SYNOPSYS®

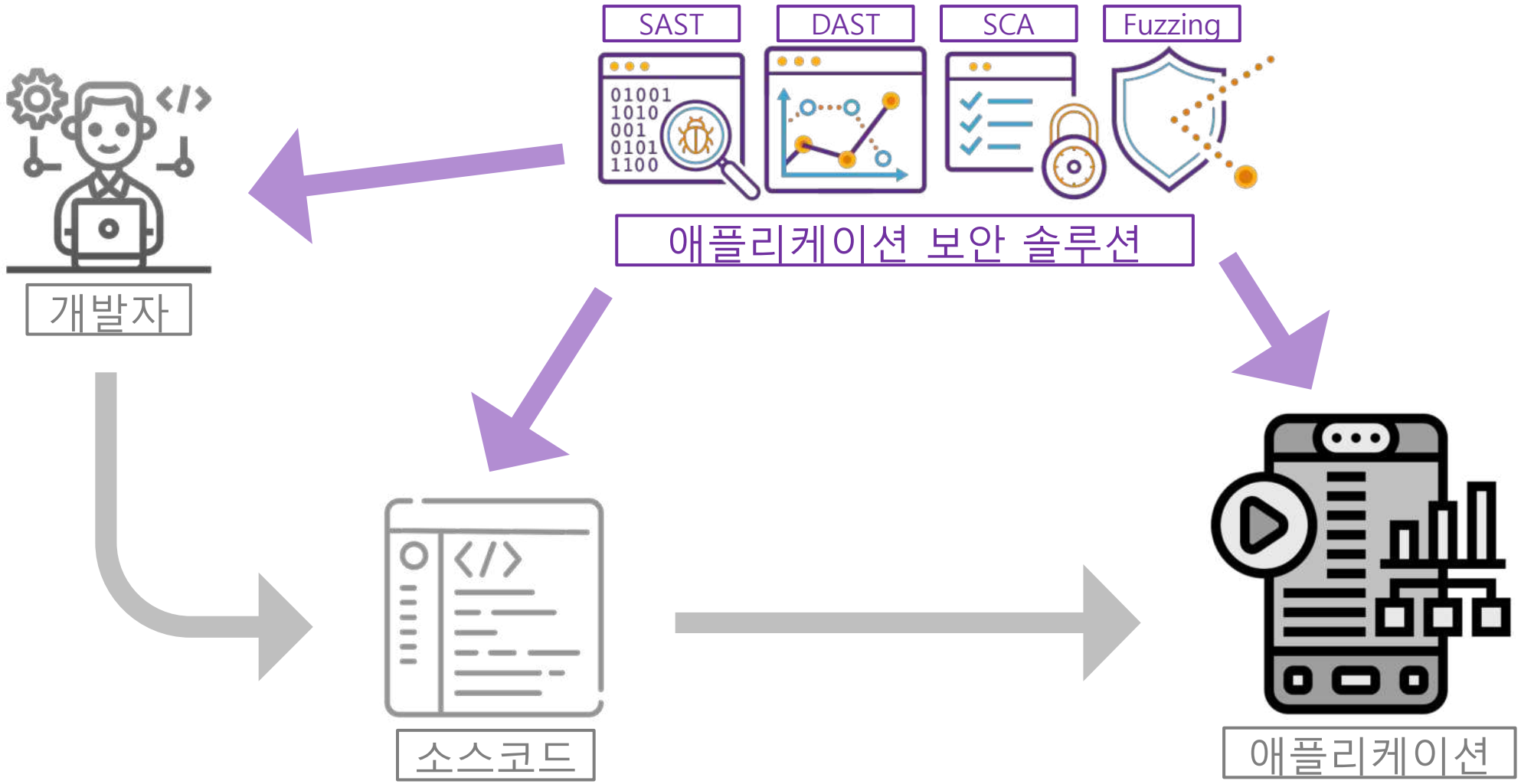
AppSec
- of the AI, by the AI, for the AI

