

Study on Application Cases of Generative AI PoC in Public Sector

Eungjoon Kim*

*Ph.D. Student, Graduate School of Public Administration, Seoul National University, Seoul, Korea.

[Abstract]

This study examines the feasibility of generative AI in public sector to enhance efficiency and service quality through a mixed-method approach. A Proof of Concept (PoC) analysis was conducted, evaluating AI performance using the F1 score and user satisfaction surveys, while employing Retrieval-Augmented Generation (RAG) for data recency. The results show that generative AI improves efficiency in repetitive tasks and receives high user satisfaction, contributing to workload reduction. Applying the Technology-Organization-Environment (TOE) framework, this study highlights key challenges in AI adoption, emphasizing the need for change management and social acceptance. These findings provide policy insights on structured AI adoption, ethical compliance, and data quality management in the public sector.

▶ **Key words:** PoC (Proof of Concept), generative AI, mixed-method, F1 score, RAG, TOE framework

[요 약]

본 연구는 공공행정 분야에서 생성형 AI 도입의 가능성과 한계를 고찰하고 이를 통해 행정 효율성 증진 및 공공 서비스의 질 향상을 모색하기 위해 혼합 연구 방법(mixed-method approach)을 적용하여 PoC(Proof of Concept) 사례를 분석하였다. AI 모델의 성능과 실질적 유효성을 평가하기 위해 정량적 지표인 F1 스코어와 정성적 방법인 사용자 만족도 조사를 함께 활용하였으며, 최신 데이터 반영을 위해 RAG(Retrieval-Augmented Generation) 기법을 적용하였다. 연구 결과, 생성형 AI는 법령 검색 및 자료 조사와 같은 반복적 업무에서 공무원의 효율성을 유의미하게 향상시키며 사용자 만족도 조사에서도 비교적 높은 점수를 획득하여 업무 경감에 실질적으로 기여하는 것으로 나타났다. 또한 TOE(기술-조직-환경) 프레임워크를 적용하여 공공부문 AI 기술 도입 시 발생할 수 있는 기술적, 조직적, 환경적 과제를 체계적으로 분석함으로써 AI 도입 과정에서의 변화 관리와 사회적 수용성 확보의 필요성을 강조하였다. 본 연구는 공공부문에서 생성형 AI의 도입 타당성을 제고하며 향후 디지털 전환 촉진을 위한 체계적 변화 관리 전략, 윤리적·법적 준수 강화, 데이터 수급 및 품질 관리의 중요성에 대한 정책적 시사점을 제공한다.

▶ **주제어:** PoC, 생성형 AI, 혼합 연구 방법, F1 스코어, RAG, TOE 프레임워크

- First Author: Eungjoon Kim, Corresponding Author: Eungjoon Kim
- *Eungjoon Kim (mrgorgeous@naver.com), Graduate School of Public Administration, Seoul National University
- Received: 2024. 12. 04, Revised: 2025. 02. 26, Accepted: 2025. 02. 26.

I. Introduction

공공부문에서 인공지능(AI)은 다양한 행정 및 서비스 부문에서의 혁신을 촉진하고 있다. 정부 지원금 및 혜택 판정[1], 시민 질문에 답변하는 챗봇(chatbots)[2], 검색 및 문서·양식 작성[3] 등 정부와 시민 간 소통 및 효율성을 높이는 새로운 패러다임으로 주목받고 있다. 이에 따라 행정 업무의 효율성을 향상시키고 공공 서비스의 질을 증진시키기 위한 AI 기술 도입이 점차 필수 과제로 부각되고 있다.

생성형 AI 기술은 대규모 언어 모델(Large Language Models)을 통해 자연어 처리의 정교함을 높이고 법률, 행정, 의료 등 전문적인 도메인에 특화된 모델을 통해 다양한 사용자 요구에 대응할 수 있다. 특히 공공부문에서 생성형 AI는 시민, 정부, 기업을 위한 혁신적 서비스의 촉진제(catalyst) 역할을 하며[4] 이를 도입함으로써 정부는 서비스 품질 향상, 생산성 증대, 비용 효율적인 서비스 확산 등의 혜택을 창출할 수 있는 것이다[5].

전 세계적으로 AI 지원에 대한 관심과 투자 역시 급격히 증가하고 있다. 유럽연합(EU) 집행위원회는 2020년부터 매년 200억 유로를 AI에 투자할 계획을 발표했고 또한 AI가 향후 10년 동안 미국, 독일, 일본의 경제 성장률을 최대 2%까지 증가시킬 수 있을 것이라는 분석도 있다[6]. 그러나 아직까지 공공 영역에서 공무원들이 생성형 AI를 어떻게 도입하고 활용하는지, 특히 공공 부문의 여러 분야에서 재량권, 활용성, 생산성에 미치는 영향을 분석한 연구는 부족한 상황이다[7]. 따라서 생성형 AI가 공공 부문에서 도입되는 과정을 살펴보고 그것이 실무자에게 미치는 영향에 대한 연구를 수행한다면 생성형 AI가 가지는 영향력을 보다 정교하게 이해하는 데 도움이 될 것이다[8].

본 연구는 TOE(기술-조직-환경) 프레임워크를 적용하여 공공부문에서 AI 도입 시 발생할 수 있는 문제를 체계적으로 분석하고 해결 전략과 시사점을 제시하는 데 목적이 있다. 이를 위해 공무원의 국유림 이용·대응 업무 대상으로 진행된 AI PoC(Proof of Concept) 과정을 사례로 선정하고 그 실제 효과는 F1 스코어와 사용자 만족도 조사를 통해 평가하였다. F1 스코어는 생성형 AI 모델의 성능을 정량적으로 측정하며 사용자 만족도 조사는 AI 기술이 현업 공무원의 업무 효율성에 미친 영향을 평가한다. 본 연구의 결과는 공공행정 분야에서 AI 도입의 효과성을 검증하고 AI 기술이 공공부문에서 어떻게 최적화되어야 하는지에 대한 실무적 시사점을 제공함으로써 AI 도입의 타당성과 가능성을 제고하는 데 기여하고자 한다.

이를 바탕으로 본 연구는 다음과 같은 연구질문을 설정하였다:

1. 생성형 AI는 공공행정에서 업무 효율성을 높이는가?
2. 생성형 AI의 F1 스코어는 실무적 활용이 가능한 수준인가?
3. 공무원들은 AI 도입이 업무 경감 효과를 제공한다고 인식하는가?
4. 공공부문 AI 도입에 있어 조직적·환경적 요인은 어떤 영향을 미치는가?

II. Preliminaries

1. Generative AI and Its Applicability in the Public Sector

생성형 AI는 공공부문에서 정책 개발, 서비스 제공, 지원 기능 강화, 규제 개선, 전략 최적화를 통해 효율성과 접근성을 높일 수 있으며 시민에 대한 서비스 품질을 개선할 것으로 기대된다[9]. 본 연구는 PoC 사례 분석과 함께 AI를 통한 법령 검색 및 자료 조사와 같은 반복적 업무에서의 시간 절감 효과를 입증한다. 이는 Carrasco 등의 연구가 제시한 AI의 공공서비스 활용 가능성을 뒷받침한다. 한편 공공부문에서 AI 도입이 효과성과 대응성을 높일 수 있지만 정치적 맥락과 공공 특수성으로 인해 제한적일 수 있다[10]. 본 연구의 PoC에서 발견한 AI 도입 단계에서 공무원마다 체감 정도에 편차가 있다는 점은 이러한 이론적 논의와 일치한다. 특히 국유림 이용 민원과 같이 규제와 이해관계가 얽힌 분야에서는 AI에 대한 신뢰가 쉽게 확보되지 않음을 확인할 수 있었다. 실제로 정부 부문의 경우 AI 도입이 제한적인 이유는 공공이라는 특수성과 정치적 맥락에서 비롯된다고도 볼 수 있는 것이다. 다만 공공 부문에서의 AI 활용은 공익을 증진해야 한다는 요구사항이 함께 있다[11]. 공공 AI 사용은 최소한의 투명성을 유지하여 AI에 대한 시민의 신뢰를 얻고 그 신뢰가 정당인지 확인할 필요가 존재한다[12].

