J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr

챗봇서비스 구현 모델의 보안요구사항 분석☆

Analysis of the Security Requirements of the Chatbot Service Implementation Model

조 규 민[1] 이 재 일[2] 신 동 규[1,3*]
Kyu-min Cho Jae-il Lee Dong-kyoo Shin

요 약

챗봇서비스는 AI서비스와 연계하여 다양한 분야에서 활용되고 있다. AI에 대한 보안 연구는 초기 단계이고, 이를 이용한 서비스 구현단계에서의 실질적인 보안 연구는 더욱 부족한 상황이다. 본 논문은 AI서비스와 연계된 챗봇서비스에 대한 보안요구사항을 분석한다. 먼저, 본 논문에서는 최근 발표된 AI보안에 대한 논문과 자료들을 분석한다. 시장에서 서비스가 제공되는 있는 챗봇서비스를 조사하여 일반적인 구현 모델을 정립한다. 구현 모델에는 챗봇관리시스템과 AI엔진이 포함된 5개의 구성요소가 포함되어 있다. 정립된 모델에 기반하여 챗봇서비스에 특화된 보호자산과 위협을 정리한다. 위협은 실제 운영중인 챗봇서비스 담당자 설문을 통해 챗봇서비스에 특화된 위협을 중심으로 정리한다. 10개의 주요 위협이 도출되었다. 정리된 위협에 대응하기 위해 필요한 보안 영역을 도출하였고, 영역별로 필요한 보안요구사항을 분석하였다. 이는 챗봇서비스 보안 수준을 검토하고 개선하는 과정에서 보안평가 기준으로 활용될 것이다.

☞ 주제어 : 챗봇, AI보안, 인공지능, 보안요구사항, 구현모델, 보안위협분석, 보안평가기준

ABSTRACT

Chatbot services are used in various fields in connection with AI services. Security research on AI is also in its infancy, but research on practical security in the service implementation stage using it is more insufficient. This paper analyzes the security requirements for chatbot services linked to AI services. First, the paper analyzes the recently published papers and articles on AI security. A general implementation model is established by investigating chatbot services provided in the market. The implementation model includes five components including a chatbot management system and an AI engine Based on the established model, the protection assets and threats specialized in Chatbot services are summarized. Threats are organized around threats specialized in chatbot services through a survey of chatbot service managers in operation. Ten major threats were drawn. It derived the necessary security areas to cope with the organized threats and analyzed the necessary security requirements for each area. This will be used as a security evaluation criterion in the process of reviewing and improving the security level of chatbot service.

☞ keyword : chatbot, AI security, artificial intelligence, security requirements, implementation model, security threat analysis, security evaluation criteria

1. 서 론

AI기술이 발전함에 따라 금융, 법률, 의료, 보안 등 많은 분야에서 AI가 활용되고 있고, 검색, 채팅, 상품 추천, 내부 의사결정 지원, 데이터 분석 등 다양한 방식으로 AI 서비스가 제공되고 있다. 사용 분야가 확대되고, 제공되는 서비스 방식이 다양화됨에 따라 AI기술의 역기능에 대한 연구도 활발해지고 있다. AI알고리즘의 공정성과 윤리성, AI기술 자체에 대한 보안성 등에 대한 연구가 진행되고 있지만, 많은 부분이 초기 단계이고, AI기술의 확대에 비해 역기능에 대한 연구는 부족한 상황이다. 특히, 알고리즘의 공정성과 윤리성, AI기술의 취약점 등에 대한

1 Department of Computer Engineering, Sejong University, Seoul, 05006, Korea.
2 CISO, Smilegate, Gyeonggi-do, 13493, Korea
3 Department of Convergence Engineering for Intelligent Drones, Sejong University, Seoul, 05006, Korea.
* Corresponding author (shindk@sejong.ac.kr)
[Received 9 October 2023, Reviewed 24 October 2023(R2 15 November 2023), Accepted 4 December 2023]
☆ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1F1A1074773). 이 논문은 2023년도 세종대학교 교내연구비 지원에 의한 논문임

연구는 활발하지만, AI기반 서비스에 대한 실재 구현단계의 보안 연구는 부족한 상황이다. 본 논문은 AI서비스 중 가장 활성화된 챗봇서비스 구현모델을 정립하고, 정립된 모델에 기반하여 보안요구사항을 분석하고자 한다.

분석과정은 다음과 같다. 실재 운영 중인 챗봇서비스 현황을 조사하여 챗봇서비스 구성요소를 정리하고, 이를 기반으로 구현 모델을 정립한다. 정립된 모델을 기반으로 보호자산을 도출하여 목록화한다. 자산에 대한 위협을 분석하기 위해 챗봇서비스를 운영하고 있는 담당자 설문조사를 시행한다. 설문 조사에 기반하여 자산별 위협을 도출한다. 도출된 위협에 대응하는 보안영역을 매칭하고, 보안영역별 보안요구사항을 분석한다. 영역별 보안요구사항은 관련 평가기준, 가이드 등을 참고한다.