AI시대의 애플리케이션 보안

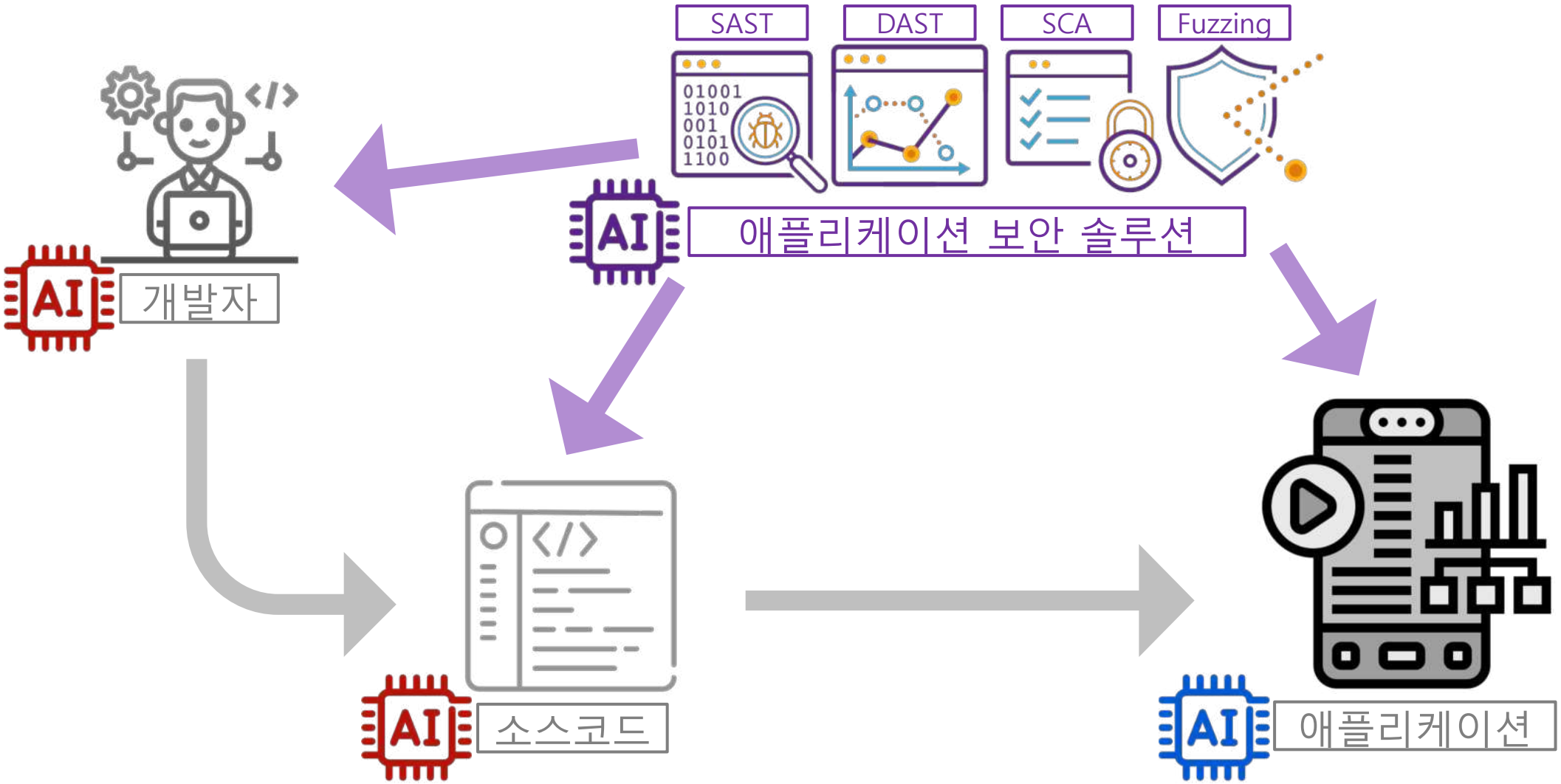
Synopsys Korea SIG, 제병주 부장

2024년 5월 28일

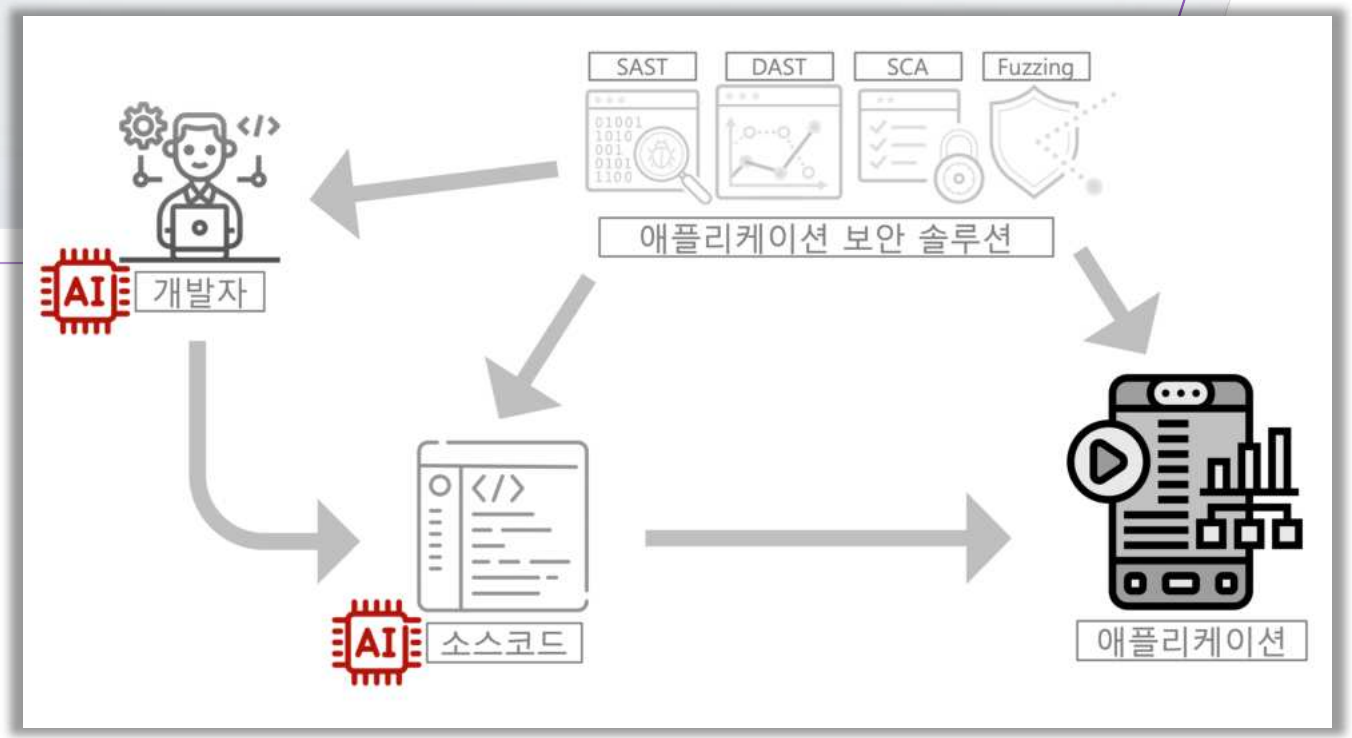
애플리케이션 보안의 일반적인 구성



애플리케이션 보안에서 AI의 활용



AI가 작성한 소스코드



AI를 통해 개발 생산성이 급격하게 증가

50% Overall productivity increase

25% Deployments increase

45% Commits increases

35% Merge request increases

Increased quality through faster unit testing, reduced code errors

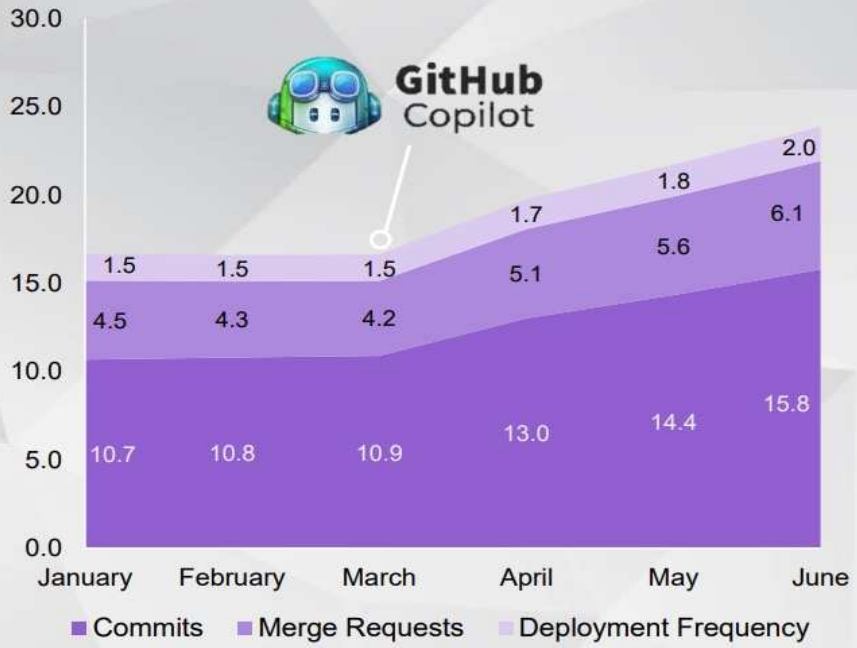
Reduced number of merge conflicts

Increased accessibility through conversational interface

Increased developer satisfaction

Context: ~100 FTE engineering team

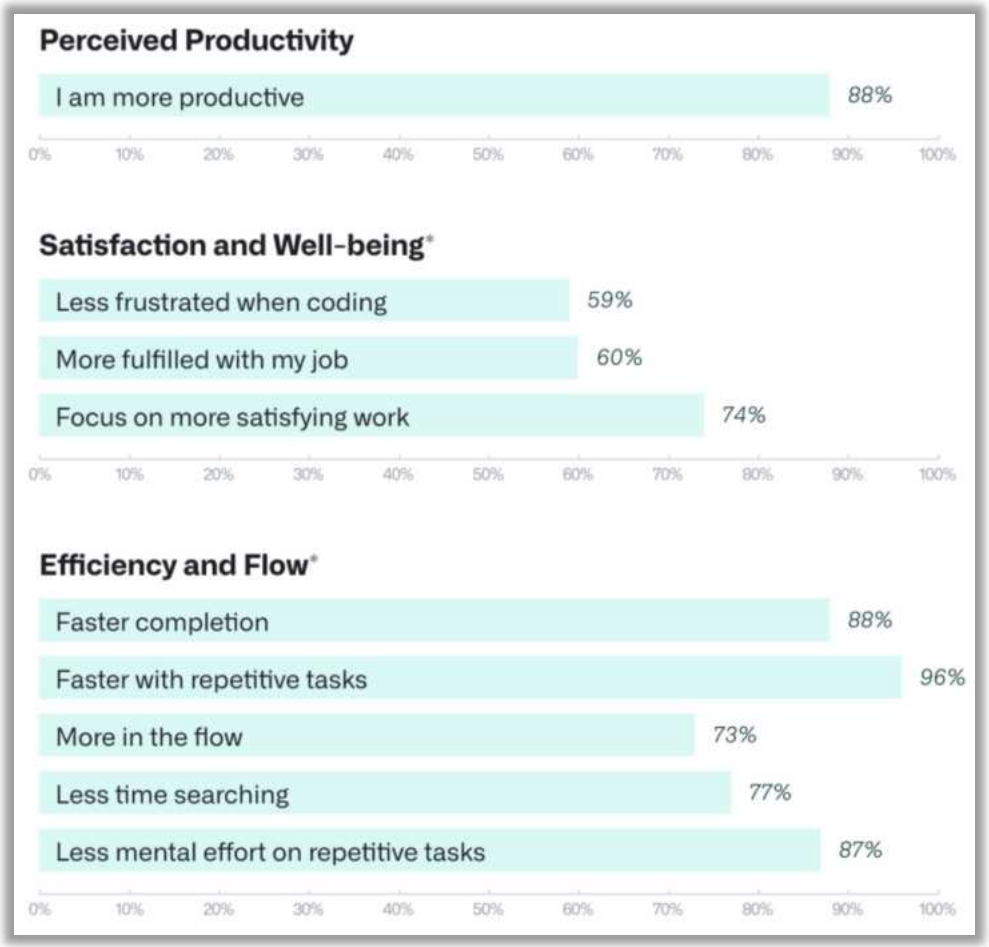
Prerequisites: Strict governance around quality gates & vulnerability testing coverage (SonarQube) implemented



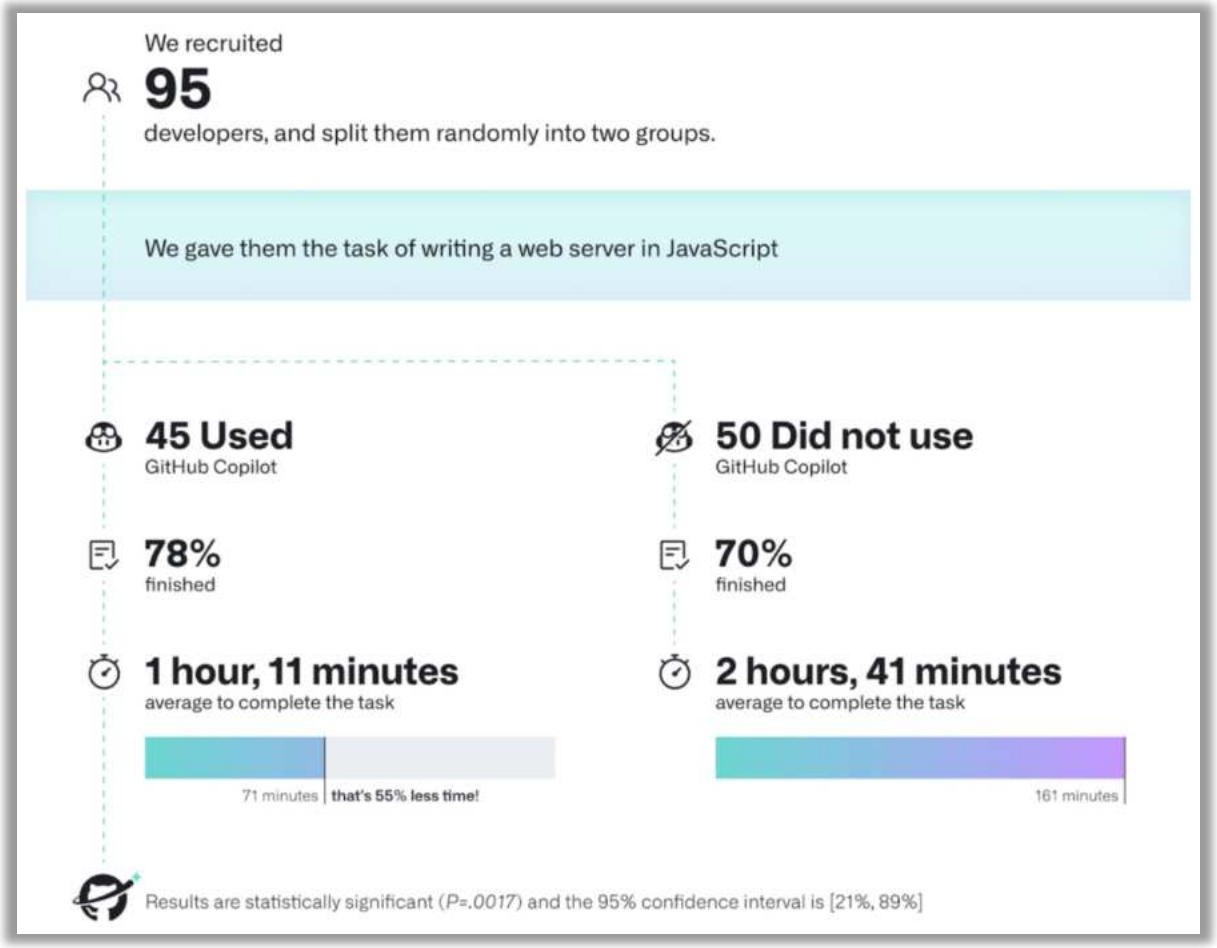
1. Github Copilot analysis https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/
2. Observations on a real case of deployment of GitHub Copilot on 100 Software Engineers

AI를 통해 개발 생산성이 급격하게 증가

개발자가 체감하는 생산성 증가



프로젝트 완수에 걸리는 시간 감소



AI가 작성한 소스코드에서 애플리케이션 보안

AI가 작성한 코드는 사이버보안에 안전한가?

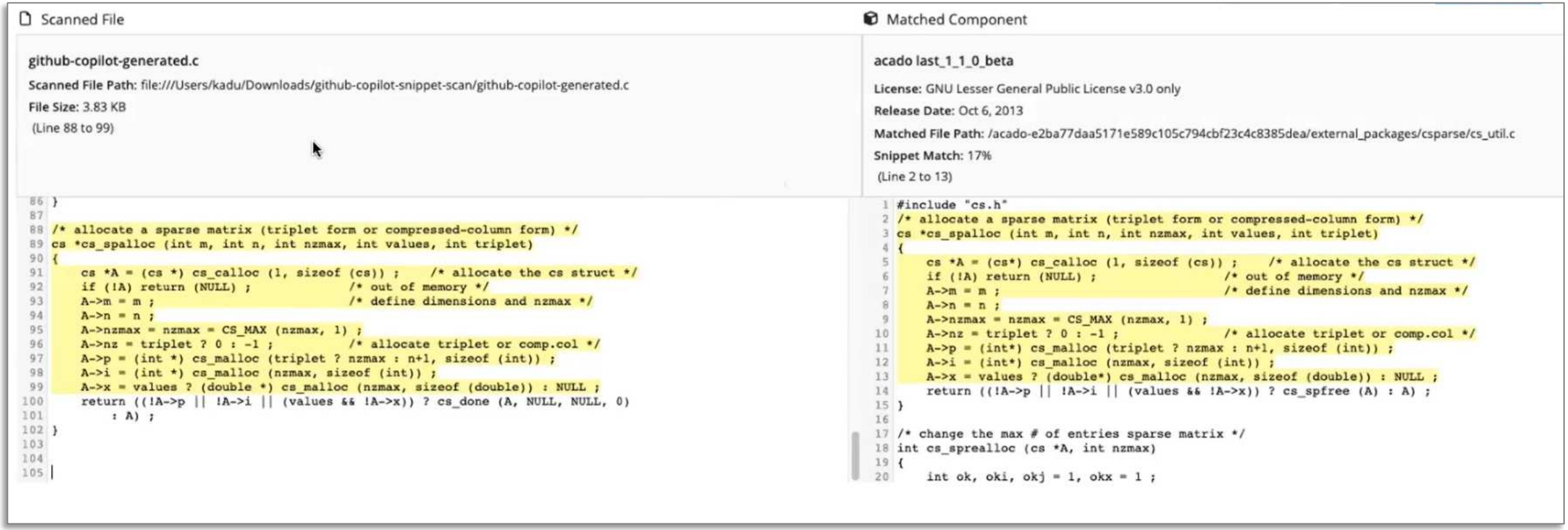
- 아직은 안전하지 않음
- AI가 보안 코딩을 한다 하더라도, 확인 및 안전을 위한 보안 활동이 필요
- 사람이 작성한 코드와 마찬가지로 애플리케이션 보안을 위한 활동이 필요

AI가 작성한 코드는 오픈소스 라이선스를 준수하는가?