국내 외의 AI 관련 선행연구를 검토한 결과 첫째, 다수의 연구들은 인공지능을 통해 현재 사회가 직면한 다양한 현안 해결이 가능할 것으로 내다보고 있다[13, 14]. 정부는 AI 연구 개발에 참여하여 예를 들면 사이버 방어, 도시화, 재난 관리 등에서 주도적 역할을 맡을 수 있을 뿐만 아니라 스타트업, 대기업과 협력하여 정부의 전통적인 영역을 넘어 민간 협력을 통한 문제 해결 분야로 역할을 확장할 수 있다. 둘째, 갈수록 복잡해질 미래의 문제를 대비하

여 고도의 AI 기술이 공공부문부터 필요할 것으로 보고 있다. 공공관리자들은 아직 완전히 다른 기술 세트를 요구하는 급진적 변화에 충분한 대비가 되어 있지 않다는 점이 한계로 작동하고 있으나[15]. 관료제의 비효율 극복을 위한 내부 효율성 향상이 AI를 채택하는 주요 동기로 자주 언급되고 있다[16]. 이외에도 복잡한 정책 문제 해결, 참여 증진, 공공 의사결정을 위한 동기에서 AI를 채택하려는 움직임도 있다[17]. 셋째, 인공지능으로 인해 발생할 수 있는 다양한 문제를 법적 테두리 내에서 해결할 수 있어야 한다고 보고 있다. 공공행정 영역에서 AI는 공정성, 개인정보 보호 등 윤리적 긴장을 동반한다[18]. AI의 적극적 사용은 막대한 이익을 제공하지만 윤리적 문제 등 사회나 공공 조직 내에 미칠 잠재적 위험을 고려해 전반적 영향을 평가할 필요가 있다[19]. 이미 여러 정부는 AI 사용에 관한 윤리 지침을 발표 중이다. 대표적으로 EUI의 윤리 지침(2019), 캐나다의 알고리즘 영향 평가(2020), 영국의 AI 지침(2019) 등이 있다. 이러한 지침에서도 정부가 점차 중요한 사용자로 자리잡고 있음은 물론 정부는 규제의 주체로서 앞으로 인공지능을 업무의 톨로 인정하는 과정에서 기존 법체계와 조화되는 방안을 고민해야 한다는 것이다[20]. 또한 새롭게 떠오를 수 있는 문제들, 예를 들면 인공지능의 성희롱 응답 생성, 장애인 혐오, 개인정보 유출 등[21]도 향후 AI를 도입하는 과정에서 정부가 윤리적 쟁점으로 고려해야 것들을 알 수 있다.

2. AI Design and Performance Evaluation

본 연구의 특화 생성형 AI 구성은 경량화 및 최적화 전략, 파인튜닝, RAG 기법, 그리고 MoM 기법 등으로 이루어졌다. MoM(모델혼합, Mixture of Models)은 여러 개의 개별 모델 또는 확률 분포를 조합하여 데이터의 복잡한 구조를 보다 정확하게 학습하는 기법으로 생성형 AI에서는 다양한 하위 모델을 결합하여 더 정교한 예측과 생성 능력을 제공하는 데 활용된다[22]. 또한 RAG(검색 증강 생성, Retrieval-Augmented Generation) 기법은 대형 언어 모델(LLM)의 한계를 극복하기 위해 외부 지식 검색(Retrieval)과 생성 모델(Generation)을 결합하여 보다 정확하고 신뢰할 수 있는 출력을 생성하는 AI 기법이다[23]. 위 구성 요소들이 상호 보완적으로 작용할 경우 질문에 가장 타당한 답변을 제공하는 특화 생성형 AI의 성능 극대화에 중요한 역할을 할 수 있게 된다[24].

AI 설계-구축 이후 성능 평가와 사용자 만족도 조사는 특화 생성형 AI가 공공부문에서 실질적인 효과를 발휘할 수 있는지 검증하는 중요한 과정이 된다. 공공부문에서 AI

기술을 도입하는 주요 목적은 행정 업무의 효율성을 높이고 공무원의 업무 부담을 줄이며 궁극적으로는 공공 서비스의 질을 향상하는 데 있다. 따라서 AI 모델의 성능이 실제 행정 환경에서 어느 정도 신뢰할 수 있는지를 평가하는 것은 필수적이다. 인공지능이 공공부문의 의사결정 지원에 효과적으로 활용되려면 모델의 신뢰성에 대한 평가가 이루어져야 하는 것이다[25].

이때 성능 지표인 F1 스코어는 모델이 실제 데이터를 다룰 때 얼마나 균형 잡힌 성능을 발휘하는지 보여준다. 데이터 세트가 불균형한 경우에도 F1 스코어는 정확한 긍정 및 부정 예측의 비율을 반영하여 생성된 응답이 균형을 유지하고 있는지 평가할 수 있다. 기술적으로 완전 자동화(full automation) 또는 AI 시스템의 출력물을 인간의 검토 없이 사용하는 경우에는 F1 점수가 1.0에 가까운 값을 요구한다[26]. 또한 사용자 만족도 조사는 공공부문에서 AI 기술 도입이 현장의 공무원들에게 얼마나 실질적 도움이 되는지를 평가하는 핵심 도구이다. AI의 성공적인 도입을 위해서는 기술적 성능뿐만 아니라 공무원들의 사용 경험과 만족도도 중요하게 고려되어야 한다. 이는 AI 도입이 단순히 기술적 혁신을 넘어 실제 조직 내에서 얼마나 효과적으로 정착할 수 있는지 판단하는 기준이 되기 때문이다. 결국 성능 평가와 사용자 만족도 조사는 공공부문에서 AI 기술의 실질적 적용 가능성을 입증하고 생성형 AI 도입이 공무원 업무와 공공 서비스 개선에 얼마나 기여할 수 있는지 근거를 제공함으로써 도입의 타당성을 제고하는 중요한 연구 과정이라 할 수 있다.

3. TOE Framework and AI Adoption in the Public Sector

TOE 프레임워크는 어떤 조직이 새로운 기술로부터 영향 받는 요인을 기술(Technology), 조직(Organization), 환경(Environment) 요소를 통해 설명한다[27]. TOE 프레임워크의 핵심 전제는 조직적, 환경적 맥락이 조직 차원에서의 기술 채택과 확산을 연구할 때 기술적 맥락만큼 중요하다는 점이다. 따라서 공공 조직 내 AI 확산을 살펴볼 때 TOE 프레임워크는 조직의 동화 패턴을 형성하는 내부 및 외부 환경과 관련된 측면을 포함할 수 있으므로 이론적 틀 측면에서 적합하다고 볼 수 있다[28].

첫 번째 기술적 요인은 AI 등 새로운 혁신 도구의 도입 과정에서 기술 측면을 고려해야 된다는 것이다. 기술적 차원은 내외부 기술장비, 시스템 설계 프로세스 및 관행 등을 탐구한다. 기술 인프라가 이미 갖추어진 상황에서는 기술의 높은 채택 비용을 낮춰 기술 채택을 촉진하는 요인으

로 작용하기도 한다. AI 채택의 맥락에서 기술적 요인은 AI 모델의 본질과 기능, 그리고 다른 옵션과 비교하여 어떤 기술이 적합한지를 포함하게 된다. 이와 함께 조직적 요인은 규모, 구조, 커뮤니케이션 메커니즘 및 의사 결정에 중점을 두고 조직과 자원을 특성화한다[29]. 공무원의 AI 활용 수준은 국가 및 기관별로 상이하지만 전반적으로 기본적인 디지털 역량은 갖추고 있으나 AI 기술의 심화된 이해와 실제 행정 업무에서의 적용 능력은 미흡한 것으로 평가된다[30, 31] 연구에 따르면 많은 공무원이 AI를 활용할 의향은 있지만 실제 행정 프로세스에서 이를 적용하는데 어려움을 겪고 있고 특히 데이터 분석, 알고리즘 이해, 자동화된 행정 절차 활용 능력에서 격차가 존재한다[32]. 따라서 공공행정에서 AI 도입은 단순한 기술적 변화에 그치는 것이 아니라 조직의 구조적 적응과 실질적 업무 효율성을 목표로 하기 때문에 변화 관리가 동반되지 않으면 AI 도입의 효과는 제한적일 수 있다. AI는 공공부문에서 데이터 기반의 의사결정을 지원하고 자동화된 행정 프로세스를 통해 거버넌스의 효율성과 투명성을 높일 수 있지만 동시에 윤리적·법적 문제를 해결하기 위한 규제와 감독이 필요할 것이다[33]. 장기적으로 AI 거버넌스는 공공 부문의 디지털 활용 역량을 높여 공공 서비스의 혁신을 촉진하면서도 시민 참여와 책임성을 강화하는 방향으로 발전해야 한다[34]. 환경적 요인은 AI 도입의 외부적 여건 등 법적 규제와 사회적 수용성 등을 포함한다. 공공행정에서는 개인정보 보호법과 전자정보 통신 관련 규제를 준수해야 하는 법적 요건이 필수적이며 이를 위반할 경우 법적 문제와 정부의 신뢰성 저하를 유발할 수 있다. 따라서 AI 기술의 활용은 공공부문 사용에 적합한 법적 규제를 준수하도록 설계되어야 하고 정부의 AI 도입이 사회적으로 수용되기 위해서는 투명성과 신뢰성이 확보되어야 할 것이다.