(표 1) 관련 연구 현황
(Table 1) Status of related research

관련연구	주요내용
AI Risk Management Framework 1.0	AI기술의 위험을 제거하기 위해 고려해야할 특성을 7가지로 분류
An Artificial Intelligence Security Framework	ISO 표준에서 제시된 AI 생명주기에 기반한 단계별 보안 위협을 정리하고 보안 프레임워크 정립
인공지능(AI) 개인정보보호 자율점검표	AI기술 및 서비스 개발 시 필요한 개인정보보호 및 정보보안 자율점검 항목을 제시
신뢰할 수 있는 인공지능 개발 안내서	AI기술 및 서비스 개발 시, 자율적으로 점검 가능한 16개의 개발 요구사항과 59개 정성적·정량적 검증 항목을 제시
Getting to know-and manage-your biggest AI risks	AI 위험에 대한 유형을 분류하고 발생 가능 영역을 데이터와 환경요소까지 확대하여 제시
챗GPT 등 생성형 AI 활용 보안 가이드라인	생성형 AI기술 개요 및 해외동향, 보안 위협 사례, 기술 사용 가이드라인, 정보화 구축방안 및 보안대책 등을 포함

2. 관련 연구

NIST에서 2023년 1월에 AI기술의 위험관리를 위한 기술 문서 'Artificial Intelligence Risk Management Framework 1.0'[1]을 발행하였다. 이 문서에서는 AI기술의 위험을 제거하기 위해 고려해야할 특성을 유효성과 신뢰성(Valid & Reliable), 책임성과 투명성(Accountable & Transparent), 안전성(Safe), 보안성과 회복탄력성(Secure & Resilient), 비편향적 공정성(Fair-with Harmful Bias managed), 설명가능성(Explainable & Interpretable), 프라이버시(Privacy-Enhanced) 등 7가지로 분류하고 있다. 이 특성들은 각각의 주제로 다양하게 연구되고 있지만, 보안성과 회복탄력성 부분은 연구가 부족한 상황이다.

2021년 발표된 'An Artificial Intelligence Security Framework'[2] 논문에서는 ISO가 제시하는 'Artificial Intelligence System Lifecycle Process'에 기반하여 단계별 보안위험을 정리하여 제시하고 있고, AI시스템이 갖추어야 할 보안 프레임워크를 6개의 보안목표(security goals), 5개의 단계별 보안 역량(security capability), 14개의 보안기술(security technique) 등 3개 계층으로 제시하고 있다.

개인정보보호위원회는 '인공지능(AI) 개인정보보호 자율점검표'[3]를 발표하였다. 자율점검표는 AI 기술·서비스 개발 시 고려해야 할 개인정보보호에 관한 의무준수 사항과 기술·서비스 환경에 적합한 자율적인 개인정보보호 이행·점검에 필요한 사항을 포함하고 있다.

과학기술정보통신부와 한국정보통신기술협회는 '신뢰할 수 있는 인공지능 개발 안내서(2022)'[4]를 발표하였다. 안내서는 AI기술 및 서비스 개발 시, 자율적으로 점검 가능한 16개의 개발 요구사항과 59개 정성적·정량적 검증 항목을 제시하고 있다. 일반적인 수준의 제시로 실제 적용을 위해서는 상세한 항목을 도출할 필요가 있다.

맥킨지의 2021년 보고서 'Getting to know-and manage-your biggest AI risks'[5]에서는 AI 위험에 대한 유형을 분류하고 발생 가능 영역을 제시하고 있다. AI 위험 유형은 개인정보보호(Privacy), 보안성(Security), 공정성(Fairness), 투명성과 설명가능성(Transparency and explaniability), 안전성과 성능(Safety and performance), 제3자 위협(Third-party risks) 등 6가지를 제시하고 있다. 발생가능 영역은 데이터(Data), 모델선택과 학습(Model selection and training), 배포와 인프라(Deployment and infrastructure), 계약과 보험(Contracts and insurance), 법률과 규제(Legal and regulatory), 조직과 문화(Organization & culture) 등 6가지 영역을 제시하고 있다. 이 보고서는 조직관점의 AI 위험을 정리하고 있지만, 위험유형들은 AI 시스템에서 고려해야할 사항과 직접 관련되어 있으며, 위험 발생 가능 영역에서는 특히, 데이터와 모델선택과 학습 부분이 시스템 구현 시 필수적으로 고려되어야 할 요소이다.

한국의 국가정보원은 2023년 6월에 '챗GPT 등 생성형 AI 활용 보안 가이드라인'[6]을 발표하였다. 보고서는 생성형 AI기술 개요 및 해외동향을 정리하고, 보안위협 사례, 기술 사용 가이드라인, 정보화 구축방안 및 보안대책 등을 포함하고 있다. 대표적인 보안 위협으로 잘못된 정

보, AI모델 악용, 유사 AI서비스 빙자, 데이터 유출, 플러그인 취약점, 확장 프로그램 취약점, API 취약점 등을 제시하고 있다. 생성형 인공지능 기술 사용 시 안전하게 사용할 수 있는 가이드라인과 생성형 AI를 활용한 정보화 사업 구축 방안 및 보안대책 등을 설명하고 있다. 가이드라인은 전반적인 사항을 제시하고 있지만, 구체적인 시스템 구현 시 필요한 사항까지 포함하고 있지 않다.

3. 챗봇서비스 구현 모델의 정립

3.1 챗봇서비스 구현 현황

챗봇서비스 구현 모델을 정립하기 위해 현재 챗봇서비스를 제공하고 있는 21개사의 현황을 조사하였다. 보안담당자 설문을 통해 서비스 목적, 사용자 챗봇 제공방식, 연계된 AI서비스, 인프라 운영방식, 보안 위협 등을 파악하였다. 보안 위협 부분은 4.2에서 정리하였다.

