KRA-AiGov 기업 내 AI 서비스 사용 기업보안 솔루션 기획 업데이트

1. 기존 기획 문서 분석 및 정리

제공해주신 2025.05.30log.docx, 2025.06.04log.docx, 2025.06.05log.docx, AI_Proxy_보안정책_20250527_v10.docx, AiGov_전체대화기록_20250609_1559.md, KRASE_AiGov_DevGuide_20250605.md 문서를 분석한 결과, KRA-AiGov 솔루션은 기업 내 AI 서비스 사용에 대한 보안 위협을 해결하고 효율적인 AI 거버넌스를 구축하기 위한 목적으로 기획되었음을 확인했습니다. 주요 내용은 다음과 같습니다.

1.1 문제 정의 및 주요 위협 시나리오

- **문제 정의**: GPT, Copilot 등 AI 도구의 업무 활용 증가에 따른 민감 데이터 외부 전송 및 학습, Shadow AI 확산, 제3자 API 통신으로 인한 데이터 주권 상실, AI 생성 코드/문서의 보안 검증 미비 등의 문제가 제기되었습니다.
- **주요 위협 시나리오**: AI 챗봇에 내부 문서 붙여넣기(민감 정보 유출), GitHub Copilot 사용(보안 취약 코드 생성), 외부 SaaS 기반 AI 활용(고객/계약 정보 노출), AI 제안 답변 사용(저작권 침해, 사실 왜곡) 등이 언급되었습니다.

1.2 보안 강화를 위한 솔루션 아이디어

- **사용 가시화 및 모니터링**: AI 사용 이력 기록 및 모니터링 시스템, Shadow AI 탐지 기능.
- **AI 접근 통제**: 사내 AI Proxy 서버 구축(입력 필터링, 민감 정보 마스킹), 사내 인증 AI 모델만 허용.
- **사용자 교육 및 정책 수립**: AI 사용 보안 교육, 부서/직무별 가이드라인, 생성형 AI 사용 신청 승인 체계.
- **AI 감사 및 컴플라이언스**: AI 사용 리포트, AI Risk Assessment Checklist, ISMS-P 기준 연동.

1.3 솔루션 구조 예시 및 확장 아이디어

- **솔루션 구조**: AI 사용 정책 및 가이드라인, AI Usage Proxy Gateway, 사용 모니터링 및 대시보드, 민감 정보 탐지 및 차단 시스템, 보안 감사 및 리스크 리포트.
- **확장 아이디어**: 사내 AI 모델 배포(LLM on Private Cloud), 문서 분류 및 민감도 자동 분류 AI 연계, AI 보안 관제센터(AI-SOC) 역할 신설, 다국적 기업 기준 AI Risk Framework 벤치마킹 (NIST AI RMF, ISO/IEC 42001).

1.4 AI Proxy 아키텍처 및 세부 기능

- **아키텍처**: 사용자 PC → AI Proxy 서버 → 외부 AI API (사내 보안 Agent, 입력/출력 필터링, 민감정보 마스킹, 사용로그 기록 포함).
- **주요 기능**: 인증 및 권한 확인, 입력 필터링, 출력 필터링, 마스킹 처리, 로그 기록, 정책 엔진, 허용 모델 목록 관리.
- **기술 요소**: DLP 연동, Zero Trust 기반 접근 제어, TLS 암호화 통신, SHA256 Hash 저장.
- **구축 방식**: Web Proxy (Nginx + Lua / Envoy), 인증 (SSO, LDAP), 필터링 엔진 (RegEx, NLP), 로그 저장 (ELK Stack), 운영 (On-prem/컨테이너).

1.5 Shadow AI 탐지 체계

- **정의 및 위험성**: 비인가 AI 도구 사용으로 인한 민감 정보 유출, 데이터 통제권 상실, 오류 책임 불분명, 컴플라이언스 미준수.
- **탐지 전략**: 네트워크 트래픽 분석(AI 관련 도메인 접근), 엔드포인트 모니터링(AI 앱 설치, 브라우저 확장 프로그램), 로그 기반 이상 행동 탐지(업무 시간 외 사용, 과도한 복사/붙여넣기), 사용자 행동 기반 탐지(AI 활용 목적, 민감 키워드).
- **시스템 구성**: 네트워크 흐름 감지 모듈, 브라우저/OS 수준 Agent, 로그 수집기, AI 탐지 시각화 대시보드.