III. The Research Method

1. Research Design

본 연구는 생성형 AI PoC 사업을 통해 공무원의 업무 경감 효과를 탐구하고 향후 AI 구축에 실무적 시사점을 제공하기 위해 질적 사례연구와 정량적 분석을 병행하는 혼합 연구 방법을 채택하였다. 사례연구를 통해 AI가 공공부문에서 어떻게 실질적으로 적용 가능한지에 대한 종합적인 이해를 돕고 F1 스코어와 사용자 만족도 조사를 통한 객관적 지표를 바탕으로 PoC의 성과와 실무적 유효성을 검증하고자 하는 목적을 가진다.

사례연구란 연구자가 연구대상의 복잡적이고 종합적인 현상을 기술하고 분석하는 과정이다[35]. 또한 질적 사례 연구의 특성 중 하나는 발견적 이해의 추구로 여기서 발견적(heuristic)이란 연구를 통해 독자의 이해를 넓혀준다는 것을 의미한다[36]. 경험적 지식을 생산하고 전달하여 독자들이 실제 경험한 것과 마찬가지로의 이해를 촉진하게 되는 것이다[37]. 본 연구는 Merriam(1988)이 제시한 질적 사례연구의 네 가지 특성인 개별성(particularistic), 기술성(descriptive), 발견성(heuristic), 귀납성(inductive)을 중심으로 연구 대상의 적합성을 평가하였다. 결론적으로 본 연구는 Yin(1989)이 제안한 사례연구 설계의 기본유형 중 하나인 유형 I(단일사례-전체적 설계)에 대해 전체적 접근방법(holistic approach)을 취하여 하나의 사례를 전체적 단위로 분석하였다[38]. 이를 통해 AI PoC 사업의 실제 추진 과정과 성과를 심층적으로 분석하고 공공부문 전반에 걸쳐 AI 도입의 잠재적 가능성을 이해하는 데 필요한 정책적 시사점을 도출하고자 한다.

정량적 해석을 위한 정량적 지표로는 AI의 성능을 객관적으로 평가하는 F1 스코어를 활용하였다. F1 스코어는 모델의 정밀도(precision)와 재현율(recall)의 조화 평균으로 PoC에 사용된 생성형 AI가 국유림 이용 관련 질의와 응답에서 요구되는 정확성을 얼마나 충족하는지 확인하기 위한 성능 평가 지표이다. F1 스코어를 통해 AI가 현업 실무에 필요한 정확성과 실효성을 갖추고 있는지 확인함으로써 연구 결과의 타당성을 높이고자 했다. 더불어 본 연구에서는 AI 활용으로 인한 공무원의 업무 효율 증대 및 반복적 업무 경감에 미치는 영향을 파악하기 위해 추가적으로 내부 업무 담당자 20명을 대상으로 서면 설문조사를 실시하였다. 설문조사는 5점 척도를 활용하여 AI의 유용성, 효율성, 업무 경감 효과에 대한 주관적 만족도를 평가하였다. 뿐만 아니라 법령 검색 및 자료 조사와 같은 반복적 작업의 시간 단축 효과도 정량적으로 측정하였다. 전체 응답자 20명 중 100%의 응답률을 확보하여 신뢰성을 높였고 만족도 조사 결과는 공무원들이 AI를 직무에 어떻게 수용하고 있는지에 대한 실무적 유효성을 보여주어 PoC 사업 성과와 더불어 AI 도입의 현실적 적용 가능성을 평가하는 데 중요한 근거가 된다.

마지막으로 TOE(기술-조직-환경) 프레임워크를 적용하여 AI PoC 사업의 도입 과정에서 발생하는 기술적(T), 조직적(O), 환경적(E) 요인을 다각도로 분석하였다. 이 프레임워크는 AI PoC의 실행 가능성을 기술적 요인에만 국한하지 않고 조직 구성원의 수용성, 내부 변화 관리 요구, 법적 규제 등 복합적 관점에서 평가할 수 있는 구조적 틀을 제공한다. 기술적 측면에서는 생성형 AI의 성능, 데이터 수급 등을 평가

하며 조직적 측면에서는 AI 도입에 대한 구성원들의 수용도, 변화관리 과정, 교육 필요성을 중점적으로 다룬다. 환경적 측면에서는 공공부문에서 AI 도입을 둘러싼 법적규제 및 사회적 수용성을 분석하여 공무원의 직무 환경이 AI 도입에 미칠 수 있는 영향을 각각도로 탐구한다. 이를 통해 AI PoC가 단순한 기술 도입을 넘어 조직과 환경 전반에 미치는 영향을 체계적으로 조망하고자 한다. 다음은 본 연구의 연구 설계와 진행 과정을 시각화한 연구 방법론 개요도이다.

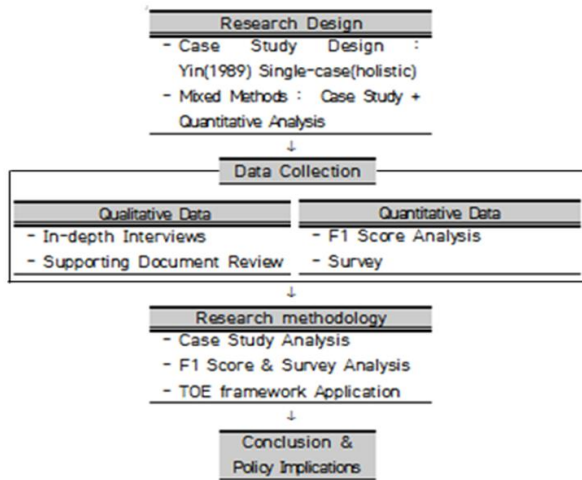


Fig. 3-1. Research Methodology Pview

2. Data Collection and Analysis

본 연구는 질적 자료 수집 과정에서 연구의 타당성과 신뢰성을 높이기 위해 삼각검증(Triangulation) 절차를 수행하였다[39]. 구체적으로 K사 담당자, 중앙행정기관의 빅데이터팀 실무자, 국유림이용 업무 실무자를 대상으로 심층 인터뷰를 진행하였으며 사업 계획서, 실무 회의록, 관련 문헌을 보조 자료로 활용하여 인터뷰 내용을 각각도로 검증하였다. 인터뷰 대상자의 인적사항은 다음과 같다.

Table 3-1. Interviewee Demographics

Affiliation	Position	Responsibilities
K Corporation	Manager	AI Strategy in the Public Sector
	Manager	Overall Management of AI Projects
	Deputy Manager	AI Development and Evaluation
Central Government Agency	Assistant Director	AI and Data Planning
	Senior Officer	AI and Data Quality Management
Subordinate Agency	Senior Officer	Handling Civil Complaints Related to National Forest Use
	Junior Officer	Handling Civil Complaints Related to National Forest Use

Table 3-1. This table summarizes the affiliations, positions, and responsibilities of the interviewees who participated in this study.

인터뷰 과정에서는 반구조화된 설문(Semi-structured interview)을 채택하여 생성형 AI의 개요, 개발 목적, 팀 협업 구조, 개발 단계, AI 구축 과정의 기술적 세부 사항(알고리즘-데이터 수집 방법 등), 윤리적 고려사항 및 규제 준수와 관련한 다양한 항목을 조사하였다.