- 라이선스 위반이 가장 큰 문제
- AI를 통해 오픈소스 라이선스를 준수한다고 해도, 확인 작업이 필요
- Black Duck의 snippet scan을 통해 오픈소스 라이선스 위반 검출 가능

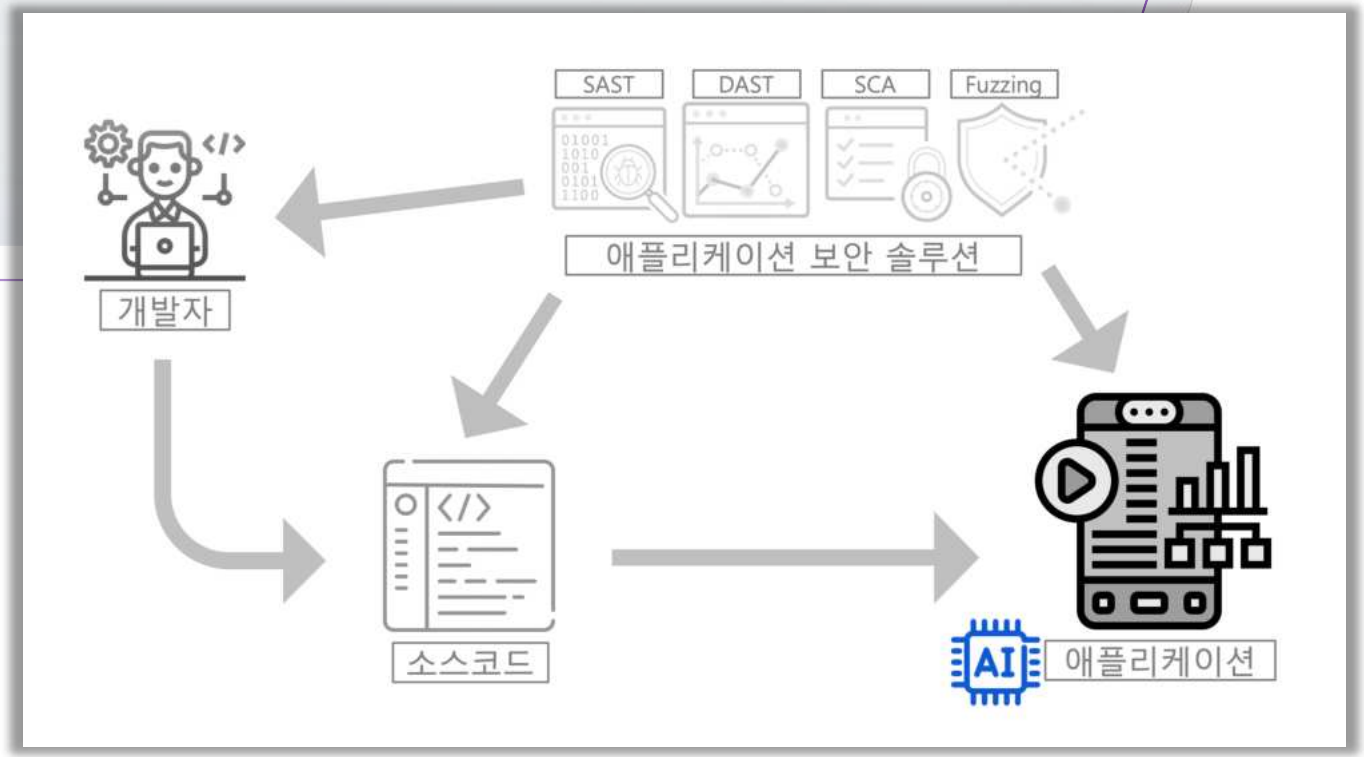
Black Duck SCA를 통한 snippet analysis 예

AI가 생성한 코드에서 오픈소스 라이선스 위험을 식별

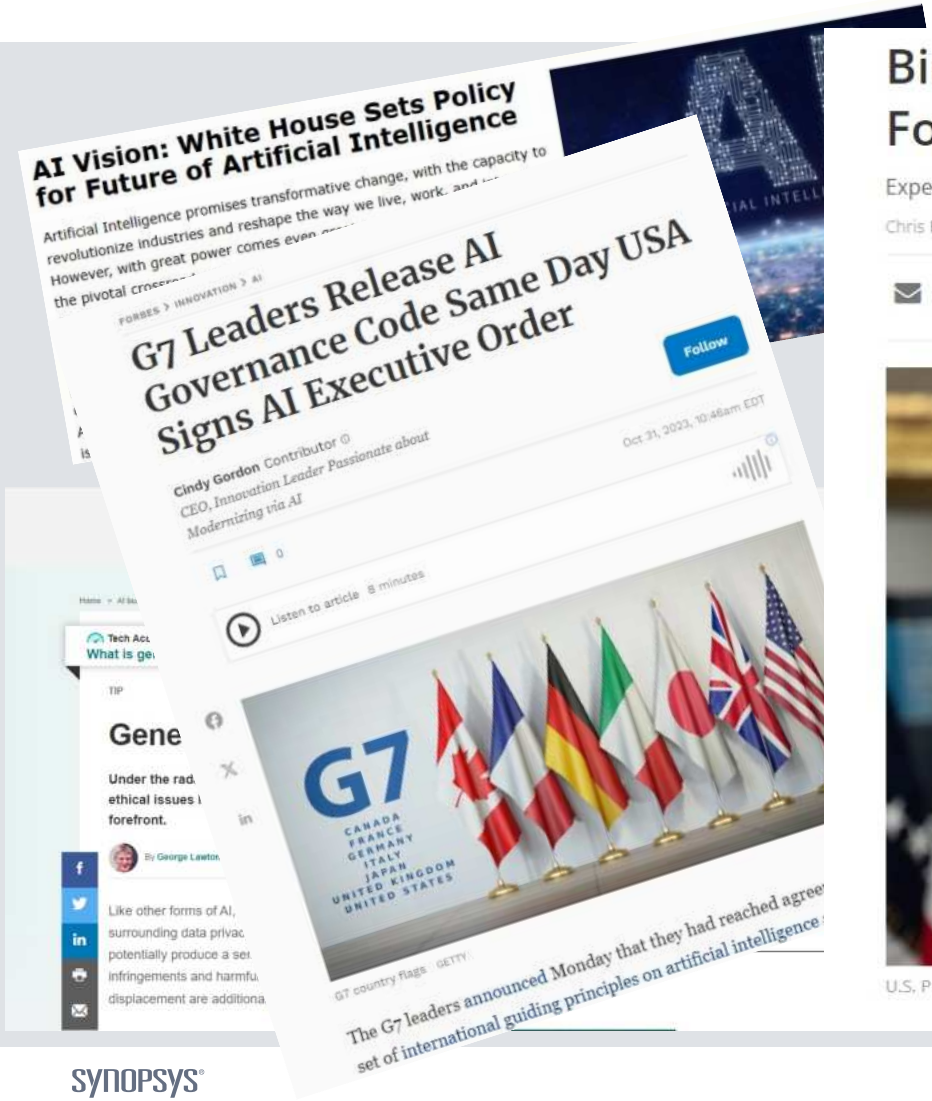


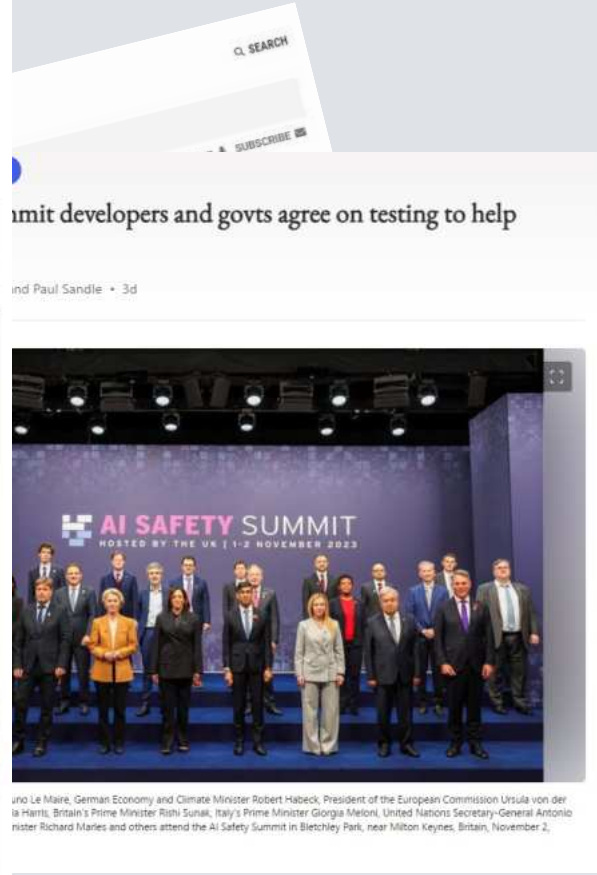
왼쪽의 AI가 생성한 코드가 오른쪽의 GNU 라이선스의 오픈소스와 동일함

AI를 활용한 애플리케이션



AI관련 뉴스 헤드라인



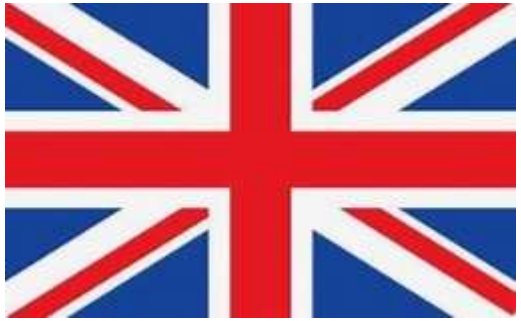


AI의 물결은 기대와 두려움을 불러옵니다!