(표 2) 설문 조사 대상 및 방법
(Table 2) Survey targets and methods

구분	내용
조사 시점	2022년 12월~23년 1월
대상 기업	은행, 보험 등 21개
대상자	보안담당 실무자
방법	설문지를 통한 조사

서비스 목적은 단순 상담 처리 50%, 상담 및 업무연계 처리 27%, 내부업무 처리 18% 등으로 조사되었다. 목적과 관계없이 질문 처리를 위한 자연어처리 기능을 AI서비스와 연계하여 구현하고 있다. 업무처리와 연계 비율은 27%로 조사되었지만, 현재 연계하지 않고 있는 기업 중 50% 이상이 향후 연계 계획이 있다고 답변하였다.

사용자 챗봇 제공방식은 앱 38%, 웹서비스 38%, SNS나 AI스피커 등 기타 24%로 조사되었다. AI서비스 구축 방식은 온프레미스 38%, 사설 클라우드 19%, 사설 클라우드와 외부 클라우드 AI서비스 연계 14%, 외부 클라우드 서비스 활용 29%로 조사되었다. 온프레미스형도 내부적으로 AI기능을 포함하고 있으며, 외부 클라우드 서비스를 활용하는 경우에도 내부 연계를 위한 최소한의 연계기능을 포함하고 있다.

조사결과를 기능적으로 정리하면, 사용자 모듈, 챗봇관리시스템, 자연어처리를 위한 AI엔진, 업무처리시스템, 연계 모듈 5가지로 정리할 수 있다. 각 기능은 통합적으로 구축될 수도 있고, 업무처리시스템은 제외될 수도 있다.

3.2 챗봇서비스 구성 요소

3.2.1 사용자 모듈

사용자 모듈은 웹이나 앱의 형태로 사용자가 사용할 수 있는 인터페이스를 제공하고, 사용자의 질문을 받아 연계 모듈을 통해 챗봇관리시스템으로 전송하고, 처리된 답변을 받아 사용자에게 알려주는 역할을 담당한다. 사용자의 질문을 발화문이라고 한다. 필요한 경우, 사용자의 업무처리 요구를 챗봇관리시스템이나 업무처리시스템에 요청하여 처리하고 결과를 사용자에게 알려주는 역할을 수행하기도 한다.

3.2.2 챗봇관리시스템

챗봇관리시스템은 AI엔진과 연계하여 사용자의 질문을 처리한다. 질문을 AI엔진에 전달하고, 답변을 받아 사용자에게 제시한다. 필요한 데이터를 학습데이터로 관리하고, 사용자 관리, 인증 등 필요한 관리 기능을 제공한다. 구성방식에 따라 자연어처리 등 AI 역할을 일부 수행할 수도 있으며, 학습데이터 및 사용자 질문/답변 등의 관리 기능만을 제공할 수도 있다. 업무처리와 연계가 필요할 경우 업무처리시스템과 연계하여 필요한 서비스를 처리할 수도 있다.

3.2.3 AI엔진

AI엔진은 챗봇서비스에 필요한 AI 기능을 제공한다. 외부 또는 내부 클라우드로 운영되는 것이 일반적이고, 경우에 따라 온프레미스 형태로 구성될 수도 있다. 챗봇서비스에서 제공되는 대표적인 AI기능은 자연어처리 기능이며, 서비스가 발전되면 생성형 AI 기능과 연계될 수도 있다. 질문과 답변 등을 이용하여 학습데이터 생성하고, 이를 학습하여 개인별 적합한 서비스를 제공 할 수도 있다.

3.2.4 업무처리시스템

업무처리시스템은 서비스 제공기업이 고객을 위한 업무를 처리하는 시스템이다. 은행이라면 이체, 계좌조회 등의 기능이 제공될 수 있고, 여행서비스라면 예약 조회/변경/취소, 여행 상품 조회 등 다양한 서비스를 제공한다. 단순 상담만을 위한 챗봇의 경우, 기능이 제공되지

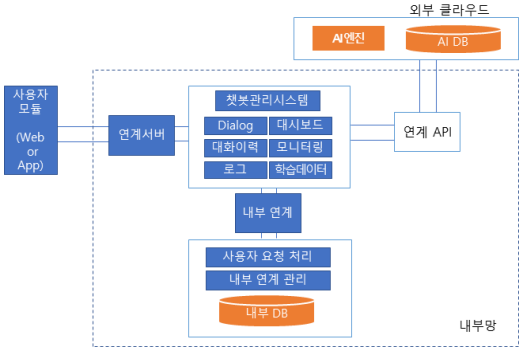
않을 수도 있지만, 챗봇서비스는 AI기능과 함께 제공되는 경우가 많고, 업무처리 서비스도로 연계되는 경우도 많다.

3.2.5 연계 모듈

챗봇서비스의 구성요소들은 독립적인 시스템으로 구성할 수도 있고, 복합적인 기능을 통합적으로 구성할 수도 있다. 일반적으로 각 요소는 운영 효율성이나 외부의 전문적인 서비스를 활용하기 위해 연계 모듈을 포함하여 독립적으로 구성된다. 독립적인 구성요소를 연계하기 위해 최소한의 연계 모듈이 필요하다.