정량분석을 함께 진행하기 위해 F1 스코어 계산에 필요한 주요 질문과 정답 쌍의 평가 데이터셋을 구성하였다. 평가 데이터셋은 국유림 대부 신청 관련 법령, 사용 허가 절차, 허가 유효 기간, 허가 취소 사유 등 공무원이 자주 직면하는 질문들을 아우르며 질의-답변 증강을 통해 총 1,000개의 질의-답변 쌍을 포함하도록 설계하였다. 이 데이터셋은 실무의 주요 주제를 고르게 반영하여 공무원 질의에 대해 신뢰성 있는 질의-답변쌍이 되도록 조작하였다. 실무자를 대상으로 한 설문조사는 Likert 5점 척도를 사용하여 응답자들이 AI에 느끼는 주관적 만족도와 객관적 효과를 구분하여 평가하도록 설계하였다. 문항은 크게 AI의 유용성, 업무 효율성, 업무 경감 효과를 측정하는 범주로 나누고 단일 항목으로 구성하였다. 설문조사는 PoC 참여자의 응답을 기반으로 진행되었고 각 항목에 대해 응답자가 자신의 업무 경험을 바탕으로 AI가 제공하는 효율성과 경감 효과를 평가하도록 하였다. 설문 문항은 기존 연구에서 IT 시스템의 효과성을 평가하기 위해 사용된 개념을 참고하여 설계되었다. Davis(1989)의 기술 수용 모델(Technology Acceptance Model, TAM)에서 제시한 Perceived Usefulness(인지된 유용성) 개념[40]과 DeLone & McLean(2003)의 IS 성공모형(Information Systems Success Model)에서 제안한 System Effectiveness(시스템 효과성) 개념[41]을 바탕으로 문항을 구성하였다. AI의 업무 경감 효과를 측정하는 항목은 주로 법령 및 자료 검색 시간을 중심으로 한 작업 효율성에 대한 체감도를 평가하도록 설계하였다. 또한 기존 연구에서도 IT 시스템이 업무 처리 시간 단축과 효율성 향상에 미치는 영향을 평가하는 방식이 사용[42]되었음을 참고하였다. 수집된 데이터는 Excel 및 Python의 Pandas 라이브러리를 사용하여 정제하고 분석하였다. 평균과 표준편차와 같은 기본적인 통계를 통해 응답자들의 만족도와 업무 경감 효과에 대한 전반적인 경향성을 파악할 수 있었다.

IV. The Research Result

1. Stakeholder Requirements Analysis

본 PoC는 온오프라인을 통해 국유림 이용 업무를 담당하는 일선 담당자들로부터 지속적으로 의견을 수렴하며

진행되었다. 실무자들의 의견을 정리하면 다음 표와 같다.

Table 4-1. Practitioner Requirements

Role	Requirements
○○ Official (Low-Skilled)	- Requires detailed guidance on national forest use permits and leasing procedures, as handovers alone are insufficient. - Expects LLM to suggest relevant laws for queries.
□□ Official (Highly-Skilled)	- Emphasizes strong security measures for sensitive data and a record-keeping function for outputs and decision-making processes. - Requires automated document generation for forms like permits and contracts. - Needs LLM to adapt promptly to policy or legislative changes due to frequent updates.
△△ Official (Highly-Skilled)	- Concerned about false outputs undermining administrative credibility. - Highlights the complexity and personalized nature of national forest-related inquiries, needing thorough legal and case law analysis. - Recommends search functionalities for procedures and relevant laws, particularly for training new employees.

Table 4-1. This table presents the key role-specific requirements and expectations regarding AI implementation in administrative tasks.

2. Design and Performance Evaluation of Generative AI

2.1 Technical Design

LLM MoM(Mixture of Models)은 다양한 언어 모델을 혼합하여 하나의 통합된 시스템을 구축하는 구성법으로 각 모델이 특정 작업에 최적화되어 높은 성능을 발휘할 수 있도록 한다. MoM 방법을 선택한 k사 개발팀에 따르면 MoM은 입력 데이터의 특성이나 작업의 요구 사항에 따라 최적화된 모델을 선택적으로 사용함으로써 시스템 성능을 극대화하고 계산 효율성을 높이는 데 기여할 수 있다. 즉 단일 모델이 모든 작업에서 최고의 성능을 발휘하는 것은 현실적으로 어렵기 때문에 모델을 혼합하여 다양한 작업 환경에서 높은 성능을 유지할 수 있도록 시도하는 것이다. 다만 각 모델의 학습과 최적화를 개별적으로 수행해야 하므로 전체 시스템의 최적화를 이루는 데 시간이 많이 소요될 수 있다는 한계점이 존재한다.

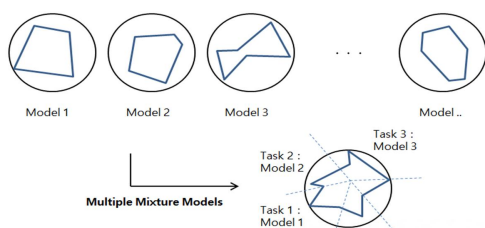


Fig. 4-1. Overview of LLM MoM

생성형 AI가 겪는 문제 중 하나가 할루시네이션(hallucination)이다. 모델이 마치 인간이 만든 것 같은(human-like) 텍스트를 재현하여 질문에 대해 자신있게 답변을 생성하지만 제한된 훈련 데이터나 일반화된 지식을 기반으로 답을 생성하는 LLM의 특성상 실은 그 답변이 부정확하거나 근거 없는 정보를 포함하는 경우가 종종 발생한다. 할루시네이션 현상은 MoM 기법으로는 해결이 어렵다. 이를 보완하기 위한 기술 중 하나가 RAG방식이다. RAG(Retrieval-Augmented Generation)는 LLM의 출력 품질을 향상시키기 위해 검색 기술을 결합한 방법으로 외부 데이터베이스를 참조하여 답변의 신뢰성을 높이는 기술적 방식을 취하게 된다. 다수 데이터를 학습하는데 있어 파인튜닝은 질문과 답변 쌍을 직접 제작한다면 RAG에 활용되는 임베딩(Embedding) 기술은 수집한 데이터를 벡터 기반 데이터베이스에 업로드하는 방식으로 이루어진다. 따라서 RAG 방식은 모델이 답변을 생성하기 전에 외부 데이터베이스에서 관련 문서를 검색하고 그 결과를 참조하도록 하여 신뢰할 수 있는 답변을 생성하게 된다. 또한 RAG 방식은 최신 정보를 반영함으로써 할루시네이션 문제를 완화하는 데 기여한다. LLM은 고정된 데이터셋을 기반으로 훈련되기 때문에 시간이 지남에 따라 최신 정보나 규정 변화를 적시에 반영하기 어렵다. 이는 법령이나 규정이 자주 변경되는 국유림이용과 같은 분야에서 할루시네이션 문제를 일으키는 기술적 원인이 된다. 반면 RAG 방식은 벡터 DB의 지속적인 업데이트를 통해 최신 정보를 반영함으로써 모델이 과거 데이터에 기반한 부정확한 정보가 아닌 최신의 정확한 정보를 토대로 답변을 생성하도록 한다.

k사는 RAG 기술을 효과적으로 구현하기 위해 다음과 같은 방식으로 시스템을 구성하였다. 먼저 기본 모델에 대한 파인튜닝을 완료한 후 RAG 기술을 적용하게 되는데 이 과정에서 국유재산 관련 법령, 국유림 이용에 관한 법률 및 산림분야 법률 전체, 국민신문고 민원 답변 데이터, 관련 판례 데이터를 활용했다. 본 PoC에서 법률 같은 경우는 한 개의 ‘조’를 하나의 청크 단위로 정하였다.

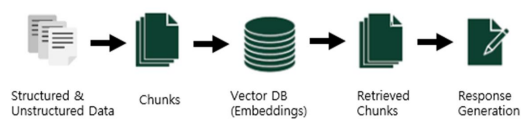


Fig. 4-2. Structure of RAG

2.2 Performance Evaluation①: F1 Score

본 연구에서는 국유림 이용 업무의 질의-답변 자동화 모델의 성능을 정량적으로 평가하기 위해 F1 스코어 테스트

트를 실시하였다. F1 스코어는 정밀도와 재현율의 조화 평균을 구하여 모델의 정답 예측 능력을 평가하는 지표로 AI 모델이 공무원의 질의에 대해 정확하고 신뢰성 있는 답변을 제공할 수 있는지 평가하기 위해 이 지표가 필요하였다. 특히 공공행정 분야에서는 답변의 정확성과 일관성이 중요하게 요구되므로 F1 스코어를 통해 평가함으로써 모델의 실질적 활용 가능성을 확인할 수 있을 것이다.