1.6 AI 사용 로그 대시보드 구성

- **목적**: AI 사용 내역 시각화, 이상 행위 및 트렌드 분석.
- **수집 항목**: 사용자 정보, 접속 시간/IP/디바이스, AI 사용 플랫폼/모델명, 입력/출력 텍스트 샘플, 민감정보 항목, 정책 위반 여부.
- **기술 구성**: 로그 수집기(Filebeat, Winlogbeat), 데이터 저장소(Elasticsearch, InfluxDB), 시각화 도구(Kibana, Grafana).

1.7 내부 LLM 도입 방안

- **목적**: 외부 API 의존도 감소, 데이터 주권 확보, 사내 지식 기반 맞춤형 AI 제공, 비용 효율성 및 보안성 향상.
- **구축 전략**: 모델 선정(경량 오픈소스 LLM), 학습 및 파인튜닝(사내 문서 기반 RAG), 인프라 구성(On-Prem/Private Cloud), 인터페이스(내부 API/웹 챗봇).

1.8 ISMS-P 기반 AI 보안 감사 항목 설계

- **필요성**: AI 사용이 ISMS-P 통제 항목에 포함되지 않을 경우 인증 유지 및 규정 위반 가능성.
- **감사 항목 예시**: AI 사용 목적/범위 정의, 정책 수립/공지, 사용 로그 보관/감사, 민감정보 탐지/차단 시스템, 외부 AI 서비스 계약/보안 검토, Shadow AI 탐지 체계, 내부 AI 솔루션 보안/유지관리.

1.9 브랜딩 전략 및 명칭 (잠정 확정)

- **메인 브랜드명**: KRA-AIGov
- **부속 시스템 명칭**: PromptShield (프롬프트 필터링), Shadoweye (Shadow AI 탐지), DashAILog (로그 대시보드), FortLLM (내부 LLM).
- **상표권 검토 결과**: KRA-AIGov, DashAILog, FortLLM은 현재 상표 등록 이슈 없음. PromptShield, Shadoweye는 유사 상표 존재 가능성 있어 상표 등록 진행 또는 대체 명칭 고려 필요.

2. AI 거버넌스 및 기업보안 솔루션 시장 현황 분석

2.1 시장 규모 및 성장 전망

AI 기반 사이버 보안 시장은 2024년 약 253억 달러 규모로 평가되며, 2030년까지 연평균 24.4% 성장하여 약 937억 달러에 이를 것으로 전망됩니다. 특히 AI의 신뢰성, 위험 및 보안 관리(TRiSM) 시장은 2024년 23억 달러에서 2030년 74억 달러로 성장할 것으로 예상되어, AI 보안 분야의 중요성이 더욱 커지고 있음을 시사합니다.

2.2 주요 기술 동향

1. **AI 기반 위협 탐지 및 대응**: AI는 대규모 데이터 분석을 통해 이상 징후를 실시간으로 탐지하고, 자동화된 대응을 가능하게 합니다. 이는 기존 보안 시스템의 한계를 보완하며, 더욱 정교한 위협에 대한 방어 능력을 강화합니다.
2. **에이전틱 AI(Agentic AI)의 부상**: 자율적으로 작업을 수행하는 에이전틱 AI는 보안 운영 센터(SOC)에서의 업무 효율성을 높이는 데 기여하고 있습니다. 이는 보안 전문가의 업무 부담을 줄이고, 신속한 위협 대응을 가능하게 합니다.
3. **프롬프트 인젝션 공격 대응**: 대규모 언어 모델(LLM)을 대상으로 한 프롬프트 인젝션 공격이 증가함에 따라, 이에 대한 보안 강화 기술 개발이 활발히 이루어지고 있습니다. 이는 LLM의 오용을 방지하고, 안전한 AI 서비스 제공을 위한 필수적인 요소입니다.