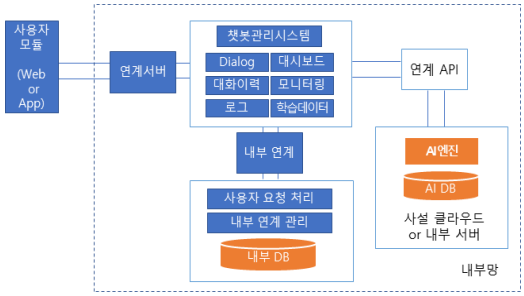
3.3 챗봇서비스 구현 모델

3.2에서 정리된 구성요소에서 가장 중요한 요소는 챗봇관리시스템과 AI엔진이다. 챗봇서비스를 제공하기 위한 최소한의 요소로 사용자 모듈과 챗봇관리시스템이 필요하고, AI서비스를 제공하기 위해서는 AI엔진이 연계되어야 한다. AI엔진은 일반적으로 외부의 전문시스템을 활용한다. 챗봇관리시스템은 일반적으로 내부에 구축되고, AI엔진은 외부나 내부에 구축할 수 있으며, 클라우드 기반으로 구축되는 경우가 많다. 외부에 구축된 AI엔진과 연계된 모델을 도식화하면 그림 1과 같다.



(그림 1) 외부 엔진 기반 챗봇서비스 구현 모델
(Figure 1) External Engine-based Chatbot Service Implementation Model

내부에 AI엔진을 구축하는 방식은 사설 클라우드와 온프레미스 방식으로 나눌 수 있다. 상담 중심의 간단한 챗봇은 온프레미스로 구축되지만, 업무연계나 사용자 특화 서비스 등 기능이 복잡해지면 클라우드 방식으로 구축된다. 내부 구현 모델을 도식화하면 그림 2와 같다.



(그림 2) 내부 엔진 기반 챗봇서비스 구현 모델
(Figure 2) Internal Engine-based Chatbot Service Implementation Model

4. 보호자산과 위협에 대한 분석

4.1 보호자산

보호자산은 소프트웨어, 하드웨어, 데이터 자산으로 분류할 수 있다. 구성요소별로 소프트웨어. 하드웨어, 데이터 관점에서 구성에 필요한 자산을 식별한다. 기존의 서비스를 위해 구축된 요소는 제외되지만, 챗봇서비스를 추가하지 위해 소프트웨어에 중요한 변경이 발생했다면 포함하였다.

각 구성요소에 사용자 모듈, 챗봇관리시스템 S/W, AI엔진 S/W, 연계 S/W, 업무처리 S/W 등 모두 소프트웨어가 구현되어 포함되어 있다. 서비스 운영에 필요한 운영체제, 미들웨어, 클라우드 가상화 모듈 등도 고려될 수 있다. 본 논문에서는 기존에 구축되어 있던 업무처리 S/W와 기본 시스템 운영에 필요한 소프트웨어는 보호자산에서 제외하고, 4가지 소프트웨어 자산을 포함하였다.

데이터 자산은 AI기능에 필요한 학습데이터, AI시스템 데이터와 챗봇관리스템 데이터, 사용자 데이터 등이 고려되어야 한다. 업무처리시스템 운영에 필요한 시스템데이터와 클라우드 운영 데이터도 포함된다. 본 논문에서는 챗봇서비스를 위해 필요한 데이터만 고려함으로 학습데이터, 챗봇관리시스템데이터, 새로운 서비스 과정에서 생성된 로그를 자산으로 분류한다.

하드웨어는 스마트폰이나 PC와 같은 사용자모듈이 운영되는 하드웨어, 시스템 S/W가 운영되는 서버, AI엔진을 위한 클라우드 인프라, 네트워크 운영에 필요한 네트워크 장비 등이 포함된다. 본 논문에서는 기존에 구축된 업무처리시스템 인프라는 제외하고, 사용자 단말, 챗봇 인프라, 클라우드 인프라, 연계 모듈 인프라를 보호자산

으로 분류한다.

최종적으로 분류된 보호자산을 아래 표 3으로 정리하였다.

(Table 3) List of Protected assets

분류	자산 항목	설명
소프트웨어	사용자 모듈	사용자용 웹 또는 앱
	챗봇관리시스템	챗봇서비스를 제공하기 위해 구현된 프로그램
	AI엔진	챗봇서비스를 위해 필요한 AI기능을 제공하는 소프트웨어
	연계 S/W	각 구성요소 간 연계를 위해 구현된 프로그램
데이터	학습데이터	챗봇서비스를 제공하는 과정에서 생성되어 AI엔진의 학습에 활용되거나 사용자별 서비스 제공을 위해 관리되는 데이터
	챗봇관리시스템데이터	챗봇관리시스템 운영에 필요한 시스템 데이터
	로그	챗봇관리시스템 운영과정에서 생성된 로그
하드웨어	사용자 단말	사용자 모듈이 운영되는 하드웨어
	챗봇 인프라	챗봇관리시스템이 운영되는 하드웨어
	클라우드 인프라	AI엔진이 운영되는 클라우드 인프라
	연계 모듈 인프라	연계 모듈이 운영되는 하드웨어

4.2 보안 위협

4.2.1 위협 조사

챗봇서비스에 특화된 보안 위협은 설문조사 결과를 활용하여 도출하였다. 챗봇서비스 구현에 따라 추가된 위협만을 포함하였고, 일반적인 서비스 구현 및 기존 위협을 제외하였다.