AI 기대와 두려움

- **빠르게 변화하는 생성형 AI 기술**
  - ✓ ChatGPT, Copilot
- **AI 가이드 레일을 둘러싼 글로벌 토론**
  - ✓ 히로시마 AI 프로세스 - G7 회원국 리더
- **AI 거버넌스**
  - ✓ AI거버넌스 탐구를 위한 UN의 **새로운 자문 기구** (2023년 10월 26일)
  - ✓ 미국 대통령의 **AI 행정명령** (2023년 10월 20일)
  - ✓ 영국 **"국제 AI 서밋"** (2023년 11월 2일)



참조:

- U.S. President Joe Biden has issued an executive order (EO) (October 30th, 2023
- Hiroshima AI Process.pdf (mofa.go.jp) (May 2023 Japan)
- Commission welcomes G7 leaders' agreement (europa.eu) (May 2023)
- The United Nations announced a new AI advisory board (October 2023
- The AI Safety Summit 2023 - GOV.UK (November 2023)

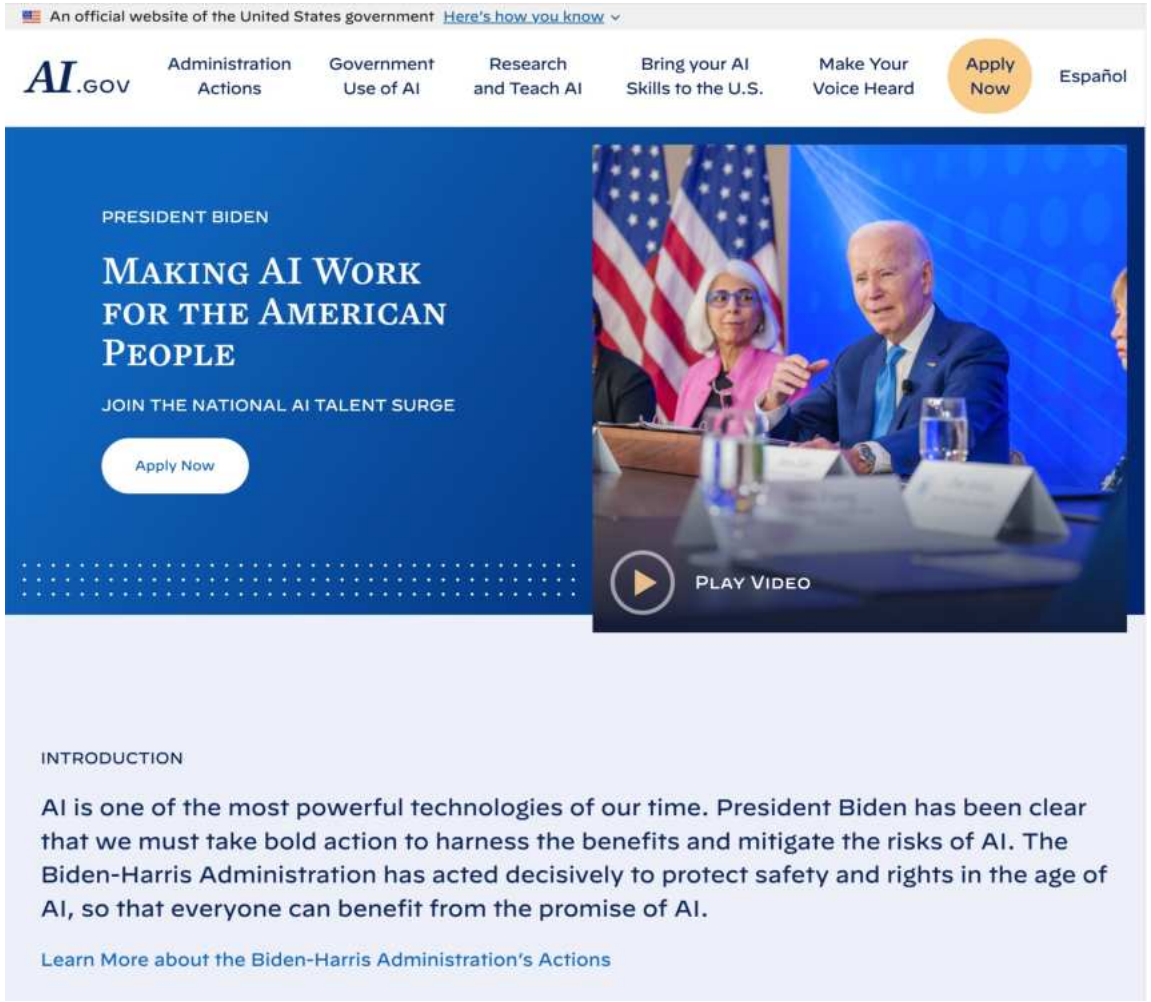
AI 행정명령?

> 행정 명령은 **AI 안전** 및 **보안**에 대한 **새로운 표준을 확립**하고, 미국인의 개인 정보를 보호하고, 형평성과 시민권을 증진하고, 소비자와 근로자를 옹호하고, **혁신과 경쟁을 촉진**하고, 전 세계에서 미국의 리더십을 발전시키는 등의 작업을 수행합니다.

참조: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

미국 행정부의 AI관련 활동 웹사이트

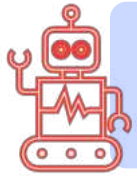
https://ai.gov



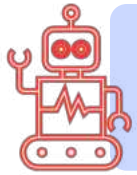
https://nairrpilot.org/

“안전하고 신뢰할 수 있는 AI”를 향한 여정

행정 명령의 일부 특징:

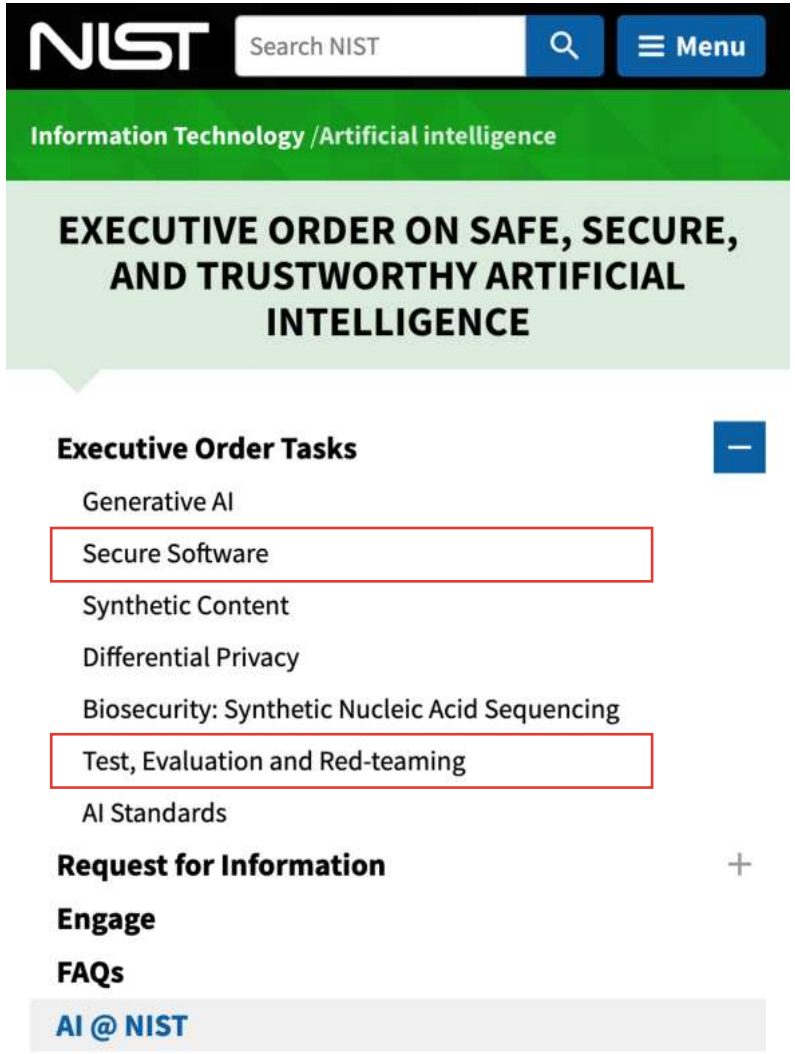


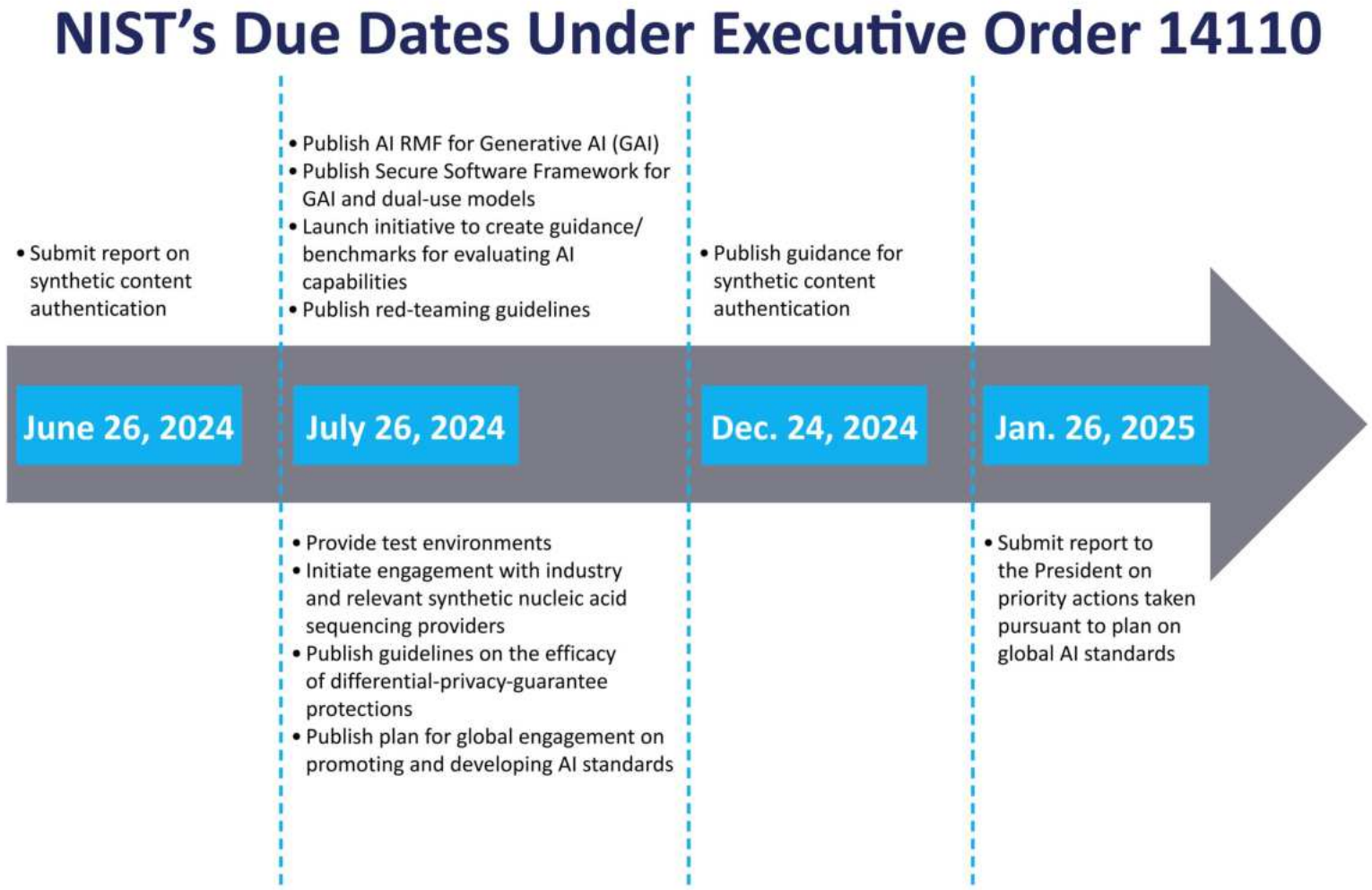
- 행정명령에는 “**강력한 AI 시스템의** 개발자는 **안전 테스트 결과를** 미국 정부와 **공유**해야 한다”고 명시
- 행정명령은 “**AI** 시스템의 **잠재적 위험**으로부터 **미국인을 보호**” 할 계획입니다.
- “새로운 **AI 안전** 및 **보안 기준**을 **국방 생산법(1950)**”에 맞춰 조정
- 또한, “의료, 정의 및 주택 분야의 **형평성**과 **시민권**, **차별** 및 편견”을 다루고 있습니다.
- **NIST**는 모델이 대중에게 공개될 수 있도록 수행해야 하는 **AI 테스팅 프레임워크**를 개발하고 있습니다.