F1 스코어를 도출하기 위해 국유림 이용 관련 질의-답변 쌍으로 평가 데이터셋을 구성하였다. 평가 데이터셋은 실무에서 자주 다루어지는 질문과 그에 대한 정답을 우선 분류하고 챗GPT를 활용해 질의-답변 쌍을 증강함으로써 총 1,000개의 데이터로 구성하였다. 여기에는 국유림 대부분 신청에 필요한 서류 및 절차와 같은 행정적 질문, 사용 허가의 유효 기간, 허가 취소 사유 등 공무원이 실무에서 빈번하게 직면하는 질문들을 포함하였다. 이외에도 질의-답변 쌍은 국유림 대부분 및 사용 허가 관련된 법령, 절차적 요건, 환경영향 평가 요구사항 등 세부 주제별로 고르게 분포하도록 하여 모델이 응답의 신뢰성과 정확성을 유지할 수 있도록 구성하였다.

F1 스코어를 산출하기 위해 평가 데이터셋을 True Positives(TP), False Positives(FP), False Negatives(FN), 그리고 True Negatives(TN)으로 분류하였다. 각 지표는 다음과 같이 도출하였다. TP는 모델이 정답이라고 예측했을 때 실제로도 정답인 경우로 본 연구에서는 TP가 570으로 집계되었다. 예를 들어 '국유림 대부분 신청에 필요한 서류는 무엇인가요?'라는 질문에 대해 실제 정답이 '신청서, 계획서, 재산세 납부 증명서'라고 할 때 모델이 정확히 이와 동일한 답변을 제공한 경우는 TP로 간주된다. FP는 모델이 정답이라고 예측했으나 실제로는 정답이 아닌 경우로 본 연구에서는 FP가 190으로 집계되었다. '국유림 사용료 산정 방식은 무엇인가요?'라는 질문의 실제 정답이 '산림청 고시 가격 기준'임에도 불구하고 모델이 잘못된 답변인 '재산세 기준'을 제시한 경우 FP로 기록된다. 본 연구에서는 총 190개의 사례에서 모델이 오답을 정답으로 잘못 예측하여 FP로 분류되었다. FN은 실제로는 정답이지만 모델이 이를 정답으로 예측하지 못한 경우로 본 연구에서는 FN이 180으로 집계되었다. '국유림 사용 허가의 유효 기간은 얼마인가요?'라는 질문에 대해 실제 정답이 '3년'인 상황에서 모델이 무응답 처리하거나 잘못된 기간을 제시한 경우 FN으로 분류된다. 본 연구에서는 총 180개의 사례에서 모델이 실제 정답을 부정으로 잘못 예측하여 FN으로 집계되었다. 마지막으로 TN은 모델이 정답이 아니라고 예측했고 실제로도 정답이 아닌 경우로 본 연구에서는 TN이 60으로 집계되

었다. TN의 경우 F1 스코어의 계산에 직접적으로 필요하지 않지만 공공행정 분야에서는 모델이 오답이 아닌 것을 잘 걸러내는 능력도 중요하게 고려할 수 있는 부분이기 때문에 모델의 전반적 성능을 평가하는 차원에서 구분하였다. 이를 바탕으로 정밀도(Precision)와 재현율(Recall)을 계산하면 다음과 같다.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{570}{570 + 190} \approx 0.75$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{570}{570 + 180} \approx 0.76$$

정밀도와 재현율을 토대로, F1 스코어는 다음과 같이 계산되었다.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.75 \times 0.76}{0.75 + 0.76} \approx 0.755$$

F1 스코어가 1에 가까울수록 모델이 높은 정밀도와 재현율을 동시에 달성하고 있음을 나타낸다. F1 스코어가 0.7에서 0.8 사이에 있는 경우 모델이 비교적 높은 수준의 정밀도와 재현율을 달성했음을 의미한다. 즉 일반적으로 F1 스코어가 0.7에서 0.8 사이일 때 모델의 성능은 '공정(fair)'한 수준으로 간주된다[43]. F1 스코어가 0.755라는 것은 특정 수준 이상의 정확성과 신뢰성을 갖춘 응답을 지속적으로 제공할 수 있음을 시사한다. 즉 빈번하게 등장하는 기본적인 질문에 대해 일관된 응답이 가능하다는 점에서 실무적으로 의미를 갖는 것이다.

2.3 Performance Evaluation②: User Satisfaction Survey

본 연구의 사용자 피드백 및 만족도 조사에 따르면 AI가 공무원의 업무 효율성을 향상시키고 반복적 업무의 부담을 경감하는 데 일정 부분 기여하고 있는 것으로 나타났다. 본 만족도 조사는 내부 업무 담당자 20명을 대상으로 수행되었고 전원이 설문문에 응답하여 조사의 신뢰성을 확보하였다. 설문은 AI 시스템의 유용성, 효율성, 및 업무 경감 효과를 평가하기 위해 5점 리커트 척도로 구성되었고 법령 검색과 자료 조사와 같은 반복 작업에서 시간 절감 효과를 정량적으로 측정하였다.

설문조사 결과에 따르면 반복적 업무 경감 효과에 대한 평균 만족도 점수는 3.8점(SD= 0.6)으로 전체 응답자의 70%가 4점 이상의 높은 만족도를 보였다. 이는 AI가 공무원의 반복적이고 비효율적인 업무 부담을 경감하는 데 실

질적 기여하고 있음을 나타낸다. 그러나 평균과 표준편차를 함께 고려할 때 응답자들이 AI의 효과를 다소 상이하게 체감하고 있음을 알 수 있다. 즉 AI가 제공하는 업무 경감 효과가 응답자의 개인적 업무 특성이나 역할에 따라 다르게 인식될 가능성이 존재함을 시사한다. 업무 효율성 항목에서도 유사한 경향이 나타났다. 평균 점수가 3.7점(SD=0.56)인 점을 고려하면 AI가 민원 처리와 같이 복잡하고 시간이 소요되는 업무에서 보조 역할을 수행하는 데 기여하고 있음을 보여준다. 표준편차가 0.56으로 응답자 간 편차가 확인되었는데 이는 효율성 증대 효과가 사용자의 업무 성격이나 숙련도에 따라 다르게 체감될 수 있음을 암시하기도 하나 4점의 분포가 약 65%를 차지하고 있음을 볼 때 검증 과정에 처음부터 완료까지 참여했던 내부직원의 특성상 가능한 긍정적으로 평가하려 했던 측면도 엿보인다. 한편 법령 및 자료 검색 작업에서의 시간 절감 효과는 상대적으로 두드러진 것으로 나타났다. 응답자들은 평균적으로 30분이 소요되던 법령 및 자료 검색 작업이 AI의 도움으로 약 23분으로 줄어 약 7분의 시간 절감 효과를 경험한 것으로 응답하였다(SD= 6분). 이는 법령 및 자료 검색 작업에서 약 23%의 효율성 향상이 이루어진 셈이지만 표준편차가 6분으로 다소 큰 편으로 사용자에 따라 효과를 덜 체감했을 가능성이 있다. AI 기술검증 과정에 참여한 실무자의 만족도 및 업무 효율성 점수 분포는 다음과 같다.

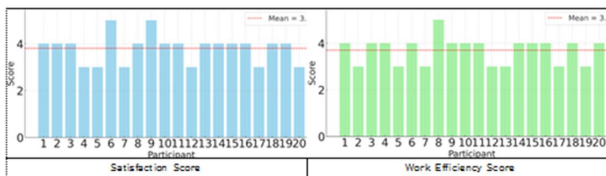


Fig. 4-3. AI Satisfaction and Efficiency Score Distribution

이상의 결과를 통해 AI가 공무원의 반복적 업무와 자료 검색 업무의 효율성 향상에 기여할 수 있음을 확인했다. 특히 법령 검색 및 자료 조사 등 시간이 많이 소요되는 작업에서 실질적인 시간 절감 효과가 나타난 점은 AI가 공무원 업무 효율성을 높이는데 유의미한 역할을 하고 있음을 시사한다. 그러나 만족도와 효율성 평가 점수에서 응답자 간 표준편차가 나타났듯 효율성 향상 효과가 사용자마다 다르게 체감되고 있어 시스템의 일관된 효과성이 실무 현장에서 완전히 보장되지는 않고 있음을 알 수 있다. AI가 일부 작업에서는 높은 성과를 발휘하지만 어떤 작업에서는 상대적으로 덜 유용하게 작용할 가능성이 있는 것이다.