AI서비스와 연계된 학습데이터와 사용자들이 입력하는 개인정보에 대한 위협이 가장 중요한 위협으로 조사되었다. AI서비스가 클라우드 기반으로 제공됨에 따라 클라우드 인프라에 대한 위협과 외부 API 연계가 확대되는 과정에서 API 및 오픈소스에 위협이 추가로 도출되었다. 조사된 위협을 정리하면 표 4과 같다.

(표 4) 챗봇서비스 보안 위협

(Table 4) Chatbot service security threats

분류	위협
데이터 관련 위협	학습데이터 오염
	학습데이터 공격
	개인정보 과다 수집
	개인정보 유·노출
소프트웨어 관련 위협	외부 프로그램 취약점
	오픈소스 취약점
	API 오류 증가
인프라 및 운영 관련 위협	클라우드 인프라 취약점
	제3자 위탁 관리 위협
	관리체계 범위 확대 미흡

4.2.2 데이터 관련 위협

학습데이터 오염 : 사용자가 요청한 질문(발화문) 등을 내부 DB나 외부 클라우드 DB에 저장하고, 학습데이터로 활용하게 된다. 학습데이터는 별도의 절차를 마련하여 관리해야 한다. 관리절차가 미흡할 경우 데이터가 오염될 수 있다.

학습데이터 공격 : 학습데이터에 대한 잘못된 입력, 적대적 예제 공격 등이 발생할 수 있다. 공격자들은 AI엔진을 혼란하게 만들거나 오류를 유도하는 다양한 방법에 관해 연구하고 있다.

개인정보 과다 수집 : 사용자는 질문 등 입력할 때, 의도하지 않은 개인정보를 입력할 수 있다. 사용자의 과도한 개인정보 제공 등으로 수집되지 않아야 할 개인정보가 수집될 수 있다.

개인정보 유·노출 : 챗봇서비스 과정에서 다양한 서비스 제공을 위한 개인정보 수집이 발생한다. 수집된 개인정보는 전송되거나 저장될 때 유출 또는 노출될 수 있다.

4.2.3 소프트웨어 관련 위협

외부 프로그램 취약점 : 챗봇서비스는 AI서비스 연계 등 외부 프로그램을 활용하여 구축된다. 외부 프로그램 취약점이 전이될 수 있다. 외부 프로그램을 도입 시, 철저한 보안 검증이 요구된다.

오픈소스 취약점 : 챗봇서비스를 효율적으로 개발·운영하기 위해 오픈소스 이용이 증가하고 있어, 검증되지 않은 오픈소스 이용에 따른 취약점이 발생할 수 있다. 오

픈소스에 대한 취약점 분석이 필요하다.

API 오류 증가 : 연계 API가 점차 증가하고 있어, API에 대한 비인가 접근, 의도치 않은 응답 출력 등의 위협이 발생할 수 있다. 업무처리시스템과 연계된 API 오류는 기업의 내부 시스템에 영향을 줄 수 있다.

4.2.4 인프라 및 운영 관련 위협

클라우드 인프라 취약점 : 자연어처리 등 AI서비스는 대부분 외부 클라우드 기반으로 제공되고, 다른 클라우드 기반 서비스도 늘어나고 있다. 외부 클라우드에 대한 비인가 접근, 관리 미흡 등 위협이 발생할 수 있다.

제3자 위탁 관리 위협 : 학습데이터 관리, 챗봇서비스 운영 등 위탁 업무가 늘어남에 따라 제3자 관리 위협도 증가한다. 개인정보에 대한 위탁 시에도 위협은 발생한다.

관리체계 범위 확대 미흡 : 챗봇서비스 등 사용자에 대한 제공 서비스가 늘어남에 따라 전사적 보안관리체계 운영범위가 늘어난다. 다양한 서비스 확대에 따라 보안관리체계 범위를 정확하게 책정하기 어렵고, 누락이 발생할 수 있다.

5. 보안요구사항의 도출

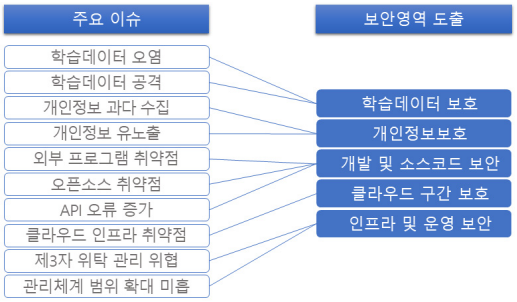
5.1 보안 위협과 보안 영역 매핑

보안요구사항을 효율적으로 분석하기 위해 도출된 보안 위협을 보안 영역으로 매핑하여 분석하였다.

학습데이터 보호 영역은 학습데이터과 관련된 2가지 위협에 대응하고, 개인정보보호 영역은 개인정보에 대한 2가지 위협에 대한 보안요구사항을 정리하는 영역이다. 도출된 위협에 대한 보안대책을 모두 포함하기 위해 클라우드구간 보호, 개발 및 소스코드 보안, 인프라 및 운영 보안 영역이 도출되었다.

정리된 보안영역별로 위협에 대응하기 위한 보안요구사항을 분석하였다. 분석된 보안요구사항은 다음절부터 정리되어 있다. 각 절에는 보안요구사항 도출을 위한 참고자료가 있는 경우 명시하였으며, 항목은 챗봇서비스에 적합하도록 용어와 관점을 수정하여 정리하였다.