참조: https://www.fema.gov/sites/default/files/2020-03/Defense_Production_Act_2018.pdf

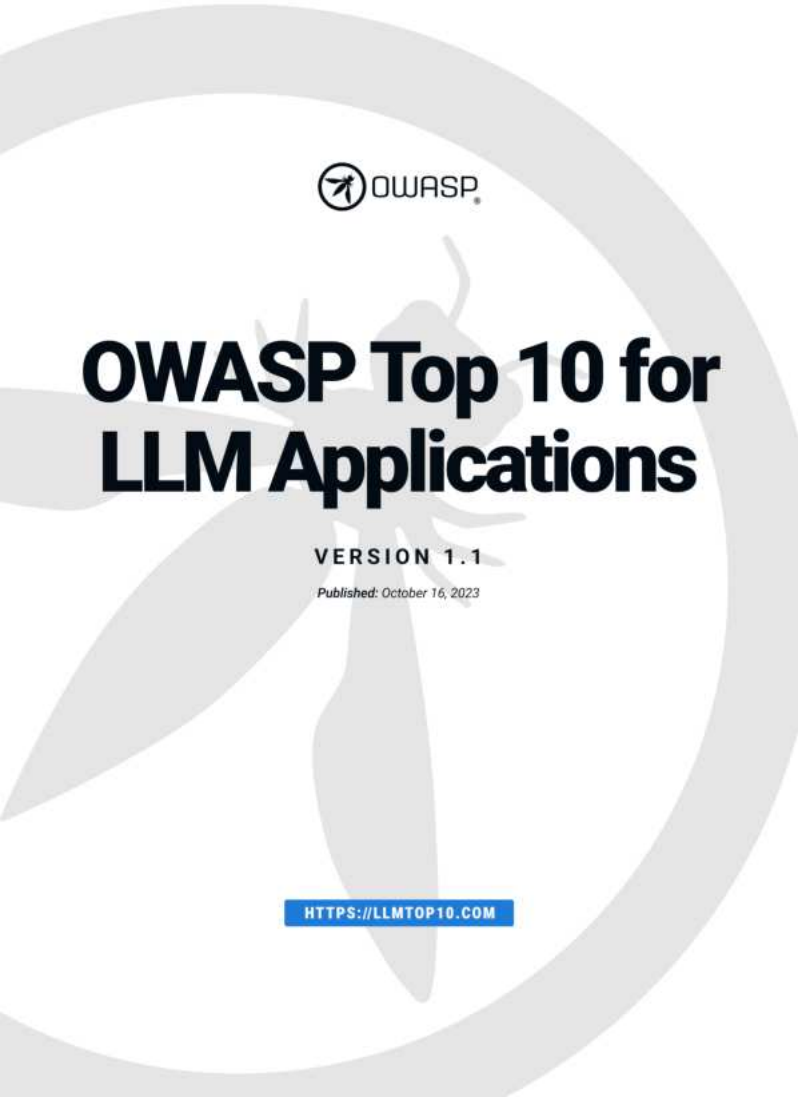
행정명령 14110 수행을 위한 NIST의 업무 및 일정





참조: https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence

OWASP Top 10 for LLM Applications



LLM01

Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02

Insecure Output Handling

This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.

LLM03

Training Data Poisoning

This occurs when LLM training data is tampered, introducing vulnerabilities or biases that compromise security, effectiveness, or ethical behavior. Sources include Common Crawl, WebText, OpenWebText, & books.

LLM04

Model Denial of Service

Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs.

LLM05

Supply Chain Vulnerabilities

LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre- trained models, and plugins can add vulnerabilities.

LLM06

Sensitive Information Disclosure

LLMs may inadvertently reveal confidential data in its responses, leading to unauthorized data access, privacy violations, and security breaches. It's crucial to implement data sanitization and strict user policies to mitigate this.

LLM07

Insecure Plugin Design

LLM plugins can have insecure inputs and insufficient access control. This lack of application control makes them easier to exploit and can result in consequences like remote code execution.

LLM08

Excessive Agency

LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.

LLM09

Overreliance

Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.

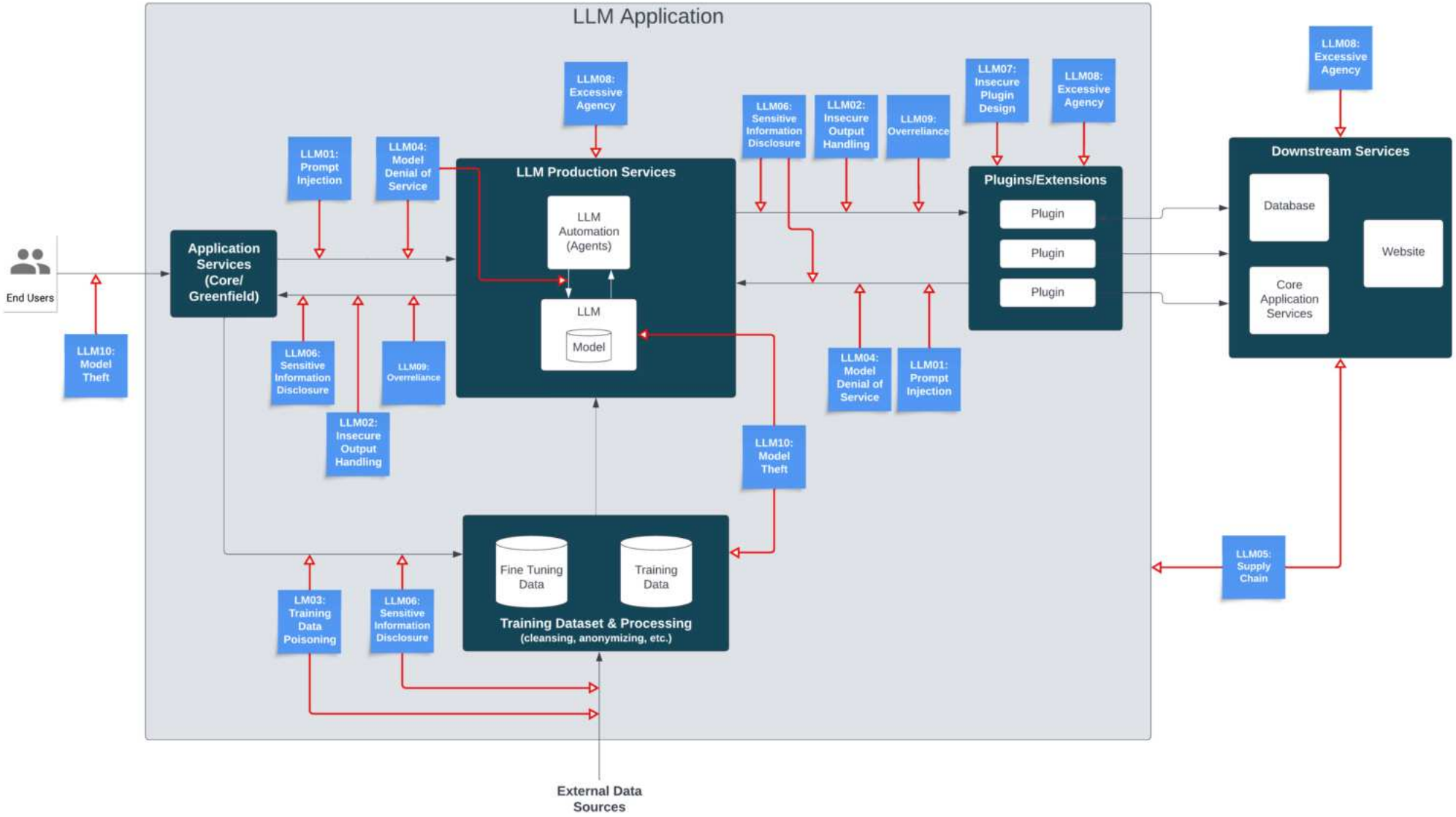
LLM10

Model Theft

This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information.

참조: https://owasp.org/www-project-top-10-for-large-language-model-applications/

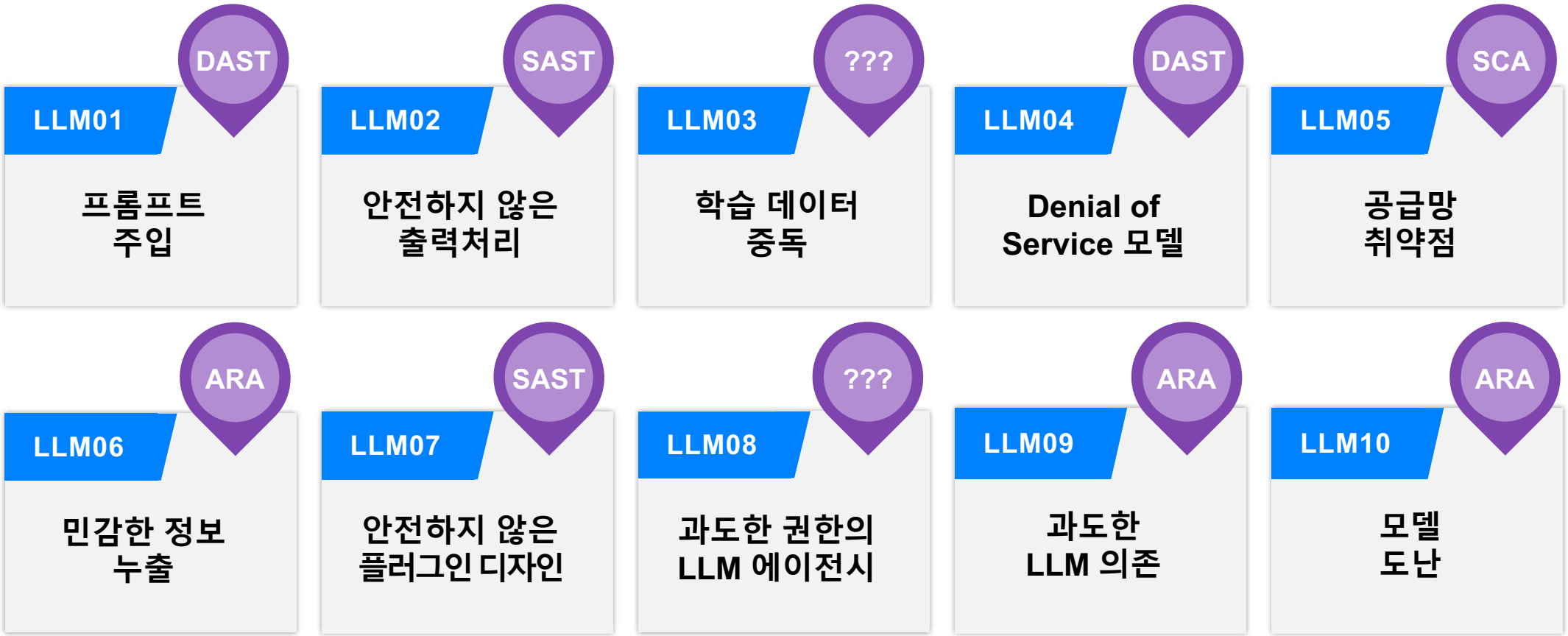
OWASP Top 10 for LLM Applications 데이터 흐름도



