**1. 서론**

최근 생성형 AI(ChatGPT, Gemini 등)의 폭발적 보급으로 인해 AI 보안 위협이 진화하고 있습니다. 2023년 Gartner 조사에 따르면, 기업의 67%가 AI 관련 사이버 공격을 경험했으며, 특히 **딥페이크(Deepfake)** 기반 피싱 공격이 전년 대비 320% 증가했습니다. 본 문서는 2024년 주목해야 할 5대 AI 보안 위협과 대응 방안을 제시합니다.

**2. 2024년 5대 AI 보안 위협**

1. **AI 모델 해킹 (Adversarial Attacks)**
   * 공격자가 이미지/텍스트에 미세한 노이즈를 추가해 AI 모델을 오작동시키는 공격.
   * 사례: 2023년 Tesla Autopilot의 표지판 오인식 사고(약 5초간 정지 표시를 속도 제한 표시로 인식).
2. **딥페이크 기반 사회 공학**
   * 음성/영상 합성 기술을 이용한 CEO 사칭(예: 2024년 3월 영국 에너지 기업 CFO 목소리 사칭으로 2,400만 달러 사기).
3. **프롬프트 인젝션(Prompt Injection)**
   * 악성 프롬프트로 LLM(Large Language Model)을 조작해 금융 정보/개인 데이터 유출.
   * 사례: ChatGPT 플러그인을 통한 비공개 PDF 문서 탈취 시도(2024년 1월).
4. **AI 훈련 데이터 오염(Data Poisoning)**
   * 공급망 공격으로 학습 데이터에 편향/오류를 주입(예: Microsoft의 2023년 보고서에서 12%의 오픈소스 데이터셋이 오염됨).
5. **AI 기반 자동화 공격(AI-powered Botnets)**
   * AI가 생성한 맞춤형 피싱 메일 + CAPTCHA 우회 기술 결합.

**3. 산업별 영향**

* **금융**: 딥페이크 음성 인증 우회(2024년 1분기 78건 보고).
* **의료**: AI 진단 모델의 오류로 인한 과잉 치료(미국 FDA, 2023년 14건 리콜).
* **제조**: 생성형 AI로 설계된 결함 있는 3D 모델 유통 증가.

**4. 대응 전략**

1. **AI 모델 검증(Red Teaming)**
   * 매월 외부 전문가가 모델 취약점 테스트(예: OpenAI의 2024년 3월 GPT-4 Red Team 결과 공개).
2. **실시간 감시 시스템**
   * AI 생성 콘텐츠 워터마킹(C2PA 표준 적용).
3. **법적 제도**
   * EU AI Act(2024년 5월 발효)에 따른 벌칙 강화(최대 글로벌 매출의 7%).

**5. 결론**

AI 보안은 이제 "기술"이 아닌 "경영 리스크"로 인식되어야 합니다. 2025년까지 글로벌 AI 보안 시장은 290억 달러 규모로 성장할 전망이며, 기업은 **Explainable AI(XAI)** 도입으로 모델 결정 근거를 투명화해야 합니다.