위협과 보안 영역을 매핑하면 그림 3과 같다.



(그림 3) 보안위협과 보안 영역 매핑
(Figure 3) Mapping Security Threats and Security Zones

5.2 학습데이터 보호

학습데이터 보호는 AI서비스 기반 챗봇서비스에 특화된 보안 영역으로 AI의 보안성과 연계하여 분석하였다. 필요한 8가지 보안요구사항을 정리하면 다음과 같다.

(Table 5) Security Requirements for Learning Date Protection

보안요구사항	세부 내용
학습데이터 관리절차 수립	학습이 안전하고 정확하게 수행될 수 있도록 담당자 지정, 학습 통제 절차 등을 수립 및 이행하고, 관련 법령 및 규제를 준수할 수 있도록 관리절차를 수립
학습데이터 신뢰성 보장	학습데이터의 출처에 대한 신뢰성 평가 기준 수립 및 평가 수행
중요 파일, DB에 대한 접근통제	챗봇 관리시스템 관련 중요 파일, DB에 대해 접근통제 방안 마련 및 구현
중요 파일 무결성 검증	챗봇서비스에 이용되는 AI 모델과 학습데이터 등 주요 파일에 대한 위·변조되는 방지 및 검증 방안
학습데이터의 건전성 보장	학습데이터에 대한 데이터 포이즈닝(data poisoning), 적대적 예제(adversarial example) 공격 등의 예방 대책 수립
발화문 원문에 대한 안전한 저장	챗봇 입력창을 통해 입력되는 발화문 원문을 저장할 때, 안전한 방안으로 저장
학습데이터 비인가 접속 탐지·차단	학습데이터의 오용 및 악용 가능성을 예방하기 위해 비인가 접속을 탐지 및 차단하는 기능 구현
학습데이터 업데이트 이력 관리	챗봇서비스에 이용되는 인공지능 시스템의 의사결정에 대해 추적 및 대응할 수 있도록 학습데이터 업데이트 이력 관리

5.3 개인정보보호

개인정보보호는 개인정보를 이용하는 챗봇서비스에 필수적인 보안 영역이다. 챗봇서비스에서 다루는 개인정보는 서비스 제공을 위한 식별 및 인증 정보와 대화에서 사용자에 의해 입력되는 개인정보로 구별할 수 있다. 식별 및 인증정보는 시스템에서 관리되어야 하며, 대화에 입력되는 개인정보는 서비스 정책에 따라 제한되거나 관리될 필요가 있다.

관련 법규와 '개인정보 안전성 확보조치 기준[7]에서는 개인정보 생명주기 전반에 관해 규정하고 있다. 챗봇서비스를 도입하는 기업들은 이미 고객서비스를 제공하고 있고, 개인정보 관리체계를 가지고 있다. 따라서, 본 논문에서는 관련 법규와 자료에서 명시된 개인정보보호 항목들 중, 챗봇서비스 구현 과정에서 적용되어야 하는 항목만을 선별하여 적용하였다. 질문 과정에서 입력되는 불필요한 개인정보를 제한하기 위한 3개 항목, 챗봇관리 시스템에서 관리하는 개인정보 암호화 관련 2개 항목, 챗봇관리시스템 등에 추가적으로 안전성 확보조치를 적용하기 위한 1개 항목을 도출하였다. 개인정보가 포함된 파일에 대한 관리가 요구되지만, 학습데이터 보안요구사항에 이미 포함되어 있어 별도록 추가하지 않았다. 도출된 6가지 보안요구사항을 정리하면 다음과 같다.

5.4 클라우드 구간 보호

AI엔진이 주로 운영되는 클라우드 관련 보안 영역으로 '금융분야 클라우드컴퓨팅 이용 가이드'[8]을 참고하여 분석하였다. 가이드에 제시된 항목 중, AI엔진이 운영되는 클라우드 시스템과 연동 구간에 적용되는 항목들만 선별하여 도출하였다.

(Table 6) Security Requirements for Privacy

보안요구사항	세부 내용
개인정보 입력 제한 안내	챗봇서비스에서 입력 시, 개인정보를 입력하지 않도록 사용자에게 안내하거나 입력을 제한
개인정보 수집 제한 조치	사용자가 입력창에 불필요한 개인정보를 입력하는 경우, 개인정보를 삭제하거나 마스킹 하는 등 수집 제한 조치 시행
개인정보 파일 탐지 및 차단	사용자가 입력한 파일 등에 대한 개인정보 탐지 및 차단 기능
안전성 확보조치 이행 확인	개인정보 관련 법과 규정에서 요구하는 안전성 확보조치에 대한 사전 점검 및 정기적 검토
중요정보 암호화	챗봇서비스에서 관리하는 중요정보 및 개인정보에 대해 암호화 대상을 식별하고 암호화
개인정보 전송 시 암호화	망을 통해 개인정보 전송 시, 전송 암호화 기능 구현

클라우드 구간에 구축된 AI엔진에 대한 내부 시스템 수준의 보호 대책, 클라우드서비스 운영을 위한 관리자 및 이용자 계정관리, 클라우드 연결 과정에서 정보 유출을 방지하기 위한 접근 통제, 클라우드 접근에 필요한 키에 대한 관리와 접근 기록 관리 등 5개 항목을 도출하였다. 클라우드 구간 보호에 필요한 5가지 보안요구사항을 정리하면 다음과 같다.