2.3 주요 보안 이슈

1. **Shadow AI의 확산**: 기업 내 승인되지 않은 AI 도구 사용이 증가하면서 데이터 유출 및 보안 위협이 심화되고 있습니다. 이는 기업의 통제 범위를 벗어나 발생하며, 민감 정보 유출, 컴플라이언스 위반 등의 문제를 야기할 수 있습니다.
2. **AI 에이전트의 보안 취약성**: AI 에이전트가 인증 정보를 노출하거나 비인가 시스템에 접근하는 사례가 보고되고 있으며, 이는 새로운 형태의 공격 경로로 악용될 수 있습니다.

2.4 국내외 주요 기업 및 솔루션 현황

2.4.1 AI를 활용한 보안 솔루션 (AI for Security)

AI 기술을 활용하여 사이버 위협을 탐지하고 대응하는 솔루션들이 국내외에서 활발히 개발되고 있습니다.

- **해외**: Darktrace (비지도 학습 기반 실시간 위협 탐지), Vectra AI (행위 기반 위협 탐지), Microsoft Security Copilot (보안 팀 지원 AI 에이전트).
- **국내**: 지니언스 (EDR, MDR, 클라우드 기반 NAC), 슈프리마 (AI 기반 얼굴인식 및 행동분석), 라온시큐어 (프리미엄 모의해킹, FIDO 기반 생체인증, 딥페이크 탐지).

2.4.2 AI 보안을 위한 솔루션 (Security for AI)

AI 시스템 자체의 보안 취약점을 보호하기 위한 솔루션들은 아직 초기 단계이며, 전통 보안 벤더들이 일부 기능을 통합하거나 시범 제공하고 있습니다. 완전한 상용화 제품은 아직 많지 않습니다.

- **클라우드 기반 서비스**: Microsoft Security Copilot (LLM 기반 보안 오퍼레이션 보조, 사용자 입력 필터링, LLM 행동 추적), Google Vertex AI with IAM & Data Loss Prevention (AI 모델 접근 제어, 민감 정보 필터링), AWS Guardrails for Amazon Bedrock (프롬프트 필터링, 응답 검열).
- **신생 기업/솔루션**: Lakera (프롬프트 인젝션 공격 방어), Protect AI (오픈소스 AI 보안 감사 도구), Robust Intelligence (AI 시스템 사전 점검 자동화), HiddenLayer (AI 시스템 보호).

2.4.3 현실적 한계 및 시사점

대부분의 상용 솔루션은 'AI 보안'보다는 'AI를 활용한 보안'에 집중되어 있습니다. AI 자체 보안을 위한 기능은 오픈소스 또는 API 기반 엔진 수준이거나, 프록시 게이트웨이, 입력 필터링, RAG 보안 강화 등은 기업별 커스터마이징 구축이 대부분입니다. 이는 KRA-AiGov와 같은 AI 거버넌스 및 보안 솔루션이 시장에서 차별화된 경쟁력을 가질 수 있는 기회가 될 수 있음을 시사합니다.

3. 경쟁사 분석 및 기술 트렌드 조사

3.1 주요 경쟁사 분석

AI 보안 솔루션 시장은 다양한 접근 방식을 가진 기업들이 경쟁하고 있으며, 크게 'AI를 활용한 보안'과 'AI 자체의 보안' 두 가지 영역으로 나눌 수 있습니다. KRA-AiGov는 후자에 더 가깝지만, 시장의 전반적인 동향을 이해하기 위해 주요 경쟁사들을 분석합니다.

3.1.1 Darktrace

- **주요 기술**: 비지도 학습(Unsupervised Learning) 기반의 AI를 활용하여 네트워크 내의 정상적인 행위를 학습하고, 이상 징후를 실시간으로 탐지합니다. '자율형 사이버 AI'를 표방하며, 인간의 개입 없이 위협을 스스로 방어하는 데 중점을 둡니다.
- **솔루션 특징**: 엔드포인트, 클라우드, SaaS, 이메일 등 전방위적인 영역에서 위협을 탐지하고 대응합니다. 특히 '자율 대응(Autonomous Response)' 기능은 공격 발생 시 자동으로 방어 조치를 취하여 피해를 최소화합니다.
- **강점**: 독자적인 AI 기술을 통해 알려지지 않은 위협(Zero-day attacks)까지 탐지할 수 있으며, 자동화된 대응으로 보안 운영 효율성을 높입니다.
- **약점**: 고도화된 AI 기술로 인해 솔루션 도입 및 운영 비용이 높을 수 있으며, AI의 판단에 대한 투명성 문제가 제기될 수 있습니다.

