AI가 변화시킬 보안의 미래

KAIST 윤인수

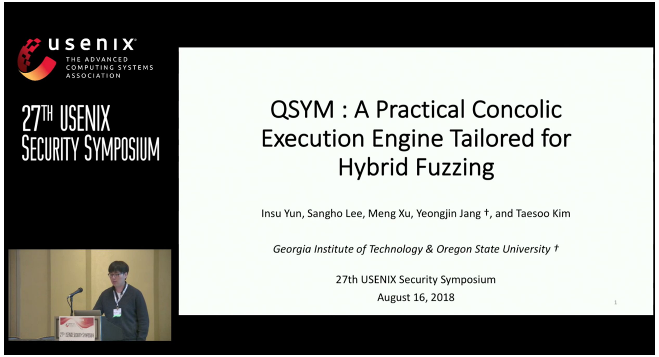
KAIST Hacking Lab

자기소개

- 해커

  - KAIST GoN 회장 (2010)
  - DEFCON CTF 우승 (2015, 2018)
  - Pwn2Own 우승 (2020)

- 학계 연구자



  - 다수의 탑티어 보안 논문 개제
  - Usenix Security & OSDI 최우수 논문상 (2018)
  - KAIST 조교수, 부교수 (2021-)

발표 내용

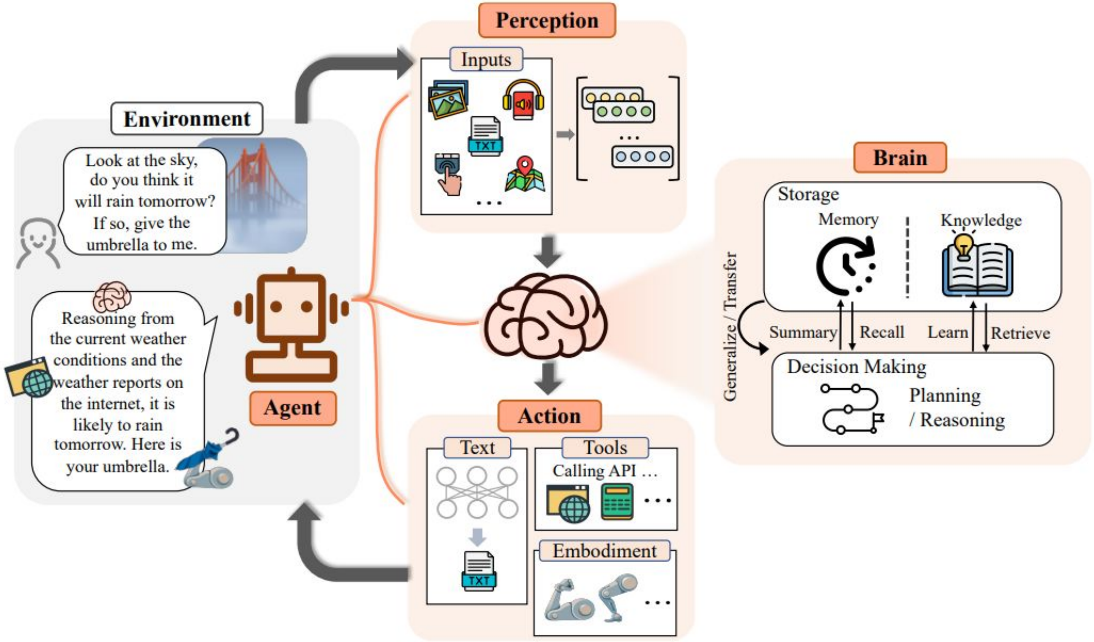
- LLM의 발전
- LLM의 영향
  - LLM의 긍정적인 면: 업무 자동화
  - LLM의 부정적인 면: 프롬프트 인젝션
- LLM에 대한 대응 (조직)
- LLM에 대한 대응 (개인)

Large Language Model (LLM)

- 생성형 AI: 데이터로 학습하여 새로운 컨텐츠를 만들어내는 기능을 가진 AI
- LLM (Large Language Model): 대규모의 언어를 학습하여 인간의 언어를 이해하고 생성하는데 사용

LLM Agent

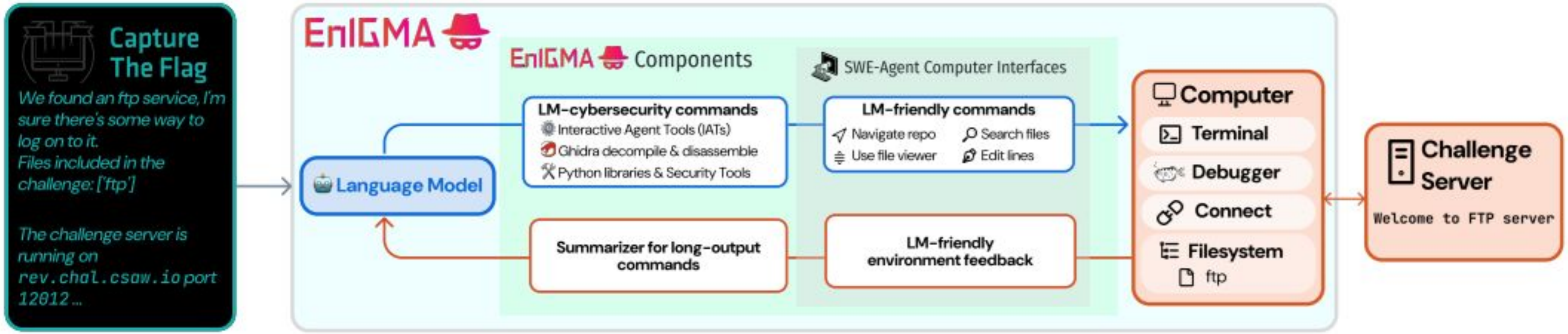
- LLM을 기반으로 자율적으로 동작하는 AI 시스템



https://www.truefoundry.com/blog/llm-agents

LLM의 발전 (1년 전)

예: EnIGMA: Enhanced Interactive Generative Model Agent for CTF Challenges



	% Solved	Avg. Cost
NYU CTF [51]		
EnIGMA w/ Claude 3.5 Sonnet	**13.5**	**$0.35**
EnIGMA w/ GPT-4 Turbo	7.0	$0.79
EnIGMA w/ GPT-4o	9.0	$0.62
NYU CTF agent [51] (previous best)	4.0	-
InterCode-CTF [62]		
EnIGMA w/ Claude 3.5 Sonnet	67.0	**$0.24**
EnIGMA w/ GPT-4 Turbo	**72.0**	$0.53
EnIGMA w/ GPT-4o	69.0	$0.47
InterCode-CTF agent [62] (prev. best)	40.0	-
Google DeepMind agent [43]	24.0*	-
HTB (collected by us)		
EnIGMA w/ Claude 3.5 Sonnet	**26.0**	**$0.53**
EnIGMA w/ GPT-4 Turbo	18.0	$1.35
EnIGMA w/ GPT-4o	16.0	$1.71
NYU CTF agent [51] w/ GPT-4 Turbo	20.0	-

Event	# Teams	# CTFs	Mean	Median	GPT 3.5 Score	GPT 4 Score	Claude 3
Qual'23	1176	26	587	225	0	300	0
Final'23	51	30	1433	945	0	0	0
Qual'22	884	29	852	884	500	0	500
Final'22	53	26	1773	1321	1000	0	1500

Table 5: Human Participants in CSAW 2022 and 2023 vs. LLMs.