(Table 7) Security Requirements for Cloud Section

보안요구사항	세부 내용
클라우드 구간 시스템에 대한 보호대책	클라우드서비스 제공자 구간에 위치한 정보처리시스템에 대한 내부망에 준하는 보호대책 수립 및 이행
클라우드서비스 구간 계정관리	클라우드서비스 관리자 및 이용자에 계정에 대한 적절한 권한 및 통제 관리
클라우드 연결구간에 대한 접근통제	클라우드 연결 시, IP 제한 등 접근통제 수단을 마련하여 시행
클라우드 접근 기록 관리 및 검토	클라우드 구간 접근 시, 접속 내역을 기록하고 주기적으로 검토
클라우드 접근키 관리	클라우드 접근키는 필요 시 복구가 가능하도록 별도의 안전한 장소에 보관하고, 접근 권한을 최소화

5.5 개발 및 소스코드 보안

개발 및 소스코드 보안에 필요한 8가지 보안요구사항을 정리하면 다음과 같다.

(Table 8) Security Requirements for Development and Source Code

보안요구사항	세부 내용
사용자 모듈 보안 점검	자체 개발 사용자 모듈에 대한 소스코드 점검 및 보안취약점 평가
승인된 오픈소스 사용	오픈소스를 사용하는 경우 보안 점검 후, 승인된 오픈소스만 사용

보안요구사항	세부 내용
챗봇 관리시스템 보안 점검	챗봇 관리시스템에 대한 소스코드 점검 및 보안취약점 평가
오픈소스 소프트웨어 보안취약점 평가	오픈소스 소프트웨어에 보안취약점 등을 정기적으로 평가하고 문제점 조치
API 목록 관리	사용하는 API의 목록을 유형별로 분류하여 식별 및 관리
비정상 API 호출 탐지차단	챗봇서비스 사용자 모듈에서 입력창을 통한 비정상적인 API 탐지 및 차단
API 변경 관리	API 변경 시, 변경 내역을 관리할 수 있도록 변경 관리절차 수립 및 이행
안전한 API 키 관리	API 토큰 또는 암호키 등은 필요 시 복구가 가능하도록 별도의 안전한 장소에 보관하고, 접근 권한을 최소화

5.6 인프라 및 운영 보안

챗봇서비스 운영 기업들은 인프라 및 운영 보안 전반에 대한 체계를 갖추고 있다. 본 논문에서 다루는 인프라 및 운영 보안영역은 챗봇서비스 구현과정에서 추가된 인프라 관련 운영 부분만을 다룬다. 기업의 보안 체계에 대한 보안 항목을 도출하기 위해 정보보호관리체계(Information Security Management System)의 요구사항을 제시하는 'ISMS-P 세부점검항목'[9]을 참고하였다.

보안 정책, 자산 식별 등 관리체계 유지 과정에 챗봇서비스 관련 요구사항을 추가하기 위한 3개 항목, 챗봇관리시스템에 대한 통제를 위해 필요한 접근 통제를 관련 2개 항목, 챗봇서비스에 대한 접근 기록 및 로그 관련 2개 항목, 추가된 구성요소에 대한 백업 및 복구 항목이 도출되었다. 챗봇서비스는 구현과정에서 AI엔진 연계, 외부API 사용 등 외부 위탁 요구가 증가하여, 외부 위탁 관리에 필요한 2개 항목이 추가되었다. 최종적으로 도출된 10가지 보안요구사항을 정리하면 다음과 같다.

(Table 9) Security Requirements for Infrastructure and Operation

보안요구사항	세부 내용
정보보호와 개인정보보호 정책 및 지침 변경	챗봇서비스와 관련하여 정보보호와 개인정보보호 정책 및 지침에 관련 보안대책 반영하여 변경
챗봇서비스 관련 위험관리 정책 마련	사용자 개인정보 확대, AI서비스 구간 확대 등에 따른 위험을 평가하고, 위험관리 정책 수립 및 이행
정보자산 식별 및 관리	챗봇 관리시스템, 챗봇 인프라 등 챗봇서비스를 구성하는 주요 정보자산을 식별하여 목록화하고 관리
챗봇 관리시스템 접근통제	챗봇관리시스템 관리자 접속 시, 접근통제 정책 수립 및 이행
챗봇관리시스템 외부 접근 제한	챗봇관리시스템에 대한 외부망에서 내부망으로의 접속은 원칙적으로 금지하고, 장애대응 등 불가피한 사유인 경우, 별도의 보안대책을 수립 및 이행
접근통제 정책 등록/변경/삭제 명세 기록 및 검토	챗봇 인프라와 챗봇 관리시스템 관련 중요 파일, DB의 접근통제 정책 등록/변경/삭제 시, 명세를 기록하고 정기적으로 토
필수 구성요소 백업 및 복구 대책	챗봇 관리시스템 및 챗봇 인프라 등 챗봇서비스 구성요소에 대하여 백업 및 복구절차를 수립하고 이행
감사 로그 생성 및 관리	챗봇서비스와 관련된 로그를 생성하여 관련 법규, 정보보호 정책 및 지침에서 명시된 기간까지 보관하고, 주기적 검토
외부자 계약 시 보안	AI서비스 및 클라우드 서비스 제공자, 관련 업무 위탁에 따른 계약 체결 시, 정보보호 요구사항 반영
외부자 보안 이행 관리	AI서비스 및 클라우드 서비스 제공자, 관련 업무 수탁자에 대한 정기적 보안 점검 및 관리

