. Zhu, "AgentOps: Enabling Observability of LLM Agents," *arXiv preprint arXiv:2411.05285*, 2024.
- [20] C. Huyen, *Designing Machine Learning Systems: An Iterative Process for Production-Ready Applications*, O'Reilly Media, 2022.

Authors



Eun-Sook Cho received the B.S. degree in Computer Science from DongEui University, Korea in 1993. She received the M.S. and Ph.D. degrees in Computer Science from Soongsil University, Korea, in 1996 and

2000, respectively. Dr. Cho joined the faculty of the Department of Software Engineering at Seoul University, Seoul, Korea, in 2005. She is currently a Professor in the Department of Software Engineering, Seoul University. She is interested in framework modeling and development, Software Quality, AI, BigData, SaaS, Cloud-Computing, and Digital Contents.