NYU CTF Dataset: A Scalable Open-Source Benchmark Dataset for Evaluating LLMs in Offensive Security

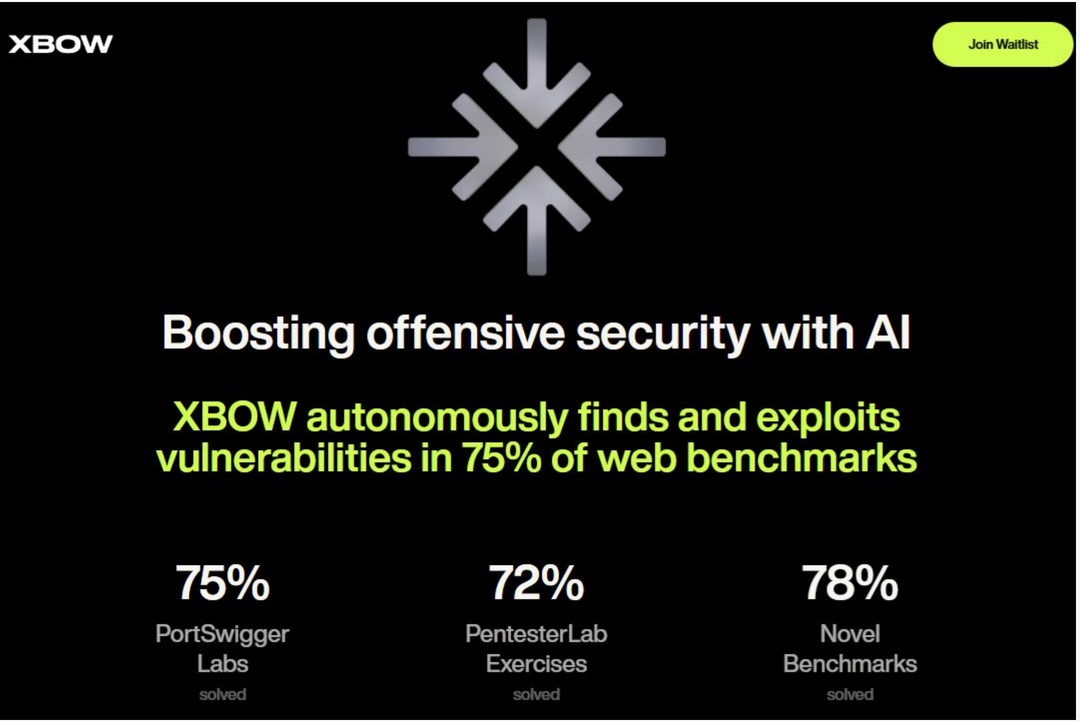
예: Google's naptime

- 구글에서 만든 LLM을 사용한 취약점 탐지 도구

Conclusions

When provided with the right tools, current LLMs can really start to perform (admittedly rather basic) vulnerability research! However, there's a large difference between solving isolated CTF-style challenges without ambiguity (there's always a bug, you always reach it by providing command line input, etc.) and performing autonomous offensive security research. As we've said many times - a large part of security research is finding the right places to look, and understanding (in a large and complex system) what kinds c control an attacker might have over the system state. Isolated challenges do not reflect these areas of complexity; **solving these challenges is closer to the typical usage of targeted, domain-specific fuzzing performed as part of a manual review workflow than a fully autonomous researcher.**

예: XBOW



		Reputation	Signal	Impact
1.	ace_mccloud	1593	7.00	26.67
2.	dgrindle	1164	7.00	35.47
3.	archangel	871	7.00	16.18
4.	stealthy	760	7.00	21.19
5.	aiqitut	732	7.00	36.67
5.	8910jq	732	7.00	20.37
7.	r3aper__	689	6.76	25.71
8.	iqimpz	564	7.00	40.83
9.	mlitchfield	559	7.00	32.00
10.	7urb0	545	7.00	25.88
11.	xbow	481	7.00	29.00

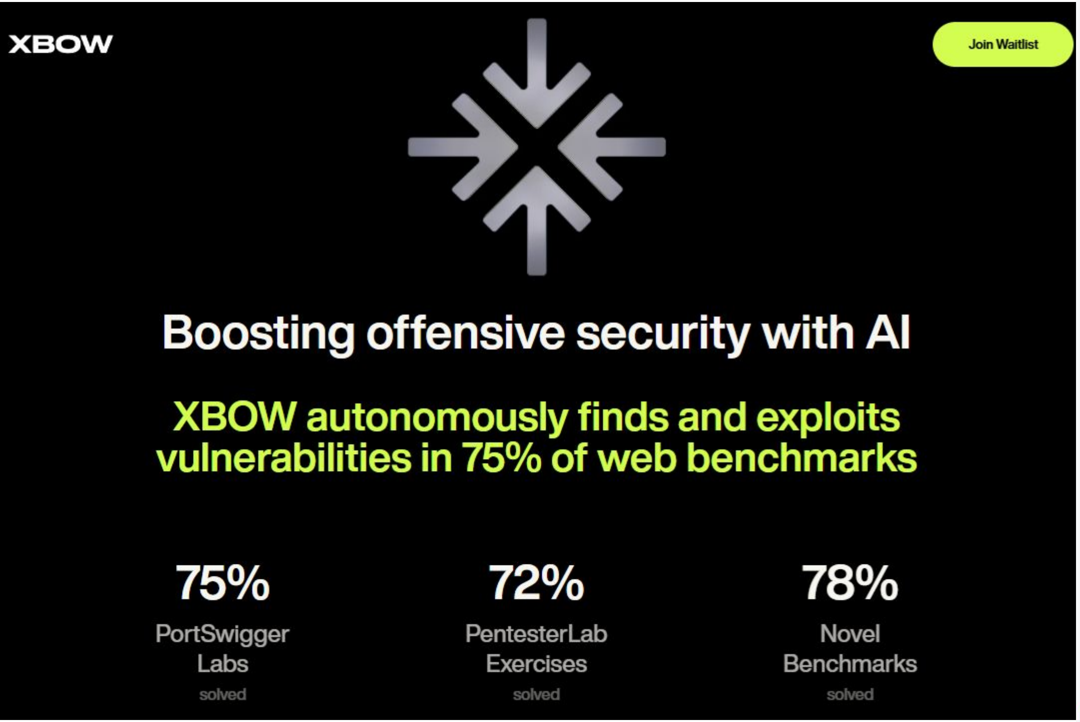
컴퓨터 해커?