6. 결 론

본 논문은 자연어처리 등 AI서비스와 연계된 챗봇서비스에 대한 보안요구사항을 분석하였다. 챗봇서비스 현황을 분석하여 다섯 가지 필수 구성요소를 정리하였고, 구성요소를 기반으로 챗봇서비스 구현 모델을 정립하였다. 정립된 모델이 기반하여 보호 자산과 보안 위협을 분석하였다. 보안 위협에 대응하는 보안영역을 매핑하여 영역별로 필요한 보안요구사항을 도출하였다. 최종적으로 5개 보안영역과 37개 보안요구사항을 도출하였다. 본 논문에서 제시된 보안요구사항은 챗봇서비스 구현과정에서 필수적으로 요구되는 보안요구사항만을 포함하고 있으며, 기존 시스템에 필요한 보안요구사항은 별도로 다루지 않았다.

본 논문에서 도출된 보안요구사항은 챗봇서비스를 신규로 도입하고자 하는 IT 및 보안 담당자들이 활용할 수 있다. 챗봇서비스 구현과정에서 기업의 보안관리체계에 추가되어야 하는 관리적 보안 요구와 개발과정에서 구현되어야 하는 보안 기능 등을 모두 포함하고 있다. 챗봇서비스 보안 수준을 검토하고 개선하는 과정에서 보안평가 기준으로 활용될 수도 있다.

향후, 실제 서비스 운영 과정에서 발생할 공격과 보안

사고에서 알려지는 보안 위협, 기술의 발전에 따라 드러나는 추가 보안 위협 등이 알려지면 연구는 보완될 필요가 있다. AI는 다양한 방식으로 적용될 수 있고, AI서비스는 기존 IT와 결합하여 새로운 분야로 확장될 것이다. 본 연구는 이런 다양한 IT서비스 구현과정에서 추가 연구로 확장될 필요가 있다.

참고문헌(Reference)

[1] "Artificial Intelligence Risk Management Framework," NIST, pp.12-18, 2023. https://doi.org/10.6028/NIST.AI.100-1

[2] Huiyun Jing & Wei Wei & Chuan Zhou, "An Artificial Intelligence Security Framework," Journal of Physics: Conference Series, 1948, 012004, pp.5-9, 2021. https://doi.org/ 10.1088/1742-6596/1948/1/012004

[3] Artificial Intelligence Personal information protection Self-checklist, Personal Information Protection Commission, pp.3-4, 2021. https://www.metaverselaw.com/wp-content/uploads/2021/09/Artificial-IntelligenceAI-Personal-Information-Protection-Self-Checklist2021.07.20.final%E2%98%85.pdf

[4] Reliable AI Development Guide, Telecommunications Technology Association, pp.1-5, 2022.

[5] Kevin Buehler & Rachel Dooley & Liz Grennan, & Alex Singla, "Getting to know-and-manage-your-bgiggest AI risks," McKinsey, pp.2-5, 2021. https://www.mckinsey.com/capabilities/quantumblack/out-insights/getting-to-know-and-manage-your-biggest-ai-risk

[6] Security guidelines for using Generative AI such as ChatGPT, Nationnal Intelligence Service, 2023. https://ncsc.go.kr

[7] Standard for measures to secure the safety of personal information, Personal Information Protection Commission, 2023. https://law.go.kr

[8] Guide to using cloud computing services in finance, Financial Security Institute, pp.55-78, 2023.

[9] ISMS-P Certification Criteria Detailed Check Items, Korea Internet & Security Agency, 2022. https://isms-p.kisa.or.kr/main/ispims/notice/

◐ 저 자 소 개 ◑

조 규 민(Kyu-min Cho)
1993년 서울대학교 계산통계학과(이학사)
2002년 동국대학교 정보보호대학원 정보보호학과(공학석사)
2015년~현재 금융보안원 데이터혁신센터 센터장
2023년~현재 세종대학교 대학원 컴퓨터공학과(박사과정)
관심분야 : 정보보호, AI보안, 금융보안, 사이버레질리언스 etc.
E-mail : gmcho69@naver.com

이 재 일(Jae-il Lee)
1986년 서울대학교 계산통계학과(이학사)
1992년 서울대학교 계산통계학과(이학석사)
2006년 연세대학교 컴퓨터과학과(공학박사)
1996년~2023년 한국인터넷진흥원
2023년~현재 스마일게이트
관심분야 : 침해사고 대응, PKI, 인증, 모바일보안, etc.
E-mail : jilee0218@gmail.com

신 동 규(Dong-kyoo Shin)
1986년 서울대학교 계산통계학과(이학사)
1992년 Illinois Institute of Technology 대학원 컴퓨터과학과(공학석사)
1997년 Texas A&M University 대학원 컴퓨터과학과(공학박사)
1998년~현재 세종대학교 컴퓨터공학과 교수
관심분야 : 머신러닝, 유비쿼터스 컴퓨팅, 생체신호 데이터처리, 정보보호, etc.
E-mail : shindk@sejong.ac.kr