3.1.2 Vectra AI

- **주요 기술**: 행위 기반 분석(Behavioral Analytics)을 통해 네트워크, 아이덴티티, 클라우드 전반에 걸쳐 공격자의 움직임을 탐지합니다. AI를 활용하여 위협의 우선순위를 지정하고, 공격 캠페인을 식별하는 데 특화되어 있습니다.
- **솔루션 특징**: 'AI 기반 위협 탐지 및 대응(NDR)' 솔루션으로, 네트워크 트래픽을 분석하여 숨겨진 위협을 찾아냅니다. 특히 공격자의 '행위'에 집중하여 오탐을 줄이고 실제 위협에 대한 가시성을 제공합니다.
- **강점**: 공격자의 전술, 기술, 절차(TTPs)를 기반으로 위협을 식별하여 정교한 탐지 능력을 제공하며, 보안 팀의 조사 시간을 단축시킵니다.
- **약점**: 네트워크 가시성에 의존하므로, 암호화된 트래픽 내부의 위협 탐지에는 한계가 있을 수 있습니다.

3.1.3 Microsoft Security Copilot

- **주요 기술**: 대규모 언어 모델(LLM) 기반의 AI 에이전트로, 보안 팀의 업무를 보조하고 위협 분석 및 대응을 돕습니다. Microsoft의 방대한 보안 데이터와 위협 인텔리전스를 활용합니다.
- **솔루션 특징**: 보안 정보 및 이벤트 관리(SIEM), 확장된 탐지 및 대응(XDR) 솔루션과 통합되어 작동하며, 자연어 기반의 질의응답을 통해 보안 전문가가 위협을 더 빠르고 효율적으로 조사하고 대응할 수 있도록 지원합니다.
- **강점**: Microsoft 생태계와의 긴밀한 통합을 통해 폭넓은 가시성을 제공하며, LLM을 활용하여 보안 전문가의 생산성을 크게 향상시킬 수 있습니다.
- **약점**: Microsoft 제품에 대한 의존도가 높으며, LLM의 한계로 인한 오정보 제공 가능성, 프롬프트 인젝션과 같은 AI 자체의 보안 취약성에 대한 고려가 필요합니다.

3.1.4 Lakera (Lakera Guard)

- **주요 기술**: LLM 기반 애플리케이션을 위한 보안 솔루션으로, 특히 프롬프트 인젝션 공격 방어에 특화되어 있습니다. API 형태로 LLM 서비스에 연동되어 작동합니다.

- **솔루션 특징**: 사용자 프롬프트와 LLM 응답을 실시간으로 분석하여 악의적인 입력이나 유해한 출력을 탐지하고 차단합니다. LLM의 안전한 사용을 위한 '가드레일' 역할을 수행합니다.
- **강점**: LLM 특화 보안에 집중하여 프롬프트 인젝션, 데이터 유출, 유해 콘텐츠 생성 등 LLM 관련 위협에 대한 전문적인 방어 기능을 제공합니다.
- **약점**: LLM 보안에만 초점을 맞추므로, 기업 내 AI 사용 전반에 대한 거버넌스 및 Shadow AI 탐지 등 포괄적인 보안 기능은 부족할 수 있습니다.

3.1.5 Protect AI (ModelScan, NB Defense)

- **주요 기술**: 머신러닝 모델 및 데이터셋의 취약점을 스캔하고 분석하는 도구를 제공합니다. 오픈소스 기반의 AI 보안 감사 도구들을 개발하고 있습니다.
- **솔루션 특징**: ModelScan은 AI 모델의 취약점(예: 모델 역공학, 데이터 추출 공격)을 식별하고, NB Defense는 노트북 환경에서의 AI 개발 보안을 강화합니다. AI/ML 개발 라이프사이클 전반에 걸친 보안을 목표로 합니다.
- **강점**: AI 모델 자체의 보안 취약점을 분석하고 개선하는 데 특화되어 있어, AI 시스템의 근본적인 보안 강화를 지원합니다.
- **약점**: 주로 개발 단계의 보안에 초점을 맞추므로, 운영 단계에서의 실시간 위협 탐지 및 대응 기능은 제한적일 수 있습니다.