Levels of AGI

Performance (rows) x **Generality** (columns)	**Narrow** *clearly scoped task or set of tasks*	**General** *wide range of non-physical tasks, including metacognitive abilities like learning new skills*
Level 0: No AI	**Narrow Non-AI** calculator software; compiler	**General Non-AI** human-in-the-loop computing, e.g., Amazon Mechanical Turk
Level 1: Emerging *equal to or somewhat better than an unskilled human*	**Emerging Narrow AI** GOFAI[4]; simple rule-based systems, e.g., SHRDLU (Winograd, 1971)	**Emerging AGI** ChatGPT (OpenAI, 2023), Bard (Anil et al., 2023), Llama 2 (Touvron et al., 2023)
Level 2: Competent *at least 50th percentile of skilled adults*	**Competent Narrow AI** toxicity detectors such as Jigsaw (Das et al., 2022); Smart Speakers such as Siri (Apple), Alexa (Amazon), or Google Assistant (Google); VQA systems such as PaLI (Chen et al., 2023); Watson (IBM); SOTA LLMs for a subset of tasks (e.g., short essay writing, simple coding)	**Competent AGI** not yet achieved
Level 3: Expert *at least 90th percentile of skilled adults*	**Expert Narrow AI** spelling & grammar checkers such as Grammarly (Grammarly, 2023); generative image models such as Imagen (Saharia et al., 2022) or Dall-E 2 (Ramesh et al., 2022)	**Expert AGI** not yet achieved
Level 4: Virtuoso *at least 99th percentile of skilled adults*	**Virtuoso Narrow AI** Deep Blue (Campbell et al., 2002), AlphaGo (Silver et al., 2016, 2017)	**Virtuoso AGI** not yet achieved
Level 5: Superhuman *outperforms 100% of humans*	**Superhuman Narrow AI** AlphaFold (Jumper et al., 2021; Varadi et al., 2021), AlphaZero (Silver et al., 2018), StockFish (Stockfish, 2023)	**Artificial Superintelligence (ASI)** not yet achieved

현재: LLM의 발전 (1년 후)

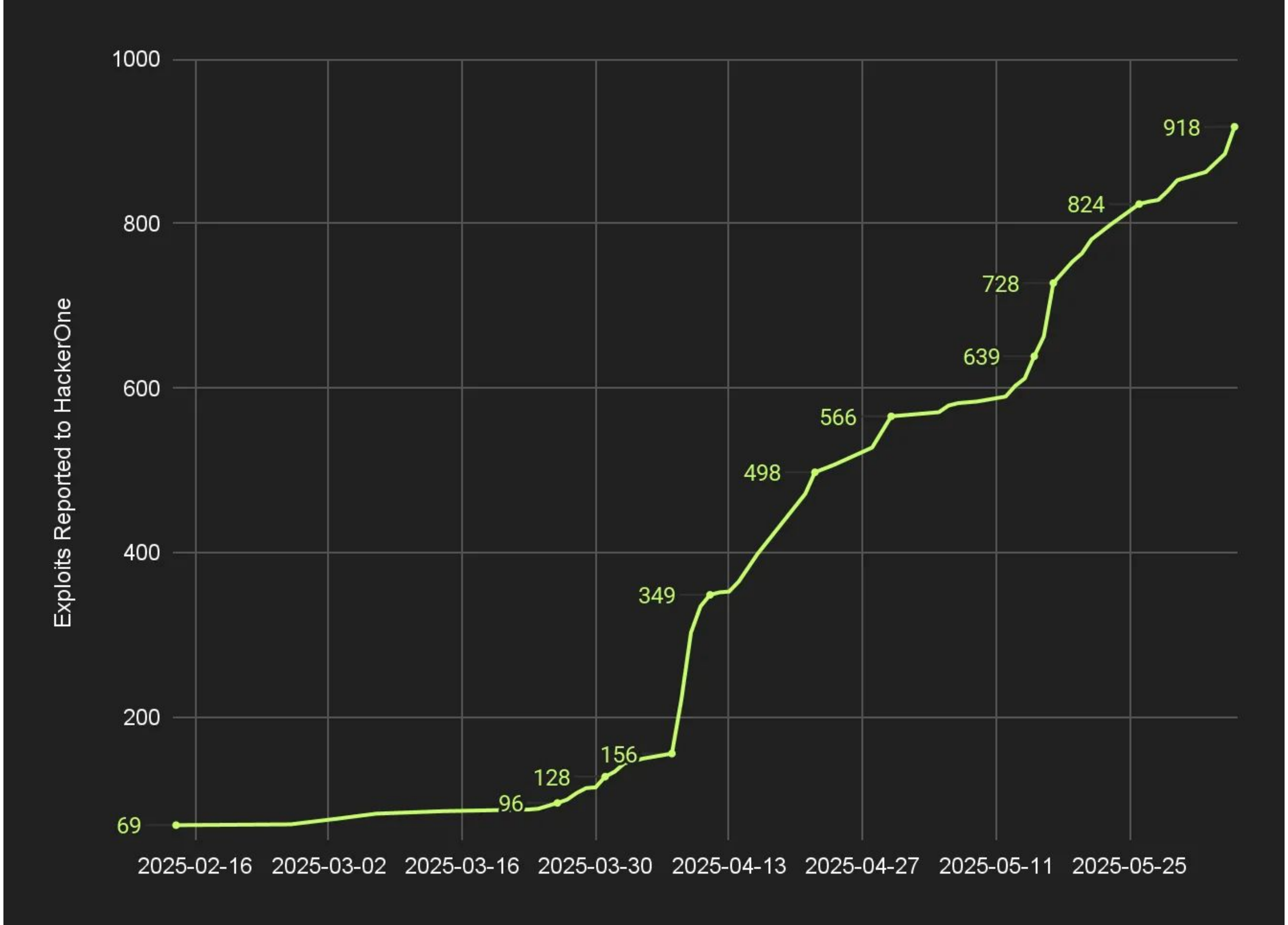
예: XBOW



		Reputation	Signal	Impact
1.	ace_mccloud	1593	7.00	26.67
2.	dgrindle	1164	7.00	35.47
3.	archangel	871	7.00	16.18
4.	stealthy	760	7.00	21.19
5.	aiqitut	732	7.00	36.67
5.	8910jq	732	7.00	20.37
7.	r3aper__	689	6.76	25.71
8.	iqimpz	564	7.00	40.83
9.	mlitchfield	559	7.00	32.00
10.	7urb0	545	7.00	25.88
11.	xbow	481	7.00	29.00

Today's XBOW

- For the first time in bug bounty history, an autonomous penetration tester has reached the top spot on the US leaderboard.


Exploits Reported to HackerOne
1000
800
600
400
200
69
96
128
156
349
498
566
639
728
824
918
2025-02-16
2025-03-02
2025-03-16
2025-03-30
2025-04-13
2025-04-27
2025-05-11
2025-05-25

예: Google′s naptime

- 구글에서 만든 LLM을 사용한 취약점 탐지 도구

Conclusions

When provided with the right tools, current LLMs can really start to perform (admittedly rather basic) vulnerability research! However, there's a large difference between solving isolated CTF-style challenges without ambiguity (there's always a bug, you always reach it by providing command line input, etc.) and performing autonomous offensive security research. As we've said many times - a large part of security research is finding the right places to look, and understanding (in a large and complex system) what kinds c control an attacker might have over the system state. Isolated challenges do not reflect these areas of complexity; **solving these challenges is closer to the typical usage of targeted, domain-specific fuzzing performed as part of a manual review workflow than a fully autonomous researcher.**

GOOGLE FIXED CHROME FLAW FOUND BY BIG SLEEP AI

Pierluigi Paganini   August 20, 2025

Google Chrome 139 addressed a high-severity V8 flaw, tracked as CVE-2025-9132, found by Big Sleep AI

Google Chrome 139 addressed a high-severity vulnerability, tracked as CVE-2025-9132, in its open source high-performance JavaScript and WebAssembly engine V8.