3. Analysis of PoC Results Using the TOE Framework

3.1 Technological Perspective

특화 생성형 AI 구축에서 데이터 수급은 모델 성능의 핵심 요소로 기능한다. 정보 품질 이론(Information quality theory)에 따르면 데이터의 품질과 양은 모델의 학습과 추론 정확도를 좌우하며 특히 정확성, 신뢰성, 최신성이 중요하게 작용한다[44]. 모델의 학습과 추론 정확도는 데이터의 품질과 양에 직접적으로 의존할 수밖에 없기 때문이다. 공무원의 국유림 사용허가 및 대부 업무를 돕기 위한 특화 모델의 경우 관련 데이터가 부족하거나 부정확할 경우 모델은 업무 관련 맥락을 제대로 이해하지 못해 부정확한 결과를 도출하게 된다. 이를 방지하기 위해 다양한 출처로부터의 데이터 수집 및 통합이 필수적이며 다음과 같은 요인들이 주요하게 고려된다.

첫째, 파인튜닝 과정에서 데이터 수급의 중요성은 모델의 전문성과 정밀성에 직접적으로 영향을 미친다. 일반적인 LLM은 대규모의 텍스트 데이터로 훈련되어 다양한 주제에 대해 기본적인 언어 이해 능력을 갖추고 있지만 특정 도메인에 특화된 지식은 없어 특화 LLM을 구축할 때에는 파인튜닝이 필요하다. 본 PoC의 경우 2만 건의 데이터를 필요로 했으나 이를 현실적으로 충족시키기에는 어려움이 있어 데이터 증강 기술을 활용하였다. 학습 과정에서 부족한 부분은 실무자와 개발자 간 끊임없는 소통을 통해 정확성을 향상시킴으로써 양적 차이를 질적 개선으로 보완하였다.

둘째, RAG 방식을 활용한 특화 생성형 AI 구축에서도 데이터 수급은 핵심적 역할을 한다. RAG 방식은 메타데이터를 기반으로 한 정리된 데이터를 필요로 하는데 이는 특정 도메인 키워드와 개념이 포함된 데이터베이스에서 정확한 정보가 검색될 수 있도록 데이터를 구조화하고 라벨링하는 과정을 요구하므로 정보의 관리 및 접근성 관점(Knowledge access framework)과도 관련이 깊다[45].

셋째, 데이터수급 과정에서 국유림 민원을 담당하는 실무자와 AI 개발자간 협업이 중요하다. 개발자가 데이터를 수집하고 선별할 때 도메인 전문가의 의견을 반영함으로써 모델이 학습할 데이터의 품질과 정확성을 향상시킬 수 있기 때문이다. 실무자들은 어떤 정보가 중요한 자료인지, 어떤 자료가 신뢰할 수 있는지 깊은 이해를 갖고 있으므로 이들의 참여는 데이터 수집 과정에서 발생할 수 있는 오류를 줄이고 신뢰할 수 있는 데이터셋을 구성하는 데 기여한다. 이는 TOE 프레임워크 상의 조직적 측면과도 연결이 된다. 파인튜닝과 RAG 방식을 통해 LLM을 구축할 때 데이터 수급은 모델의 성능, 정확성, 신뢰성에 직접적인 영향을 미친다. 따

라서 데이터의 양과 질, 최신성과 포괄성을 확보하고 도메인 전문가와 협업하여 데이터를 수집·준비해야 한다. 이를 통해 LLM이 국유립 민원 등 공무원 질의응답 대응 분야에서 최상의 성능을 발휘할 수 있도록 해야 한다.

3.2 Organizational Perspective

본 PoC는 LLM 도입의 실효성을 검증하고 사업 주체인 중앙행정기관 내 다양한 이해관계자들이 변화에 적응할 수 있도록 지원하는 것을 목적으로 한다. 조직 내에서 AI 기술 도입은 여전히 많은 도전에 직면하고 있다. 실제로 AI 도입과 같은 거대한 규모 및 문화적 변화가 일어나는 곳에서는 격렬한 저항이 발생한다[46]. 저항의 원인을 명확히 분석하고 효과적인 변화관리 전략을 수립하는 것이 향후 추가적인 AI 도입과 공무원 업무 효율화를 위해 필수적으로 요구된다. 이때 변화관리(change management)라 함은 신기술의 도입으로 지적 변화가 수반될 경우 책임과 지원을 늘려 저항을 줄이는 과정을 일컫는다[47].

PoC 초기 단계에서 발생할 수 있는 주요 저항의 원인 중 하나는 LLM 기술에 대한 불안감이다. 공공기관의 특성상 정확성과 신뢰성이 필수적이기 때문에 LLM이 실제 환경에서 얼마나 정확한 답변을 제공할 수 있는지 우려가 나타날 수 있다. 특히 국유립 민원과 같이 이견이 첨예하게 대립하는 분야에서는 LLM이 부정확한 답변을 제공할 가능성이 더 높다. 이러한 상황에서 핵심적인 변화 관리 방법은 명확하고 일관된 커뮤니케이션 전략이 될 수 있다. PoC 과정에서의 목표, 진행 상황, 예상 효과 등을 조직 내 모든 이해관계자에게 투명하게 공유함으로써 변화의 이유와 기대되는 결과에 대한 명확한 설명이 이루어지면 조직 구성원들은 변화의 불확실성을 스스로 줄이고 새로운 기술 도입의 정당성을 인식하게 된다. 실제로 목표설정이론(goal setting theory)에 따르면 목표의 구체성(goal specificity)이 조직 구성원의 동기유발에 영향을 미치기 때문이다[48].

변화관리의 목적은 변화에 소요되는 시간과 노력 및 구성원의 저항을 최소화하고 미래의 목표는 최대화하는 것이다[49]. 저항을 효과적으로 관리하기 위한 세부 전략으로 조직 내 상향식과 하향식의 각기 다른 레벨에서의 변화 프로그램을 추진하는 것이 기술 실험을 성공시키는 방법이자 프로세스 그 자체로도 훌륭한 방법이 될 수 있다[50]. 또한 명확한 상호소통 전략의 수행이 요구된다. 많은 연구에서 변화 관리의 핵심은 내부자 간 커뮤니케이션임을 강조하고 있다. 이는 단순히 실행 단계에서 뿐만 아니라 변화를 위한 최초 협의로부터 합의를 이끌어내는 데서도 마찬가지인 것이다[51].

3.3 Environmental Perspective

특화 생성형 AI 구축을 위한 데이터학습 과정에서 개인 정보 보호와 관련한 문제를 해결하기 위해 두 가지 접근 방식이 가능했다. 첫 번째 방법은 데이터를 전처리하는 과정에서 민감한 개인정보를 수작업으로 삭제하는 것이다. 이러한 접근 방식은 비교적 안전하게 개인정보를 보호할 수 있지만 대규모 데이터셋에 적용하기는 많은 시간과 비용이 소모되는 단점이 있다. 두 번째 방법은 개인정보를 특정 개인과 직접 연관지을 수 없도록 가명처리 하는 것이다. 예를 들어 이름을 임의의 문자열로 대체하거나 주소를 특정 범위 내에서 일반화하는 방법이 있다. 가명처리된 데이터는 본래의 데이터를 재구성할 수 없도록 설계하는 것이 핵심이며 데이터가 대규모라 수작업으로 개인정보를 삭제하기 어려운 경우에 유용하다. 본 PoC 사업에서는 담당 공무원들이 직접 문서를 살피고 개인정보를 삭제하는 방식으로 진행하였다. 개인정보 보호는 본 PoC에서 가장 중요한 윤리적 고려사항 중 하나로 민원 및 판례 데이터를 활용함에 있어 개인정보 보호법을 준수해야 했다. 따라서 데이터 학습 과정에서는 모든 개인정보를 비식별화 처리하여 사용하거나 담당자가 문서를 하나씩 확인하여 개인정보를 삭제함으로써 개인정보의 노출 및 오용을 방지하고 민원인의 권리를 보호하기 위한 윤리적 조치를 진행하였다.

V. Conclusions

본 연구는 TOE(기술-조직-환경) 프레임워크를 활용해 특화 생성형 AI PoC 사업을 분석하여 공공행정 분야에의 AI 기술 도입 과제를 살펴보고 이를 바탕으로 행정 효율성 및 공공 서비스의 질 향상을 위한 실질적 시사점을 제시하고자 하였다. 이 과정에서 AI 시스템의 정량적 평가와 내부 사용자 대상의 만족도 조사를 함께 진행하였다. 연구 결과는 다음의 주요 시사점을 제공한다.