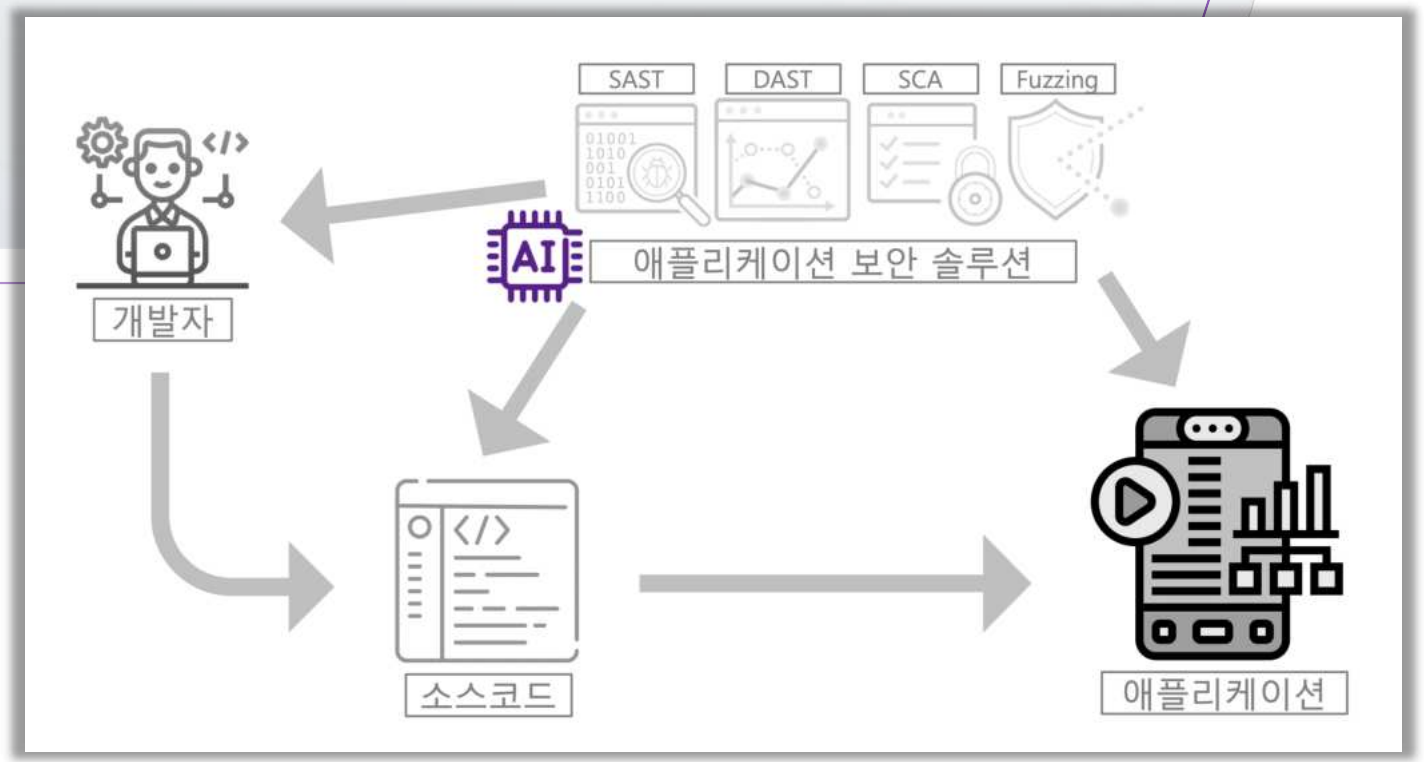
OWASP Top 10 for LLM Applications 항목

LLM01 프롬프트 주입	**LLM02** 안전하지 않은 출력처리	**LLM03** 학습 데이터 중독	**LLM04** Denial of Service 모델	**LLM05** 공급망 취약점
LLM06 민감한 정보 누출	**LLM07** 안전하지 않은 플러그인 디자인	**LLM08** 과도한 권한의 LLM 에이전시	**LLM09** 과도한 LLM 의존	**LLM10** 모델 도난