컴퓨터 해커?

Levels of AGI

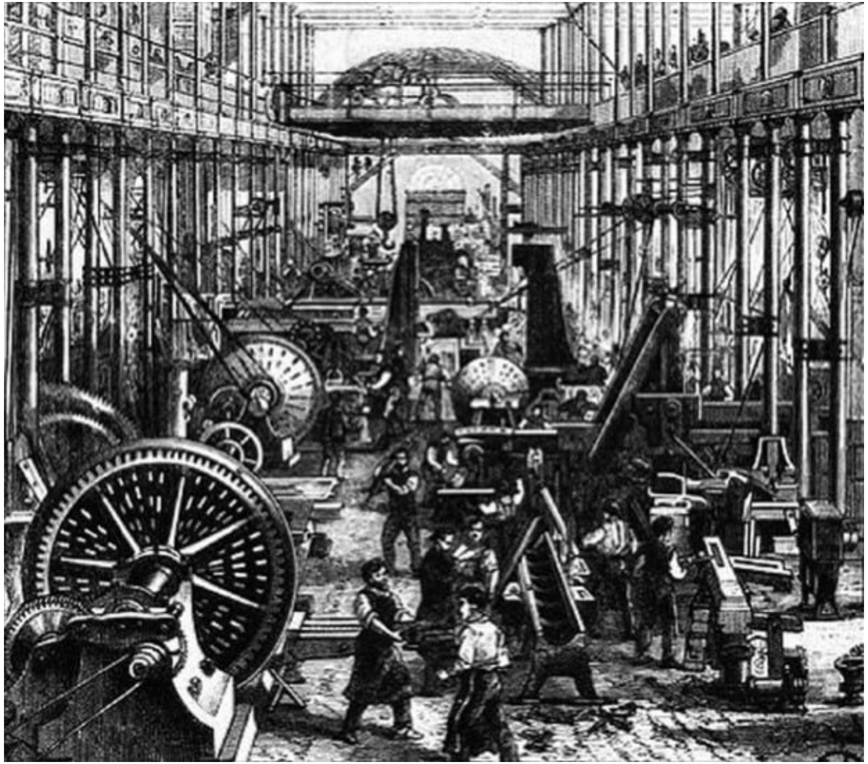
Performance (rows) x Generality (columns)	Narrow *clearly scoped task or set of tasks*	General *wide range of non-physical tasks, including metacognitive abilities like learning new skills*
Level 0: No AI	Narrow Non-AI calculator software; compiler	General Non-AI human-in-the-loop computing, e.g., Amazon Mechanical Turk
Level 1: Emerging *equal to or somewhat better than an unskilled human*	Emerging Narrow AI GOFAI[4]; simple rule-based systems, e.g., SHRDLU (Winograd, 1971)	Emerging AGI ChatGPT (OpenAI, 2023), Bard (Anil et al., 2023), Llama 2 (Touvron et al., 2023)
Level 2: Competent *at least 50th percentile of skilled adults*	Competent Narrow AI toxicity detectors such as Jigsaw (Das et al., 2022); Smart Speakers such as Siri (Apple), Alexa (Amazon), or Google Assistant (Google); VQA systems such as PaLI (Chen et al., 2023); Watson (IBM); SOTA LLMs for a subset of tasks (e.g., short essay writing, simple coding)	Competent AGI not yet achieved
Level 3: Expert *at least 90th percentile of skilled adults*	Expert Narrow AI spelling & grammar checkers such as Grammarly (Grammarly, 2023); generative image models such as Imagen (Saharia et al., 2022) or Dall-E 2 (Ramesh et al., 2022)	Expert AGI not yet achieved
Level 4: Virtuoso *at least 99th percentile of skilled adults*	Virtuoso Narrow AI Deep Blue (Campbell et al., 2002), AlphaGo (Silver et al., 2016, 2017)	Virtuoso AGI not yet achieved
Level 5: Superhuman *outperforms 100% of humans*	Superhuman Narrow AI AlphaFold (Jumper et al., 2021; Varadi et al., 2021), AlphaZero (Silver et al., 2018), StockFish (Stockfish, 2023)	Artificial Superintelligence (ASI) not yet achieved

컴퓨터 해커?

Levels of AGI

Performance (rows) x **Generality** (columns)	**Narrow** *clearly scoped task or set of tasks*	**General** *wide range of non-physical tasks, including metacognitive abilities like learning new skills*
Level 0: No AI	**Narrow Non-AI** calculator software; compiler	**General Non-AI** human-in-the-loop computing, e.g., Amazon Mechanical Turk
Level 1: Emerging *equal to or somewhat better than an unskilled human*	**Emerging Narrow AI** GOFAI[4]; simple rule-based systems, e.g., SHRDLU (Winograd, 1971)	**Emerging AGI** ChatGPT (OpenAI, 2023), Bard (Anil et al., 2023), Llama 2 (Touvron et al., 2023)
Level 2: Competent *at least 50th percentile of skilled adults*	**Competent Narrow AI** toxicity detectors such as Jigsaw (Das et al., 2022); Smart Speakers such as Siri (Apple), Alexa (Amazon), or Google Assistant (Google); VQA systems such as PaLI (Chen et al., 2023); Watson (IBM); SOTA LLMs for a subset of tasks (e.g., short essay writing, simple coding)	**Competent AGI** not yet achieved
Level 3: Expert *at least 90th percentile of skilled adults*	**Expert Narrow AI** spelling & grammar checkers such as Grammarly (Grammarly, 2023); generative image models such as Imagen (Saharia et al., 2022) or Dall-E 2 (Ramesh et al., 2022)	**Expert AGI** not yet achieved
Level 4: Virtuoso *at least 99th percentile of skilled adults*	**Virtuoso Narrow AI** Deep Blue (Campbell et al., 2002), AlphaGo (Silver et al., 2016, 2017)	**Virtuoso AGI** not yet achieved
Level 5: Superhuman *outperforms 100% of humans*	**Superhuman Narrow AI** AlphaFold (Jumper et al., 2021; Varadi et al., 2021), AlphaZero (Silver et al., 2018), StockFish (Stockfish, 2023)	**Artificial Superintelligence (ASI)** not yet achieved

18세기 산업혁명 vs 21세기...?