3.1.6 Robust Intelligence (RIME)

- **주요 기술**: AI 시스템의 신뢰성, 공정성, 보안성을 검증하고 자동화하는 플랫폼을 제공합니다. AI 인프라에 보안 인터셉터를 삽입하여 모델의 오작동이나 공격을 방지합니다.
- **솔루션 특징**: 모델 배포 전후의 지속적인 모니터링을 통해 데이터 드리프트, 모델 편향, 적대적 공격 등을 탐지하고 경고합니다. AI 모델의 '품질 보증' 및 '보안 강화'를 동시에 수행합니다.
- **강점**: AI 모델의 라이프사이클 전반에 걸쳐 신뢰성과 보안을 확보하는 데 강점을 가지며, 특히 적대적 공격에 대한 방어 능력이 뛰어납니다.
- **약점**: 복잡한 AI 시스템에 대한 통합이 필요하며, 초기 도입 비용 및 운영 난이도가 높을 수 있습니다.

3.1.7 HiddenLayer

- **주요 기술**: AI 모델의 공격 표면을 실시간으로 모니터링하고, AI 기반 위협을 탐지하는 소프트웨어 솔루션을 제공합니다. 'AI 보안 플랫폼'을 표방합니다.
- **솔루션 특징**: AI 모델에 대한 악의적인 입력(예: 적대적 예제)이나 모델 탈취 시도 등을 탐지하고 방어합니다. 클라우드 및 온프레미스 환경 모두에서 배포 가능하며, 기존 보안 스택과 통합될 수 있습니다.
- **강점**: AI 모델에 대한 특화된 보안 기능을 제공하여, AI 시스템의 무결성과 가용성을 보호합니다.
- **약점**: AI 모델 자체의 보안에 집중하므로, Shadow AI 탐지나 AI 사용 정책 관리 등 거버넌스 영역은 별도의 솔루션이 필요할 수 있습니다.

3.2 기술 트렌드 및 시사점

AI 보안 시장은 'AI를 활용한 보안'에서 'AI 자체의 보안'으로 점차 확장되고 있으며, 특히 LLM의 확산과 함께 다음과 같은 트렌드가 두드러집니다.

1. **LLM 보안의 중요성 증대**: 프롬프트 인젝션, 데이터 유출, 유해 콘텐츠 생성 등 LLM 특유의 보안 취약점에 대한 방어 기술이 핵심 트렌드로 부상하고 있습니다. Lakera와 같은 전문 솔루션의 등장이 이를 뒷받침합니다.
2. **AI 거버넌스 및 가시성 강화**: Shadow AI의 확산과 AI 사용에 대한 규제 강화로 인해, 기업 내 AI 사용 현황을 모니터링하고 통제하는 거버넌스 솔루션의 필요성이 커지고 있습니다. KRA-AiGov의 'PromptShield', 'Shadoweye', 'DashAILog'와 같은 기능은 이러한 시장 요구에 부합합니다.
3. **AI 모델 라이프사이클 전반의 보안**: AI 모델의 개발부터 배포, 운영에 이르는 전 과정에서 보안을 고려하는 'MLSecOps' 개념이 중요해지고 있습니다. Protect AI, Robust Intelligence, HiddenLayer와 같은 기업들은 이러한 요구를 충족시키는 솔루션을 제공합니다.
4. **AI 기반 자동화된 보안 운영**: Microsoft Security Copilot과 같이 AI를 활용하여 보안 운영의 효율성을 높이고, 보안 전문가의 업무 부담을 줄이는 방향으로 기술이 발전하고 있습니다.
5. **하이브리드 및 멀티 클라우드 환경 지원**: 다양한 클라우드 환경과 온프레미스 환경에서 AI가 활용됨에 따라, 이들을 모두 아우를 수 있는 유연한 보안 솔루션의 중요성이 커지고 있습니다.