OWASP Top 10 for LLM Applications 항목별 예방조치



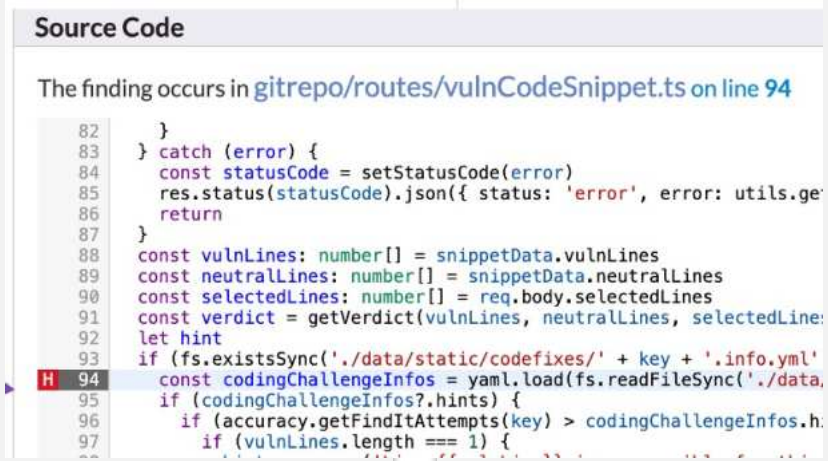
AI를 활용한 보안 솔루션



보안 솔루션에서 AI 활용 방안

EXPLANATION

왜 보안취약점이 발생하는 지에 대해 일반적인 설명이 아니라 해당 문제에 맞춤형 설명을 제공



FIX GUIDE

어떻게 보안취약점을 해결하는 지에 대해 가이드를 제공

FIX CODE

보안취약점을 해결한 코드를 생성하여 제시

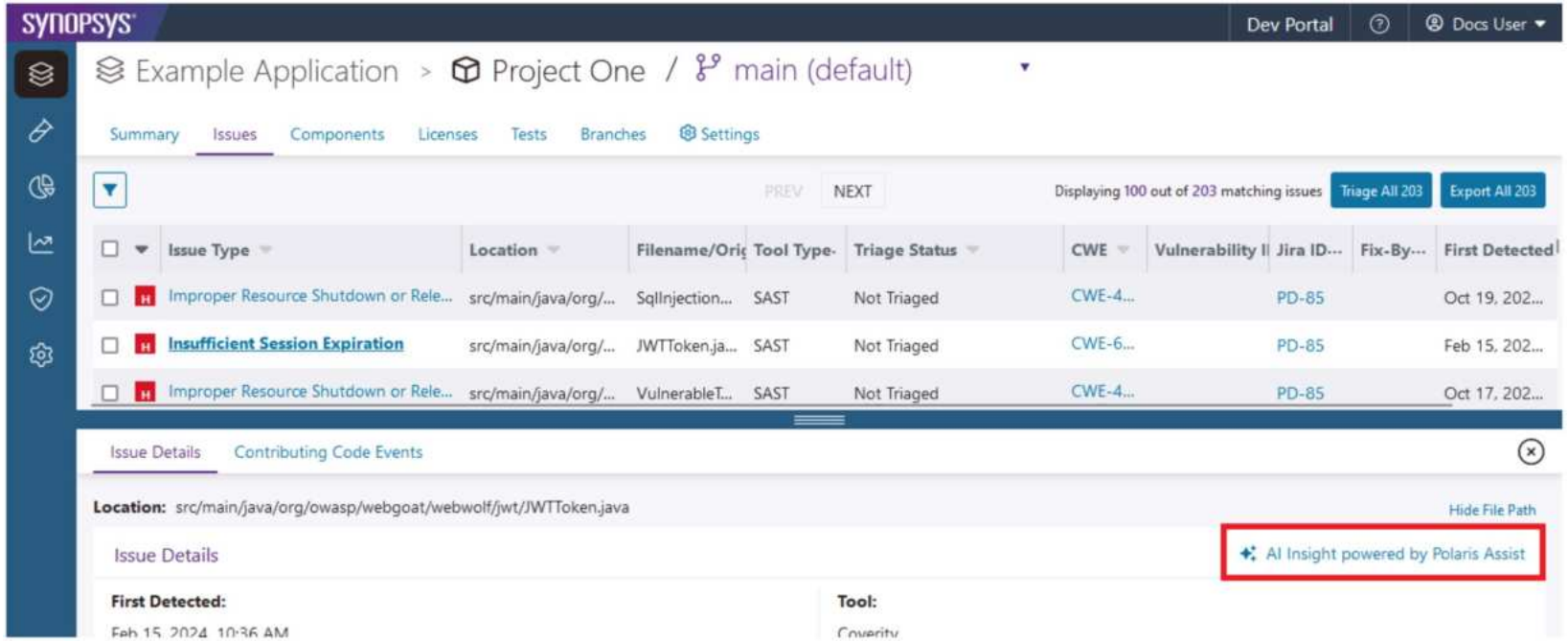
FP REMOVAL

애플리케이션 보안 도구에서 가장 문제가 되는 '오탐'을 AI를 활용하여 사전에 제거

CORRELATION

발견된 보안취약점 중에서 동일한 원인에 의해 발생한 문제들을 묶어서 하나의 아이템으로 표시

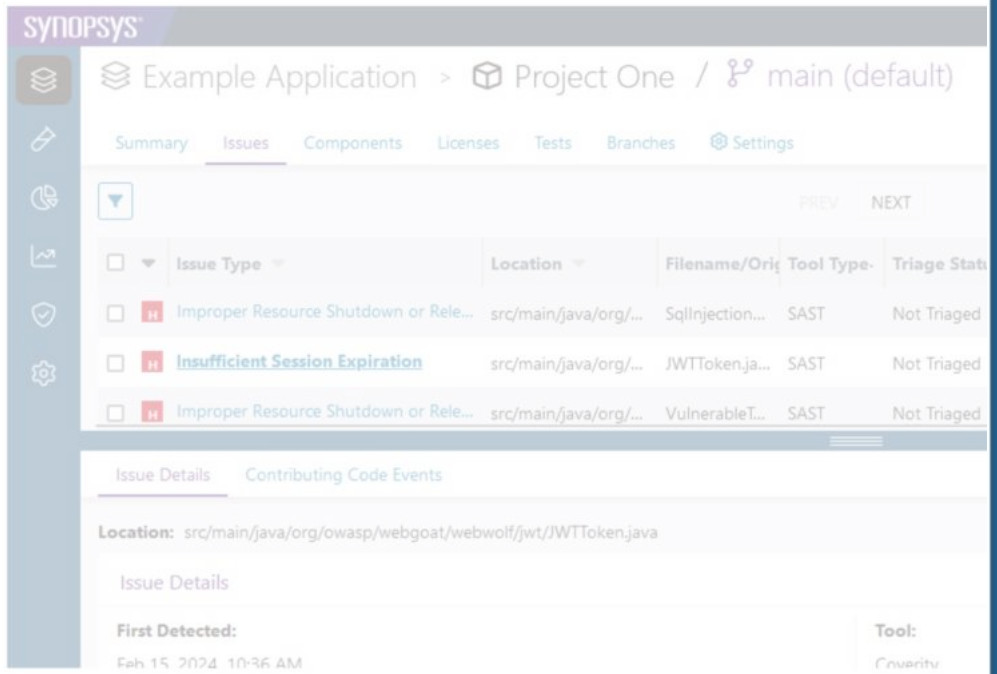
Polaris Assist: AI를 활용한 가이드 제공

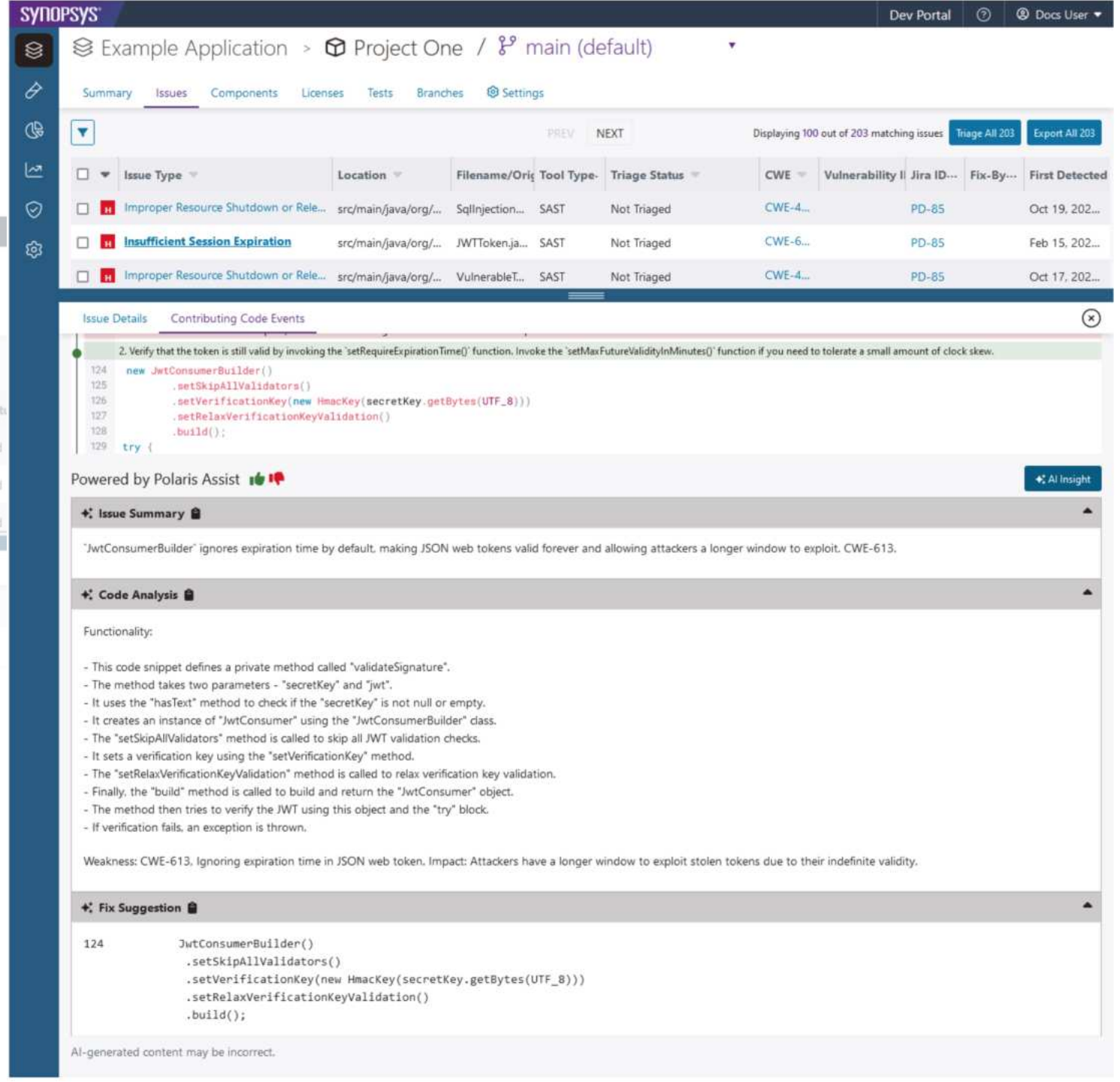


POLARIS
Software Integrity Platform®

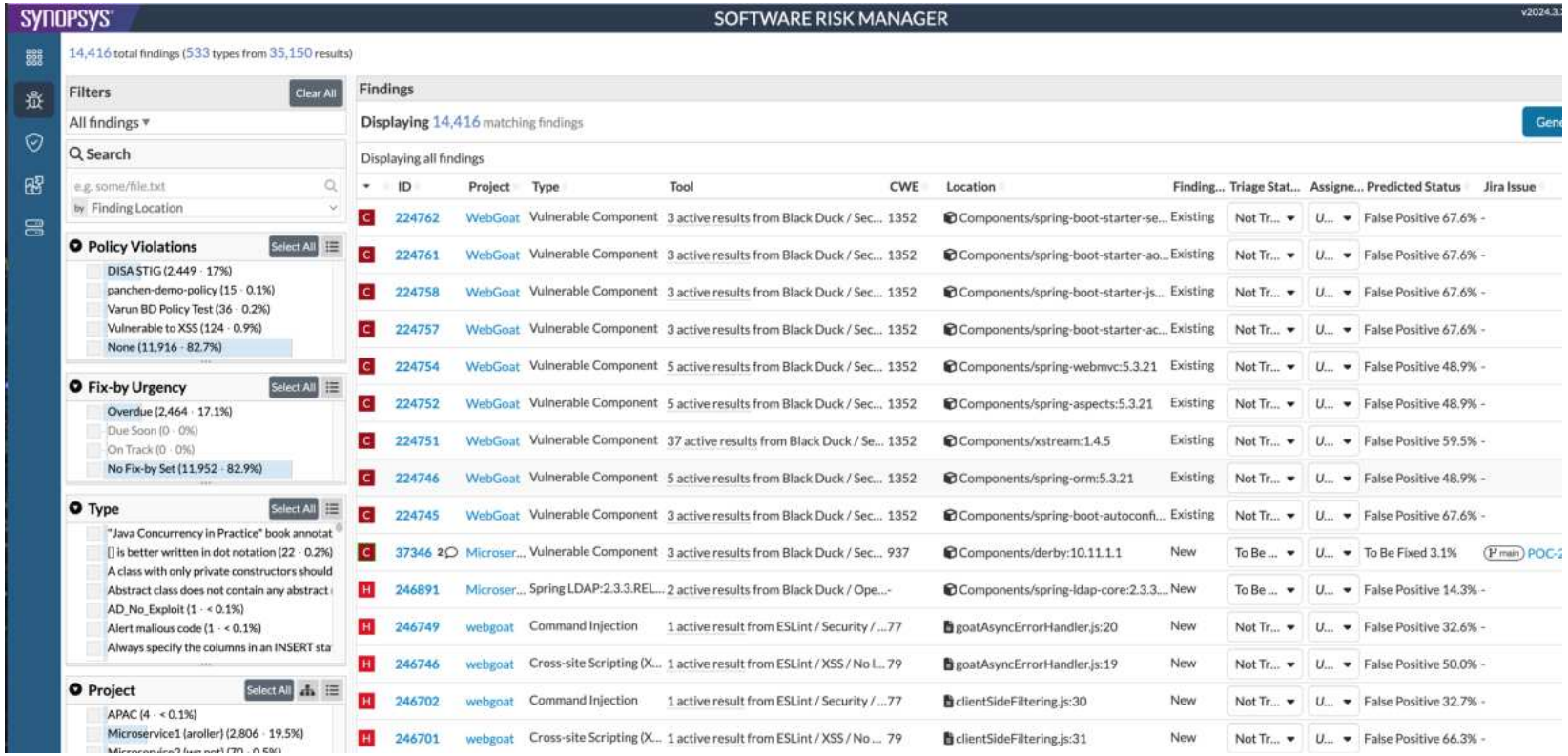
SaaS 기반으로
SAST, SCA, DAST
테스트를
통합 제공하는
Synopsys SIG의
차세대 솔루션