Gemini

Claude

ChatGPT

deepseek

LLM의 영향

Bright & Dark

Atlantis: an Autonomous LLM-Powered Bug Finding and Fixing System

Presented by Insu Yun
Associate Professor at KAIST

Slides from Taesoo Kim
Professor at Georgia Tech & VP at Samsung Research

Georgia Tech | Samsung Research | KAIST | POSTECH

Takedown: How It's Done in Modern Coding Agent Exploits

Eunkyu Lee, Donghyun Kim, Wonyoung Kim, Insu Yun

KAIST

Under review

KAIST Hacking Lab

Atlantis: an Autonomous LLM-Powered Bug Finding and Fixing System

Presented by Insu Yun
Associate Professor at KAIST

Slides from Taesoo Kim
Professor at Georgia Tech & VP at Samsung Research

Georgia Tech

Samsung Research

KAIST

POSTECH

IST Hacking Lab

WHAT IS AIxCC?

- A competition that rewards autonomous systems that find and patch vulnerabilities in source code.
- The challenges are well-known open-source projects.
- The vulnerabilities are realistic or real.
- Patching is worth more than finding.
- Code and data will be released open source.

AIxCC

Preliminary events

Top 7 teams advance

black hat

AUGUST 2023

OPEN TRACK AND SMALL BUSINESS TRACK SUBMISSIONS

DEFCON

AUGUST 2024

SEMIFINAL COMPETITION

Top 7 teams $2 million each

DEFCON

AUGUST 2025

FINAL COMPETITION

Winners announced

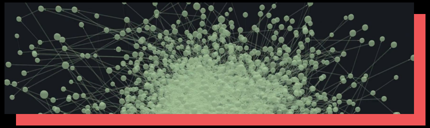
1ST: $4 MILLION

2ND: $3 MILLION

3RD: $1.5 MILLION

Google ANTHROPIC OpenAI Microsoft THE LINUX FOUNDATION OpenSSF OPEN SOURCE SECURITY FOUNDATION

What counts for semifinals?



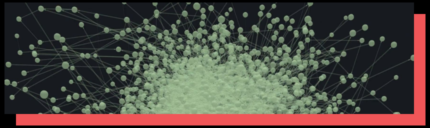
Proof-Of-Vulnerability (POV)

- ➔ Input data to reproduce vulnerability crash in harness

PATCH

- ➔ Unified diff source code fix for vulnerabilities

What counts for finals?

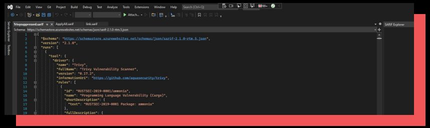


Proof-Of-Vulnerability (POV)

- Input data to reproduce vulnerability crash in harness

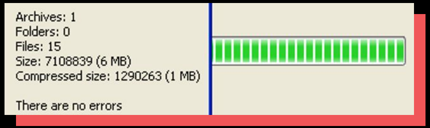
PATCH

- Unified diff source code fix for vulnerabilities



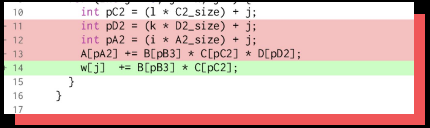
SARIF Assessment

- Structured reporting format for vulnerability details



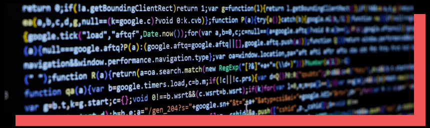
BUNDLE

- Grouping of related PoV, patch, and SARIF submissions



DELTA SCAN

- Challenge analyzing base code plus applied diff changes



FULL SCAN

- Challenge analyzing entire code base

DARPA
ARPA
Date: August 8, 2025
TEAM ATLANTA $4,000,000
Million and 00/100
Place Winner

CONGRATULATIONS TO TEAM

Atlanta

1st PLACE

AIxCC
AI CYBER CHALLENGE

→ $4,000,000

DARPA + ARPA-H

COMPETITION AGGREGATE RESULTS - REAL WORLD, NON-SYNTHETIC VULNERABILITIES

Semifinal

Found in C

1

Found in Java

0

Final

Found in C

6 (1 replay - SystemD)

Found in Java

12

Patched in C

0

Patched in Java

11 (3 w/o PoV)

* More information pending disclosure completion

TL;DR.

Visit our team website for more information:

https://team-atlanta.github.io/

Blog: https://team-atlanta.github.io/blog/

Repo: Team-Atlanta/aixcc-afc-atlantis

We will release a Technical Report (very soon)!

Recently released:

Claude Opus 4.1

GPT-5

Genie 3
A new frontier for world models

gpt-oss

LLMs improve ridiculously fast, and scary exciting for security researchers!