KRA-AiGov는 'AI 자체의 보안'과 'AI 거버넌스'라는 두 가지 핵심 축을 중심으로 경쟁력을 확보할 수 있습니다. 특히 국내 시장에서 Shadow AI 탐지 및 AI 사용 정책 관리에 대한 명확한 솔루션이 부족한 상황에서, KRA-AiGov의 포괄적인 접근 방식은 큰 강점이 될 수 있습니다. 또한, 내부 LLM 도입 방안(FortLLM)을 통해 데이터 주권 확보 및 맞춤형 AI 서비스 제공이라는 차별화된 가치를 제공할 수 있습니다.

4. MVP 모델 기능 및 개발 계획 업데이트

기존 기획 문서 분석, AI 거버넌스 및 기업보안 솔루션 시장 현황, 그리고 경쟁사 분석 결과를 바탕으로 KRA-AiGov의 MVP(Minimum Viable Product) 모델 기능 및 개발 계획을 현실성 있고 개발 진행이 용이하도록 업데이트합니다.

4.1 MVP 모델 기능 정의

KRA-AiGov의 MVP는 기업 내 AI 서비스 사용에 대한 가시성 확보, 기본적인 통제 및 보안 위협 탐지에 중점을 둡니다. 초기 단계에서는 핵심 기능에 집중하여 빠른 시장 출시와 사용자 피드백 확보를 목표로 합니다.

핵심 기능:

1. **PromptShield (AI 프롬프트 필터링 프록시)**
    - **기능**: 사용자의 AI 서비스 요청(프롬프트)을 가로채어 사전 정의된 정책에 따라 필터링 및 마스킹 처리합니다. 외부 AI API로의 직접적인 데이터 전송을 통제하고, 민감 정보 유출을 방지합니다.
    - **세부 기능**: 민감 정보(개인 식별 정보, 기업 기밀 등) 탐지 및 자동 마스킹, 특정 키워드/문구 차단, AI 서비스별 접근 제어 (화이트리스트/블랙리스트).
    - **기술 스택 고려**: Nginx + Lua 또는 Envoy Proxy 기반의 경량 프록시 구현. 정규표현식(RegEx) 및 간단한 키워드 매칭을 통한 필터링.
2. **DashAILog (AI 사용 로그 대시보드)**
    - **기능**: PromptShield를 통해 처리된 AI 서비스 사용 로그를 수집하고 시각화하여 관리자에게 AI 사용 현황에 대한 가시성을 제공합니다.
    - **세부 기능**: 사용자별/부서별 AI 사용량, 시간대별 사용 트렌드, 필터링/마스킹 이력, 정책 위반 시도 기록. 간단한 검색 및 필터링 기능.
    - **기술 스택 고려**: 로그 수집을 위한 Filebeat/Fluentd, 데이터 저장을 위한 Elasticsearch 또는 경량 DB (SQLite/PostgreSQL), 시각화를 위한 Kibana 또는 Grafana (초기에는 간단한 웹 UI 개발).
3. **Shadoweye (Shadow AI 탐지)**
    - **기능**: 기업 네트워크 내에서 승인되지 않은 AI 서비스 사용 시도를 탐지하고 관리자에게 알림을 제공합니다. 초기 MVP에서는 네트워크 트래픽 분석에 집중합니다.
    - **세부 기능**: AI 서비스 관련 도메인(예: openai.com, anthropic.com) 접근 탐지, 비인가 AI 서비스 접속 시도 알림.
    - **기술 스택 고려**: DNS 로그 분석, 네트워크 트래픽 모니터링 (예: Zeek, Suricata의 경량화된 규칙 적용). 초기에는 네트워크 장비의 로그를 활용하는 방식.

MVP 제외 기능 (향후 개발 고려):

- **FortLLM (내부 LLM)**: 초기 MVP에서는 외부 AI 서비스 통제에 집중하며, 내부 LLM 구축은 다음 단계로 미룹니다. (높은 구축 비용 및 복잡성)
- **고급 정책 엔진**: 부서/직무별 세분화된 정책 제어는 초기에는 수동 설정 또는 간단한 규칙 기반으로 운영하고, 복잡한 정책 엔진은 추후 고도화합니다.
- **ISMS-P 연동 자동화**: 감사 항목 설계는 포함하되, ISMS-P 시스템과의 자동 연동은 다음 단계로 미룹니다.
- **DLP 연동**: 기존 DLP 솔루션과의 실시간 연동은 복잡성이 높으므로, 초기에는 PromptShield 자체의 마스킹 기능에 집중합니다.
- **엔드포인트 기반 Shadow AI 탐지**: 네트워크 트래픽 분석에 우선 집중하고, 에이전트 설치를 통한 엔드포인트 모니터링은 추후 고려합니다.