Polaris Assist





SRM: 오탐 확률 / 여러 도구의 결과 통합

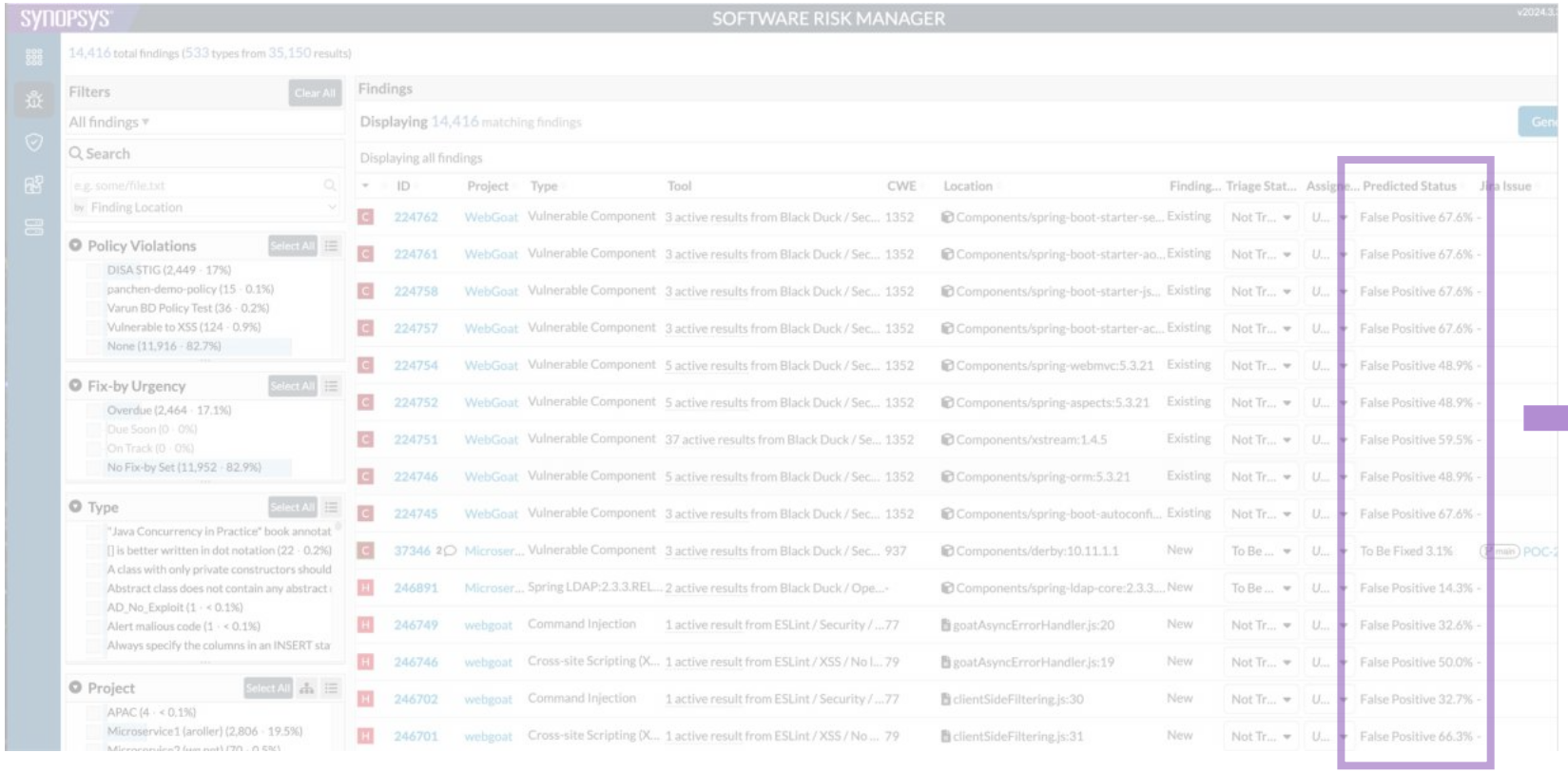


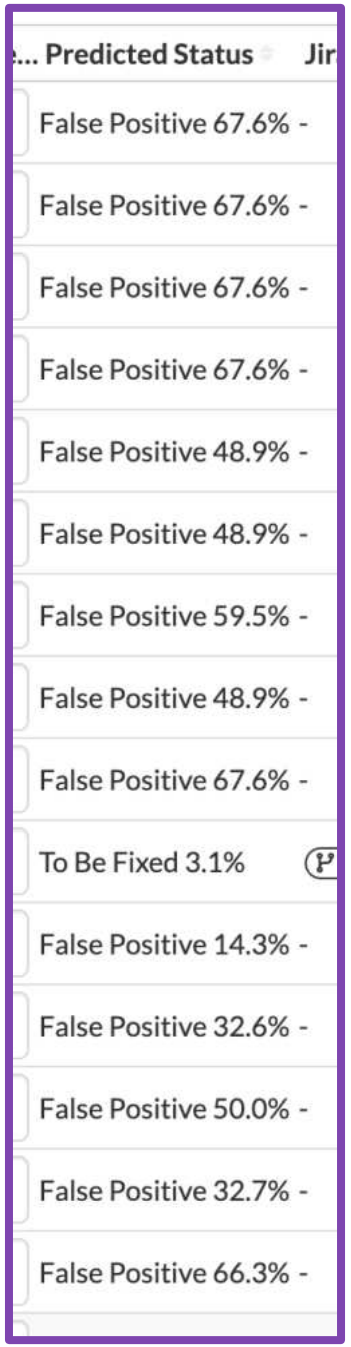
Software Risk Manager

여러 종류의
애플리케이션 보안
도구의 결과를
통합하여 보여주는
On-prem 기반의
ASPM 솔루션

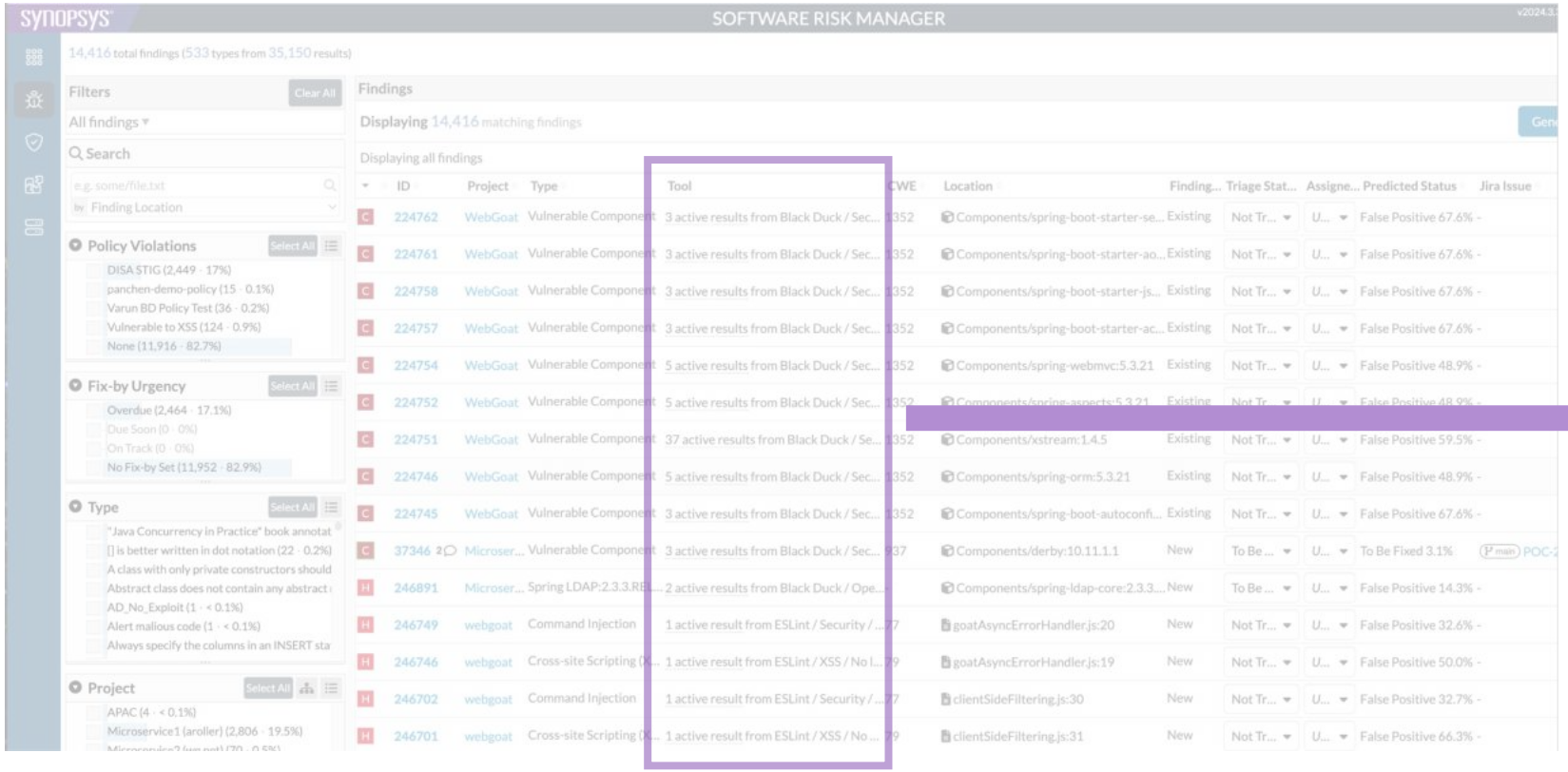
**** ASPM : Application Security Posture Management***

SRM: 오탐 확률

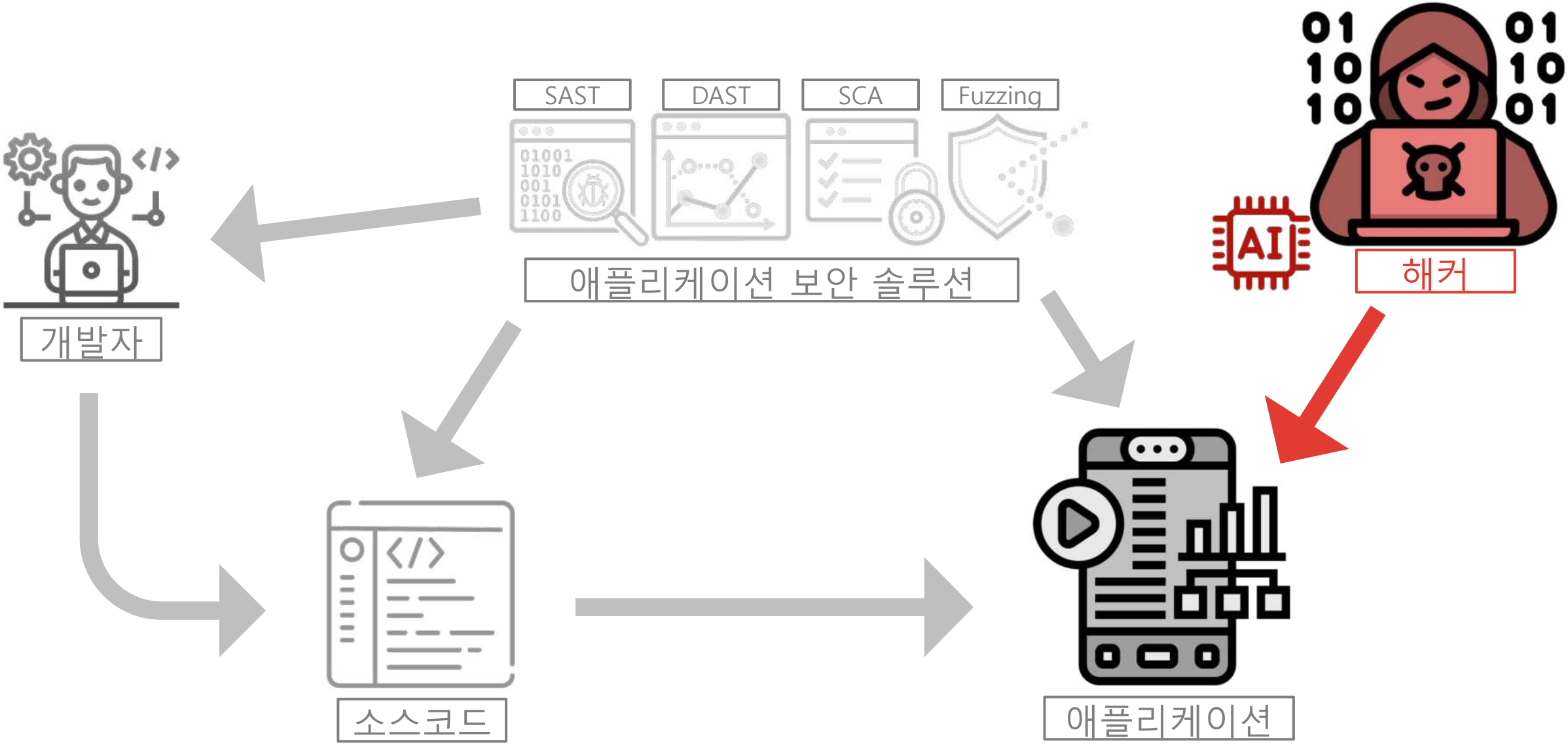




SRM: 여러 도구의 결과 통합



애플리케이션 보안에서
제일 많이 AI가 활성화 된 분야는…?



Summary

AI가 소프트웨어 개발 전반에서 활성화되는 시대,
애플리케이션 보안은 더욱 더 사이버보안을 위해
주요해 지고 있습니다.

다각적이면서 엄격하고 효율적인 애플리케이션
보안 활동과 이를 위한 애플리케이션 보안 솔루션이
필요합니다.

Thank You

SYNOPSYS®