ChatGPT

금융보안원 해커톤 테스트 데이터 분석

해커톤 개요

금융보안원과 다양한 금융기관들이 주최하는 **2025 금융 AI Challenge**는 금융보안 업무에 적합한 생성형 AI 모델을 찾기 위한 해커톤이다. 대회는 FSKU(Financial Security Knowledge Understanding) 평가지표에서 제공하는 객관식·주관식 문제에 대해 정확한 답을 내놓는 모델을 평가한다. 대회 설명에 따르면 참가 팀들은 FSKU 평가지표를 기반으로 **금융보안 실무에 적합한 AI 모델의 성능을 경쟁**하고 [1], **객관식과 주관식 응답을 모두 생성할 수 있는 단일 LLM**을 개발해야 한다 [2]. 이는 금융보안 영역의 전문지식과 금융관련 법규를 포괄하는 모델의 정확도를 겨루는 대회다.

데이터셋 구성

주어진 테스트 데이터는 515개의 질문으로 구성되어 있다. 질문 중 500개(97.1%)는 선택지와 함께 제공되는 **객관식** 문제이며, 15개(2.9%)는 자유롭게 답변을 작성해야 하는 **주관식** 문제이다. 각 질문은 법률, 개인정보보호, 금융 산업, 네트워크 보안, 악성코드 등 다양한 금융/보안 주제를 아우른다. 주요 법률명과 등장 횟수를 세어보면 `개인정보보호법` 언급이 40건, `전자금융거래법` 25건, `정보통신망법` 24건, `전자서명법` 16건, `신용정보법` 15건으로 나타났다. 이는 FSKU가 금융보안 업무에서 빈번하게 활용되는 법규를 집중적으로 다루고 있음을 시사한다.

분야별 분포 분석

질문을 분야별로 분류하기 위해 법규명과 핵심 키워드를 바탕으로 11개의 상위 영역을 정의했다. 법률명을 포함하는 질문은 **법규/규제** 영역으로 묶었고, 나머지는 개인정보보호, 금융/투자, 정보보호 거버넌스, 네트워크 보안, 사이버 위협, 암호화·인증, 접근 통제, 위험관리·재해복구, 보안 솔루션, 기타 등으로 분류하였다. 아래 표는 각 영역에 속한 질문 수와 전체에서 차지하는 비중을 요약한 것이다.

영역(분야)	설명(핵심 키워드)	질문 수	비율
법규/규제	개인정보보호법, 전자금융거래법, 정보통신망법, 전자서명법, 신용정보법 등 금융보안 관련 법규·규정	145	28.2%
금융/투자	금융산업·투자·보험·대출·예금·선불전자지급수단 등	56	10.9%
접근 통제/권한	계정·권한·인증·비밀번호·로그관리 등 접근통제 및 권한관리	50	9.7%
개인정보/신용정보보호	개인정보 처리, 전송요구권, 고유식별정보, 개인정보관리 전문기관 등	42	8.2%
정보보호/거버넌스	정보보호 관리체계, 거버넌스, 정책 수립, 감사, 사이버 복원력 등	38	7.4%
네트워크/시스템 보안	방화벽, VPN, IDS/IPS, 이메일·도메인 보안(SPF), NAT, VLAN 등	37	7.2%
사이버 위협/악성코드	트로이 목마, 랜섬웨어, 피싱, APT, 취약점 스캐닝, 하이브리드 위협 등	34	6.6%
암호화/인증	암호화 방식, 전자서명·전자문서, 키 관리, PKI, OTP 등	33	6.4%

영역(분야)	설명(핵심 키워드)	질문 수	비율
위험관리/재해복구	위험 분석, 재해 복구 계획, BCP, 비상절차 등	6	1.2%
보안 솔루션	SBOM, EDR, SIEM, 안티바이러스 등 보안솔루션·도구	5	1.0%
기타	위 분류에 명확히 속하지 않는 일반 정보보안/IT 주제(SDLC, 클라우드, IoT 등)	69	13.4%

질문 유형별 분포

대부분의 범주에서 객관식 문제 비율이 높지만, 사이버 위협/악성코드, 네트워크/시스템 보안, 접근 통제 같은 기술 영역에서는 주관식 문제도 일부 포함되어 있다. 다음 표는 각 분야에서 객관식·주관식 문제 수를 나열한 것이다.

분야	객관식	주관식	비고
법규/규제	143	2	금융보안 관련 법령·조항을 묻는 선택형이 대부분
금융/투자	55	1	금융산업·투자·전자지급수단 관련 규제와 실무
개인정보/신용정보보호	42	0	개인정보 전송요구권, 신용정보 처리 등
정보보호/거버넌스	37	1	관리체계 수립, 정책 수립, 거버넌스, 사이버 복원력
네트워크/시스템 보안	34	3	네트워크 구성 및 프로토콜, 이메일 인증, VLAN 등
사이버 위협/악성코드	31	3	악성코드 특징, 하이브리드 위협, 공격 대응 전략
암호화/인증	32	1	암호화 방식, 전자서명, 키 관리
접근 통제/권한	47	3	계정·권한 관리, 다중 인증, 접근제어
위험관리/재해복구	6	0	위험평가 방법론, 재해 복구 계획
보안 솔루션	5	0	SBOM, 안티바이러스 등
기타	68	1	SDLC, 클라우드, IoT, 금융보안 거버넌스 등

시사점 및 데이터 수집 전략 제안

테스트 데이터의 약 28 %가 금융보안 관련 법률과 규제를 묻는 문제이다. 특히 **개인정보보호법**, **전자금융거래법**, **정보통신망법**, **전자서명법**, **신용정보법**이 자주 등장한다. 이는 모델을 튜닝할 때 해당 법률의 조문·해설·FAQ 등 **정확한 법령 정보를 포함한 학습 자료**가 필수적임을 의미한다. 또한 개인정보·신용정보 보호(8.2 %), 접근통제/권한관리(9.7 %) 등도 비중이 크므로, 인증·권한 부여 절차와 개인정보보호 가이드라인에 대한 자료를 확보해야 한다.

기술적인 보안 영역에서는 네트워크/시스템 보안(7.2 %), 사이버 위협/악성코드(6.6 %), 암호화/인증(6.4 %)과 같은 주제가 고르게 분포한다. 따라서 **네트워크 프로토콜**, **침해사고 대응**, **악성코드 탐지·분석**, **암호화 방법론** 등에 대한 한글 자료(매뉴얼, 표준, 교육 자료)를 수집해 모델의 기술적 이해도를 높이는 것이 좋다. 정보보호 거버넌스(7.4 %)와 위험관리·재해복구(1.2 %) 관련 질문은 정책 수립과 조직 차원의 대응 역량을 다루므로, **ISO 27001/ISMS-P/금융보안원 지침**과 **업무 연속성 계획(BCP)** 관련 문서도 유용하다.

마지막으로 기타 항목(13.4 %)에는 SDLC, 클라우드 마이그레이션, IoT 융합서비스, 딥페이크와 같은 신기술 이슈가 포함된다. 모델의 범용성을 높이기 위해 이러한 **최신 IT 동향과 보안 이슈에 관한 자료**도 함께 수집·학습하는 것을 권장한다.

---

[1] [2] 2025 금융 AI Challenge : 금융 AI 모델 경쟁 - DACON
https://dacon.io/competitions/official/236527/overview/description