4.2 개발 계획 (로드맵)

MVP 개발은 3단계로 나누어 진행하며, 각 단계별 목표와 주요 작업을 정의합니다.

단계 1: 핵심 프록시 및 로깅 시스템 구축 (1-2개월)

- **목표**: PromptShield의 기본 필터링 및 마스킹 기능 구현, AI 사용 로그 수집 및 저장.
- **주요 작업**:
  - PromptShield 프록시 서버 아키텍처 설계 및 구축 (Nginx + Lua 또는 Envoy).
  - 기본적인 민감 정보(주민등록번호, 이메일 등) 탐지 및 마스킹 로직 구현.
  - AI 서비스 요청 및 응답 로그 수집 모듈 개발.
  - 로그 저장소 (경량 DB 또는 Elasticsearch) 설정 및 연동.
  - 간단한 웹 기반 DashAILog 대시보드 (사용자별/시간대별 사용량) 구현.
- **기술 스택**: Python/Node.js (필터링 로직), Nginx/Envoy, Docker, SQLite/PostgreSQL, HTML/CSS/JS (대시보드).

단계 2: Shadow AI 탐지 및 대시보드 고도화 (1-2개월)

- **목표**: Shadow AI 탐지 기능 추가 및 DashAILog 대시보드 기능 확장.
- **주요 작업**:
  - Shadoweye 네트워크 트래픽 분석 모듈 개발 (AI 서비스 도메인 탐지).
  - 비인가 AI 서비스 접속 시도 알림 기능 구현.
  - DashAILog 대시보드에 Shadow AI 탐지 현황 및 알림 기능 추가.
  - 필터링 정책 관리 UI 개선 (간단한 정책 추가/수정).
- **기술 스택**: Python (네트워크 분석), 기존 스택 활용.

단계 3: PoC 및 피드백 반영 (1개월)

- **목표**: MVP 솔루션의 사내 PoC(Proof of Concept) 진행 및 사용자/관리자 피드백 수집, 개선.
- **주요 작업**:
  - 선정된 부서/팀에 MVP 솔루션 배포 및 적용.
  - 사용자 교육 및 가이드라인 제공.
  - PoC 기간 동안 데이터 수집 및 성능 모니터링.
  - 수집된 피드백을 바탕으로 기능 개선 및 버그 수정.
  - 향후 개발 로드맵 구체화.
- **기술 스택**: 기존 스택 활용.

총 개발 기간: 약 3-5개월 (초기 MVP 출시까지)

4.3 개발 인력 및 자원 계획

- **개발 인력**: 백엔드 개발자 1-2명, 프론트엔드 개발자 1명, 보안 엔지니어 1명 (총 3-4명).
- **필요 자원**: 개발 서버 (가상 머신 또는 클라우드 인스턴스), 테스트 환경, 버전 관리 시스템 (Git), 협업 도구.

4.4 성공 지표 (KPI)

- **AI 서비스 사용 가시성**: DashAILog를 통한 AI 서비스 사용 현황 파악률 (예: 90% 이상).
- **민감 정보 유출 방지**: PromptShield를 통한 민감 정보 마스킹/차단 성공률 (예: 99% 이상).
- **Shadow AI 탐지율**: Shadoweye를 통한 비인가 AI 서비스 탐지율 (예: 80% 이상).
- **사용자 만족도**: PoC 후 사용자 및 관리자 만족도 조사 (예: 4점/5점 이상).

이러한 MVP 모델 기능 및 개발 계획은 KRA-AiGov가 시장에 빠르게 진입하고, 실제 사용자들의 요구 사항을 반영하여 지속적으로 발전할 수 있는 기반을 마련할 것입니다.