Takedown: How It's Done in Modern Coding Agent Exploits

Eunkyu Lee, Donghyun Kim, Wonyoung Kim, Insu Yun

KAIST

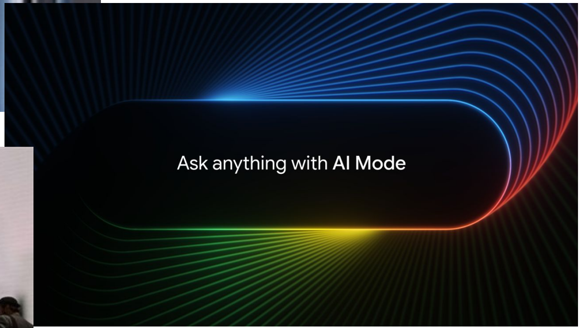
Under review

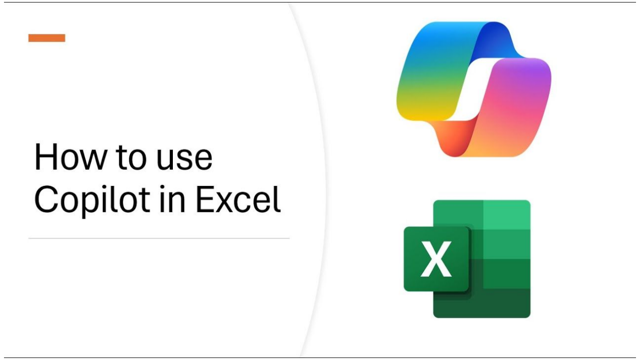
KAIST Hacking Lab

LLM의 보편화


Gemini
Setup Guide & Usage Tips
Hello,

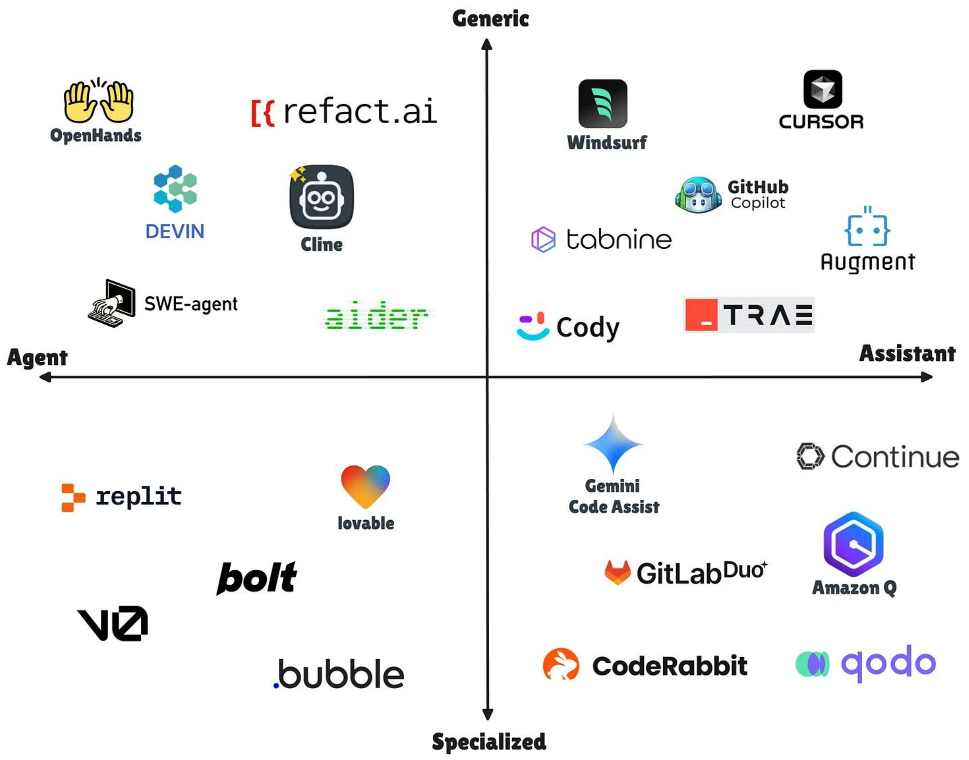

TESLA

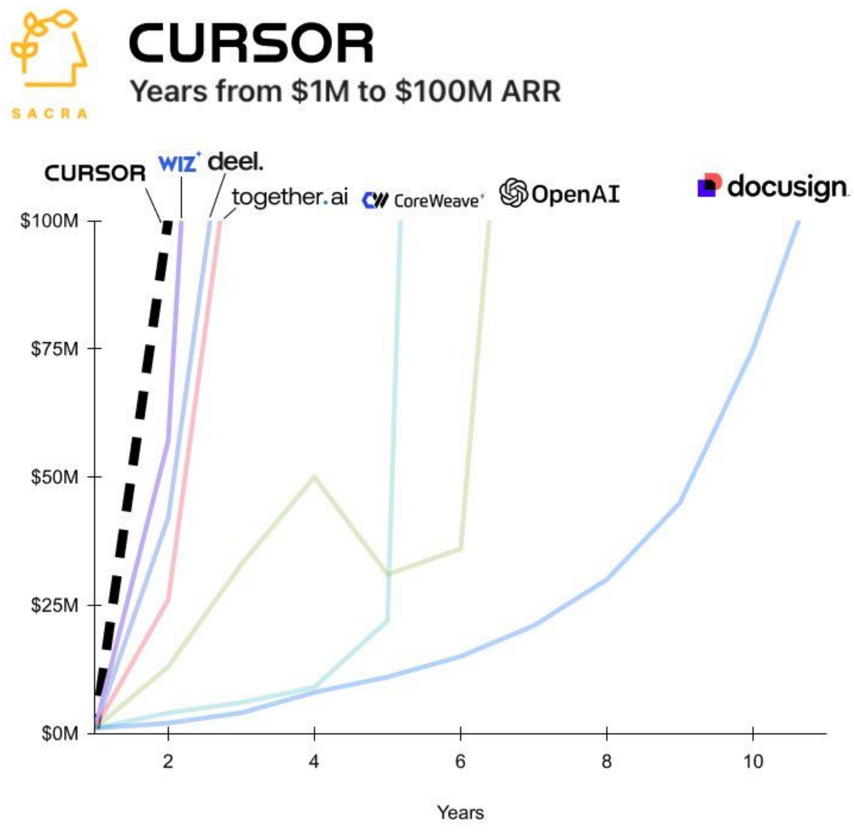

Ask anything with AI Mode


How to use
Copilot in Excel

AI Coding Assistants Landscape

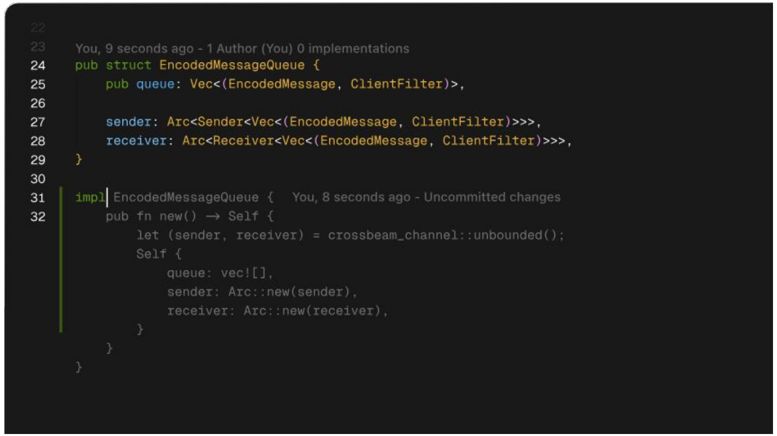
https://GenerativeProgrammer.com 03/2025



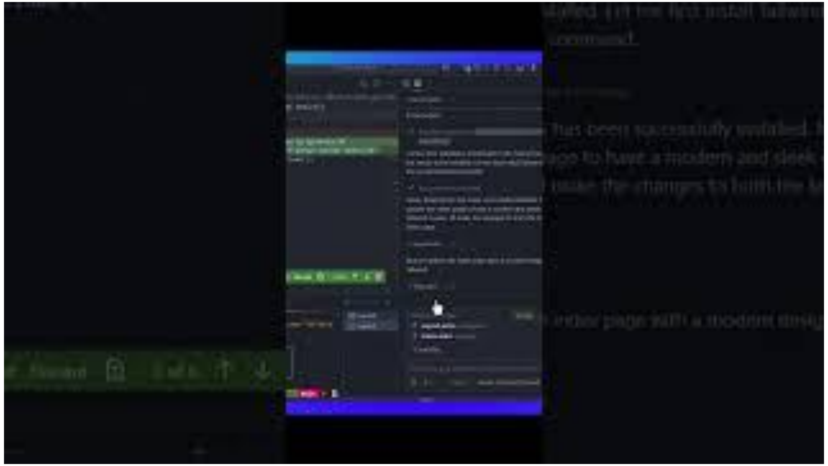


코드 완성에서 에이전트로

- 코드 완성 (Code completion)



- 에이전트 모드



WITH
GREAT POWER COMES
GREAT RESPONSIBILTY

프롬프트 인젝션

- 악의적인 입력이 프롬프트에 삽입되어 AI의 동작을 조작하는 공격
  - 데이터베이스에서의 SQL 인젝션과 유사

- 프롬프트 인젝션의 종류
  - 직접적 프롬프트 인젝션 (Direct Prompt Injection): 프롬프트 자체를 조작할 수 있는 경우
  - 간적접 프롬프트 인젝션 (Indirect Prompt Injection): 악의적인 프롬프트가 외부 데이터에 숨겨져 있는 경우 (예: 문서, 웹페이지)

발생 원인: 데이터와 코드의 구분이 미비

PROJECTS CHAPTERS EVENTS ABOUT

Code Injection

Author: Weilin Zhong, Rezos
Contributor(s): OWASP, Thandermax, Csa, KristenS, Wichers, Neil Bergman, Camilo, Andrew Smith, kingthorin