첫째, 본 연구의 PoC 사업을 통해 공공부문에서의 특화 생성형 AI 기술 도입이 반복적인 행정 업무에서 의미미한 효율성을 제공할 수 있음을 확인하였다. 생성형 AI 기반의 질의-응답 자동화 모델은 법령 및 자료 검색과 같은 고빈도 작업에서 공무원의 행정 부담을 경감하는 역할을 수행했다. 공공부문에서 법령과 규정, 판례 등을 빠르게 참조하는 능력은 의사결정 속도와 업무 정확성을 좌우하기 때문에 이러한 절차가 단축된다면 공무원들은 보다 중요한 의사결정과 대민 서비스에 집중할 수 있게 된다. 본 연구에서 제시된 방법론과 결과는 국유립 외의 다른 부처나 공

공 서비스 업무로 확장 가능성이 있다. 예를 들어 민원 처리에서는 AI를 통해 시민 질의에 대한 신속한 응답 체계를 구축할 수 있고 재난 대응에서는 실시간 데이터 분석으로 신속한 의사결정을 지원할 수 있다. 또한 정책 홍보 분야에서는 AI를 활용한 시민 맞춤형 정보 제공이 가능하다. 이러한 확장 가능성은 AI 기술이 반복적이고 단순한 업무를 넘어 다양한 공공행정 영역에서 디지털 전환을 촉진할 수 있는 잠재력을 시사한다. 물론 PoC 진행 과정에서 AI 도입으로 인한 능률 향상이 모든 공무원에게 동일하게 체감되지는 않는다는 한계 또한 확인되었다. 사용자에 따라 AI의 효율성이나 정확성에 대한 만족도가 다르게 나타났으며 이는 업무의 성격과 숙련도, 개별 사용자의 디지털 도구 친숙도, 필요로 하는 정보의 복잡성에 따라 달라지는 것으로 분석된다.

둘째, 본 연구는 생성형 AI의 공공부문 도입 시 성능 평가뿐만 아니라 사용자 경험과 수용성을 높이는 변화 관리 전략이 중요함을 강조한다. 공공부문은 관료제적 특성으로 인해 새로운 기술 도입에 있어 변화에 대한 저항이 클 뿐만 아니라 AI가 기존 행정 절차와 직무 체계에 미칠 잠재적 영향에 대한 불안감을 안고 있다. 이러한 맥락에서 정부 내 PoC 사업에서는 AI 도입의 목적과 기대되는 효과, 도입 과정과 변화 방향을 투명하고 일관되게 전달하는 커뮤니케이션 전략을 통해 AI가 공공부문에서 필요로 하는 효율성 증대에 어떻게 기여할 수 있는지 제시할 필요가 있다. 본 PoC에서는 AI 기술의 도입과 수용을 촉진하기 위해 조직 내 서로 다른 직급 간 의견을 수집하고 이를 시스템 개선에 반영하는 상향식 및 하향식 변화 관리 접근을 병행하였다. 결국 AI 등 혁신 기술이 공공부문에서 효과적으로 자리 잡기 위해서는 신뢰 형성과 조직 구성원의 수용을 우선하는 변화 관리 접근이 필요함을 시사한다.

셋째, 본 연구는 TOE 프레임워크를 통해 공공부문에서 성공적으로 AI를 도입하는데 필요한 주요 과제와 실행 방향을 제시하였다. 기술적으로 AI의 성능을 높이기 위해서는 데이터 공급과 품질 관리가 필수적인 영역으로 특히 공공부문과 같이 법령 및 민원 처리가 주를 이루는 분야에서는 데이터의 정확성과 최신성 확보가 무엇보다 중요함을 강조하였다. 또한 환경적 측면에서 개인정보 보호는 공공부문 AI 도입 시 반드시 검토해야 할 중요한 과제였다. AI가 민원 응답과 법령 정보를 학습하고 처리하는 과정에서 개인정보가 포함될 가능성이 높기 때문에 데이터 전처리 단계에서부터 민감 정보를 비식별화하여 수집된 데이터는 특정 개인을 식별할 수 없도록 제공하는 방식이 필요했다. 본 연구에서는 개인정보 보호 조치를 위해 원천 데이터를

수작업으로 검토하여 주민등록번호, 주소 등 개인식별 정보를 삭제한 후 제공하는 방식을 채택했다. 이를 통해 AI가 공공부문에 안정적으로 도입되기 위해 개인정보 보호법 등 관련 법령을 준수하고 시민들이 신뢰할 수 있는 데이터 관리 체계를 확보하는 것이 중요하다는 점을 강조한다.

위 연구 결과를 종합해 볼 때 향후 AI 기술이 공공행정의 전반적 혁신을 선도하기 위해서는 기술적 완성도뿐만 아니라 사용자 특성에 맞춘 맞춤형 설계와 변화 관리를 중점적으로 고려해야 한다는 점을 강조하며 나아가 법적·윤리적 요구에 부합하는 데이터 관리 체계의 확립이 필요함을 제안한다. AI 행정혁신 거버넌스는 기술 도입 단계에서부터 AI 도입 목표와 구현 전략을 명확히 설정하고 이를 지속적으로 평가·관리할 수 있는 체계를 포함해야 할 것이다. 예산 편성 단계에서는 AI 기반 솔루션 개발 및 운영을 위한 장기적 투자 계획이 필요하며 조직 재설계에서는 기술 도입으로 인해 변화하는 직무 구조와 역할을 재조정하고 디지털 역량 강화를 위한 교육 프로그램을 병행해야 한다. 이러한 제도적 기반은 공공부문에서 AI 기술의 성공적인 정착과 지속적인 발전을 위한 필수 요건이다.

본 연구는 공무원의 업무경감 목적의 특화 생성형 AI의 설계와 PoC 과정을 분석하였으나 몇 가지 한계점을 지니고 있다. 첫째, 본 연구의 PoC 사업은 국유림 이용이라는 업무 영역에 집중되어 있어 그 결과가 공공부문의 다른 행정 업무에 일관되게 적용될 수 있는지에 한계가 있다. 둘째, 본 연구에서 활용한 F1 스코어와 사용자 만족도 조사는 AI 도입의 효과를 일정 부분 입증하는 지표로 활용되었으나 실제 업무 환경에서의 장기적 효과성이나 경제적 타당성을 충분히 평가하지는 못했다. 셋째, 데이터 공급과 개인정보 보호 문제는 연구에서 다각도로 다루어졌으나 민감한 개인정보와 최신 법령 정보를 지속적으로 업데이트하고 관리하는 체계 구축이 완전하지 않아 실제 운영 단계에서의 위험성을 완전히 배제하기 어려운 상황이다. 위 한계점들을 바탕으로 본 연구는 후속 연구가 AI 도입의 범용성과 장기적 효과성, 데이터 보호 체계의 고도화를 포함하여 보다 확장된 관점에서 진행될 필요성을 제시한다.

REFERENCES

- [1] Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M.-F. (2020). Government by algorithm: Artificial intelligence in federal administrative agencies. Report submitted to the Administrative Conference of the United States. NYU School of Law.

- https://www.law.ox.ac.uk/sites/files/oxlaw/government_by_algorithm_acus_report.pdf
- [2] Ojo, A., Zeleti, F. A., & Mellouli, S. (2019). A realist perspective on AI-era public management. *ACM International Conference Proceeding Series*, 159-170. <https://doi.org/10.1145/3325112.3325261>
- [3] Mehr, H. (2017). Artificial intelligence for citizen services and government. Harvard Ash Center. https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf
- [4] Kankanhalli, A., Charalabidis, Y., & Mellouli, S. (2019). IoT and AI for smart government: A research agenda. *Government Information Quarterly*, 36(2), 304-309. <https://doi.org/10.1016/j.giq.2019.02.003>
- [5] Wirtz, J. (2020). Organizational ambidexterity: Cost-effective service excellence, service robots, and artificial intelligence. *Organizational Dynamics*, 49(3), 1-9. <https://doi.org/10.1016/j.orgdyn.2019.04.005>
- [6] Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—applications and challenges. *International Journal of Public Administration*, 42(7), 596-615. <https://doi.org/10.1080/01900692.2018.1498103>
- [7] Bright, J., Enock, F. E., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2024). Generative AI is already widespread in the public sector. arXiv preprint, arXiv:2401.01291. <https://doi.org/10.48550/arXiv.2401.01291>
- [8] F. Selten, M. Roberer, and S. Grimmelikhuijsen, “‘Just like I thought’: Street-level bureaucrats trust AI recommendations if they confirm their professional judgment,” *Public Administration Review*, vol. 83, no. 2, pp. 263-278, 2023. DOI: 10.1111/puar.13602
- [9] M. Carrasco, et al. (2023). *Generative AI for the Public Sector: From Opportunities to Value*. Boston Consulting Group
- [10] Gesk, T. S., & Leyer, M. (2022). Artificial intelligence in public services: When and why citizens accept its usage. ; *Government Information Quarterly*. <https://doi.org/10.1016/j.giq.2022.101704>
- [11] C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, L. Floridi (2018) Artificial intelligence and the “good society”: The US, EU, and UK approach, *Science and Engineering Ethics*. DOI: <https://doi.org/10.1007/s11948-017-9901-7>
- [12] J. Bryson, A. Winfield (2017). Standardizing ethical design for artificial intelligence and autonomous systems. DOI: 10.1109/MC.2017.154
- [13] Korea Institute of Science and Technology Policy. (2023). "Activation Strategies for Citizen-Created AI to Address Social Issues."
- [14] Korea Electronics and Telecommunications Research Institute. (2020). "Path to the Intelligent Information Society: Technology Development Roadmap 2035."
- [15] P. K. Agarwal. (2018). *Public Administration Challenges in the World of AI and Bots*. DOI: 10.1111/puar.12979
- [16] A. Androutsopoulou, N. Karacapilidis, E. Loukis, Y. Charalabidis (2019). Transforming the communication between citizens and government through AI-guided chatbots
- [17] K. de Fine Licht, J. de Fine Licht (2020). Artificial intelligence, transparency, and public decision-making. <https://doi.org/10.1007/s00146-020-00960-w>
- [18] M. Kuziemski, G. Misuraca, (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. <https://doi.org/10.1016/j.telpol.2020.101976>
- [19] R. Medaglia, J.R. Gil-Garcia, T.A. Pardo (2021). Artificial intelligence in government: Taking stock and moving forward, *Social Science Computer Review*. DOI: 10.1177/08944393211034087
- [20] Kang, O., Kim, H., & Park, J. (2023). "Development of a Customized Information Provision System Using Vector-Based Databases and Generative AI."
- [21] Son, H. (2022). Artificial Intelligence and Gender Technology: Focusing on the Iruda 1.0 Controversy.
- [22] Deshmukh, A. A., Sonar, S. D. B., & Ingole, R. V. (2023). Satellite image segmentation for forest fire risk detection using Gaussian mixture models. *IEEE Xplore*. <https://doi.org/10.1109/ICAIC56838.2023.10140399>
- [23] Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., et al. (2024). A survey on RAG meeting LLMs: Towards retrieval-augmented large language models. *Proceedings of the 30th ACM*. <https://doi.org/10.1145/3637528.3671470>
- [24] Marjan Ajevski, Kim Barker, Andrew Gilbert, Liz Hardie & Francine Ryan. (2023). “ChatGPT and the Future of Legal Education and Practice”, *The Law Teacher*. <https://doi.org/10.1080/03069400.2023.2207426>
- [25] Lee, C. (2023). "Concept of Artificial Intelligence and Application Cases in the Public Sector: Introduction to Major Approaches and Suggestions for Future Research Directions." DOI: 10.14329/apjis.2014.24.4.443
- [26] Ball, R., & Dal Pan, G. (2022). "Artificial Intelligence" for pharmacovigilance: Ready for prime time? *Drug Safety*, 45(5), 469-471. <https://doi.org/10.1007/s40264-022-01157-4>
- [27] Gahoe, K., & Kim, J. (2014). "A Study on Factors Influencing Big Data Adoption Intention: Focusing on Strategic Value Perception and the TOE (Technology Organizational Environment) Framework." *Asia Pacific Journal of Information Systems*. DOI: 10.14329/apjis.2014.24.4.443
- [28] J. Baker (2012) The technology-organization-environment framework. https://doi.org/10.1007/978-1-4419-6108-2_12
- [29] G.M. Aboelmaged (2014). Predicting e-readiness at firm-level: An analysis of technological, organizational and environmental (TOE) effects on e-maintenance readiness in manufacturing,

- International Journal of Information Management. <https://doi.org/10.1016/j.ijinfomgt.2014.05.002>
- [30] Ahn, M. J., & Chen, Y. C. (2022). Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. *Government Information Quarterly*, 39(1), 101679. <https://doi.org/10.1016/j.giq.2021.101679>
- [31] Popa, I., Cioc, M. M., Breazu, A., & Popa, C. F. (2024). Identifying sufficient and necessary competencies in the effective use of artificial intelligence technologies. *Amfiteatru Economic*, 26(63), 112-128. <https://doi.org/10.24818/EA/2024/63/112>
- [32] Daradkah, A., Awais, B., Telfah, E., & AlKhatib, F. (2024). Applying Artificial Intelligence (AI) and Digital Competencies to Drive Organizational Excellence. *Applied Mathematics & Information Sciences*, 18(2), 235-248. <https://doi.org/10.18576/amis/180203>
- [33] Henman, P. (2020). Improving public services using artificial intelligence: Possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration*, 42(4), 209-221. <https://doi.org/10.1080/23276665.2020.1816188>
- [34] Djeflal, C. (2020). Artificial intelligence and public governance: Normative guidelines for artificial intelligence in government and public administration. *Regulating artificial intelligence*. Springer. https://doi.org/10.1007/978-3-030-32361-5_12
- [35] Wilson, S. (1979). Explorations of the usefulness of case study evaluation, *Evaluation quarterly*, 3, 446-459. DOI: 10.1111/puar.12979
- [36] Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco, Jossey-Bass Publishers.
- [37] Stake, R. E. (2008). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln(Eds), *Strategies of qualitative inquiry*(3rd ed).
- [38] Yin, Robert K. (1989). *Case Study Research: Design and Methods*(Revised), Sage Publications.
- [39] Hong, J., & Won, H. (2017). "Exploring Qualitative Evaluation Methods for Education Programs." <https://dx-doi-org.libproxy.snu.ac.kr/10.13000/JFMSE.2017.29.1.306>
- [40] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. <https://doi.org/10.2307/249008>
- [41] DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9-30. <https://doi.org/10.1080/07421222.2003.11045748>
- [42] Rai, A., Lang, S. S., & Welker, R. B. (2002). Assessing the validity of IS success models: An empirical test and theoretical analysis. *Information Systems Research*, 13(1), 50-69. <https://doi.org/10.1287/isre.13.1.50.96>
- [43] KDnuggets. (2020). Performance evaluation metrics for classification: Precision, recall, F1-score & more. KDnuggets. Retrieved February 9, 2025, from <https://www.kdnuggets.com/2020/04/performance-evaluation-metrics-classification.html>
- [44] Wang, J., & Moon, J. (2021). "Analysis of Factors Influencing Public Officials' Acceptance of Intelligent Information Technology: A Comparison of Individual and Organizational-Level Factors."
- [45] Davenport & Prusak, (1998). "Working Knowledge: How Organizations Manage What They Know"
- [46] Osborne, David, and Peter Plastrik (1997). *Banishing Bureaucracy: The Five Strategies for Reinventing Government*. Reading MA: Addison-Wesley.
- [47] Castle, D. & Sir, M. (2001). *Organization Development: A Framework for Successful Information Technology Assimilation*. *Organization Developmental Journal*. DOI: 10.1080/08874417.2008.11646036
- [48] Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal Setting and Task Performance: 1969-1980. *Psychological Bulletin*. DOI: 10.1007/s11948-017-9901-7
- [49] Lewin, K. (1951). *Field Theory in Social Science: Selected Theoretical Papers* (Cartwright, D. (Ed.)). NY: Harper & Row.
- [50] Lim, S. (2005). "Organizational Context and Change Management Analysis in the Reform of Science and Technology Administrative Systems."
- [51] Issac-Henry, Kester, & Chris Painter. (1991). *The Management Challenge in Local Government-Emerging Themes and Trends*. *Local Government Studies*. DOI: 10.1007/s11948-017-9901-7

Authors



Eungjoon Kim received his B.S. in Public Administration from Korea University in 2015 and his M.S. from the Graduate School of Public Administration at Seoul National University.

He is currently pursuing a Ph.D. at the same institution. Since joining a central government agency, he has been working in the Big Data Team, focusing on data analytics and AI in public administration.