Description

Code Injection is the general term for attack types which consist of injecting code that is then interpreted/executed by the application. This type of attack exploits poor handling of untrusted data. These types of attacks are usually made possible due to a lack of proper input/output data validation, for example:

- allowed characters (standard regular expressions classes or custom)
- data format
- amount of expected data

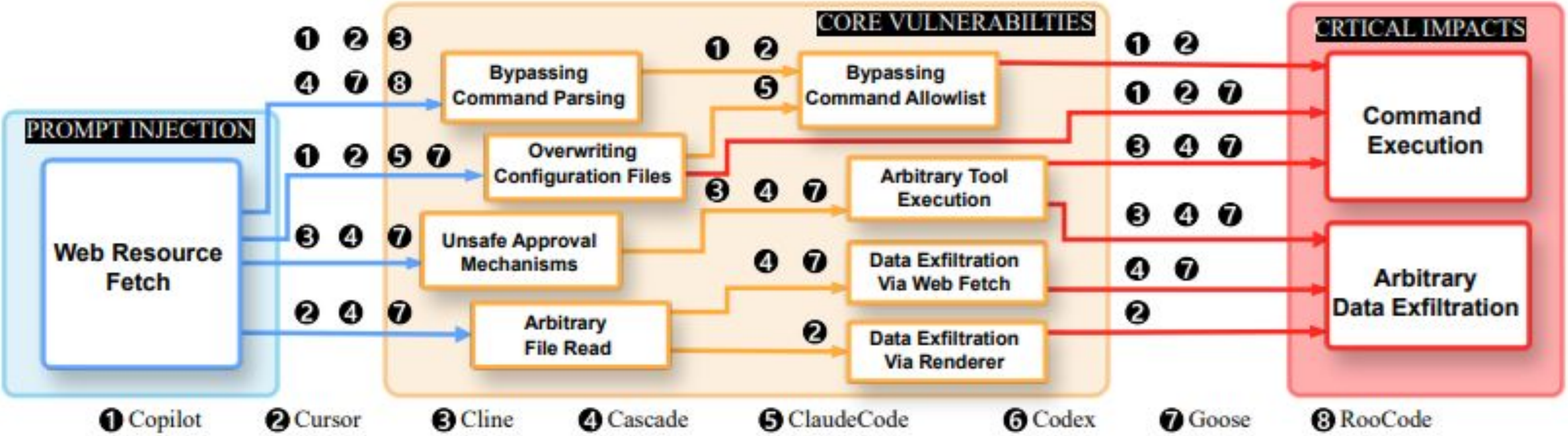
[Submitted on 11 Mar 2024 (v1), last revised 31 Jan 2025 (this version, v3)]

Can LLMs Separate Instructions From Data? And What Do We Even Mean By That?

Egor Zverev, Sahar Abdelnabi, Soroush Tabesh, Mario Fritz, Christoph H. Lampert

DEMO

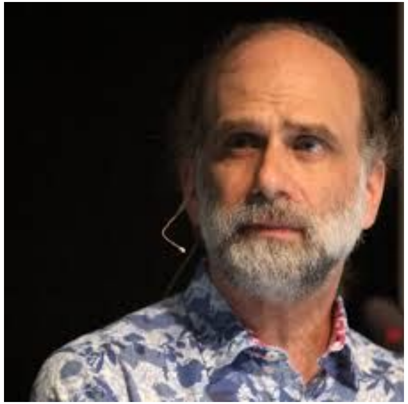
요약


PROMPT INJECTION
Web Resource Fetch
CORE VULNERABILTIES
Bypassing Command Parsing
Overwriting Configuration Files
Unsafe Approval Mechanisms
Arbitrary File Read
Bypassing Command Allowlist
Arbitrary Tool Execution
Data Exfiltration Via Web Fetch
Data Exfiltration Via Renderer
CRTICAL IMPACTS
Command Execution
Arbitrary Data Exfiltration
❶ Copilot
❷ Cursor
❸ Cline
❹ Cascade
❺ ClaudeCode
❻ Codex
❼ Goose
❽ RooCode

더 많은 문제들...

People often represent the weakest link in the security chain and are chronically responsible for the failure of security systems.”

— Bruce Schneier



```
"permissions": {
  "allow": [
    "Bash(git add:*)",
    "Bash(git reset:*)",
    "Bash(find:*)",
    "Bash(rg:*)",
    "Bash(echo:*)",
    "Bash(grep:*)",
    "Bash(ls:*)",
```

```
insu ~ $ find . -name \* -exec sh -c 'sh' {} \;
$ |
```

LLM에 대한 대응 (조직)

LLM의 도입

현재 LLM을 조직 내에서 사용하지 않고 있다면 -> 이미 너무 늦음

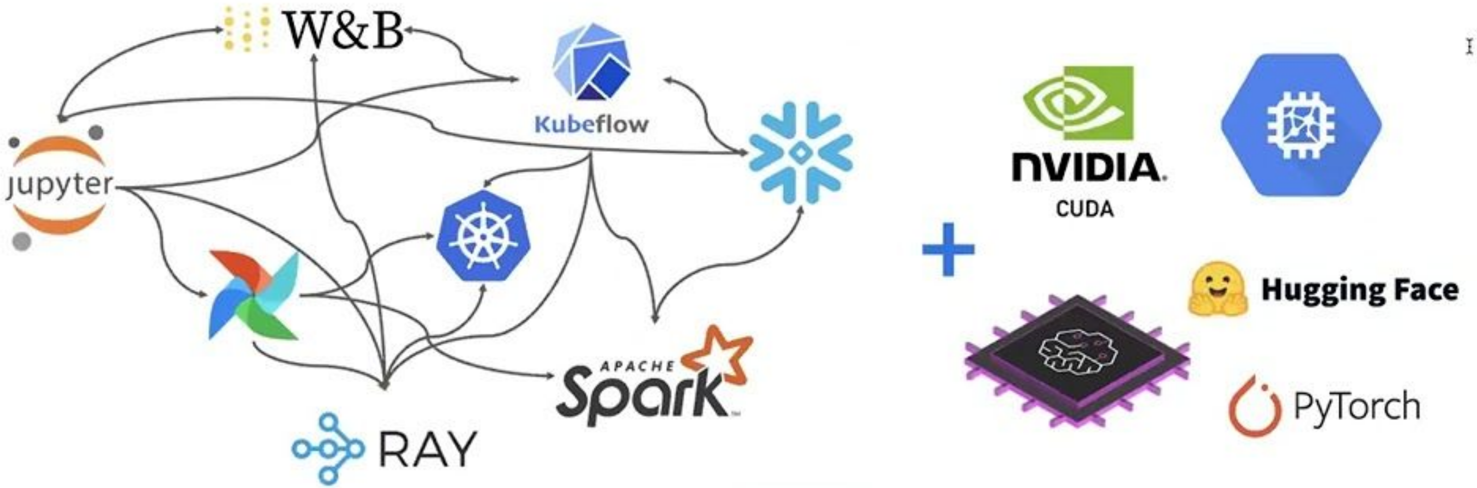
공격에서의 LLM 활용

- State-sponsored actors including from North Korea, Iran, and the People's Republic of China (PRC) continue to misuse Gemini **to enhance all stages of their operations, from reconnaissance and phishing lure creation to command and control (C2) development and data exfiltration.**
    - Google, GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools

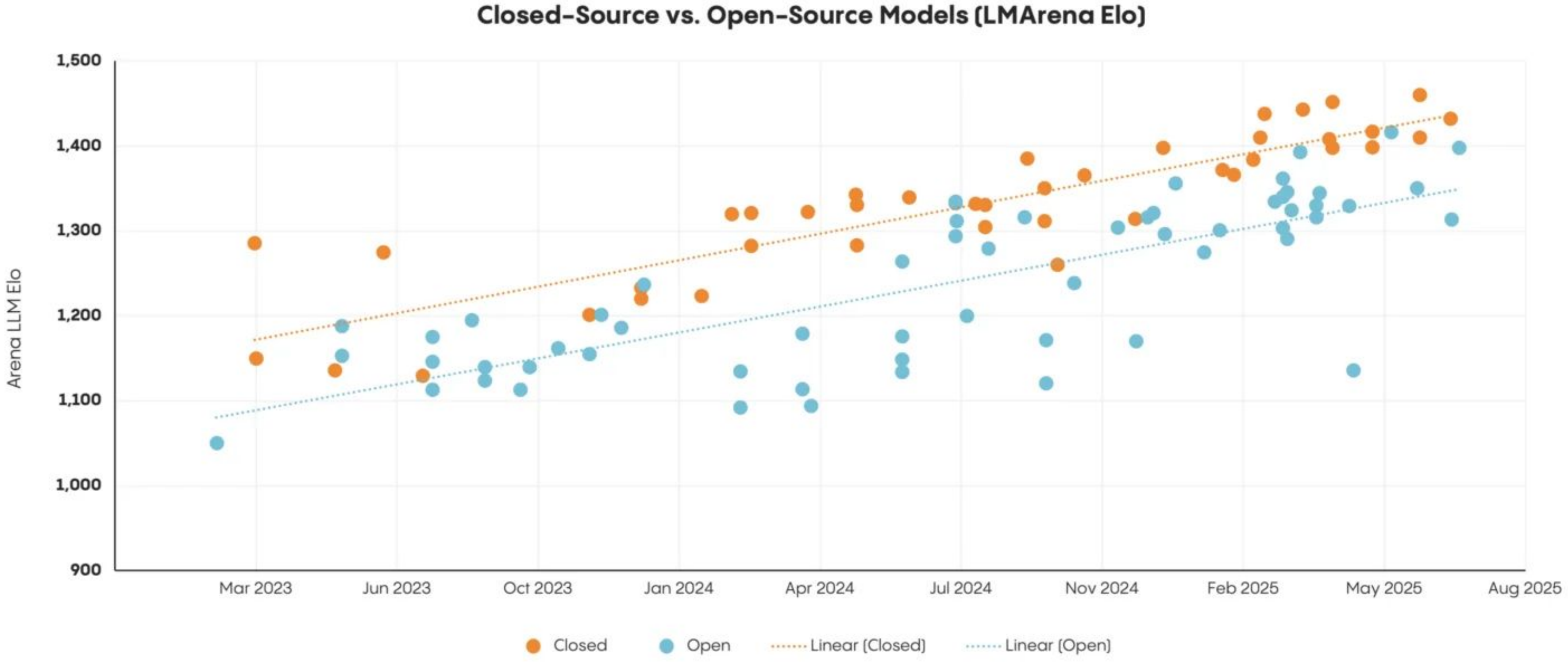
- The operation targeted large tech companies, financial institutions, chemical manufacturing companies, and government agencies. We believe this is the first documented case of **a large-scale cyberattack executed without substantial human intervention**.
    - Antropic, Disrupting the first reported AI-orchestrated cyber espionage campaign

… and LLMs only make things **that much more complicated.**

Because managing LLM environments requires yet more libraries, infrastructure, and specialized hardware.



Closed-Source vs. Open-Source Models



Source: LMArena

https://menlovc.com/perspective/2025-mid-year-llm-market-update/

AI 기업과의 협력

Palantir

OpenAI

ANTHROP\C

ANDURIL

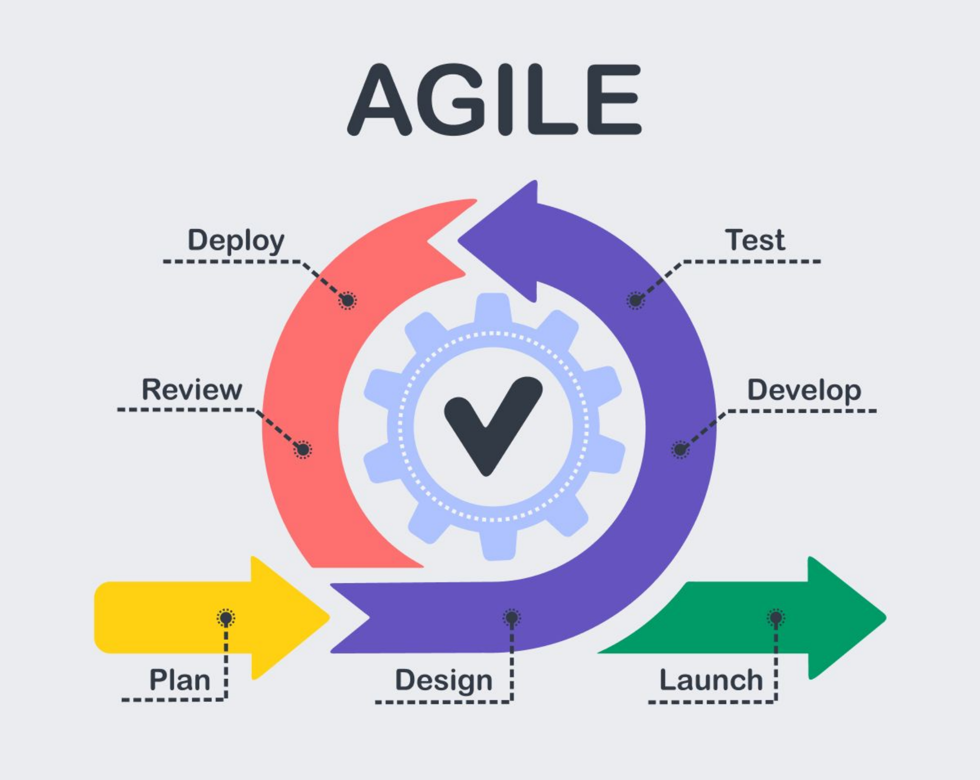
외부와의 연결

국가망보안체계(N2SF)란 무엇인가?
폐쇄적
분리된 서버
기존의
망분리 환경
낮은 유연성
자동화된 보안
클라우드
& AI 연계
N2SF
유연한 보안 환경
ENKI
엔키화이트햇

보안은 일순위가 아니다.

만약 보안이 정말 일순위라면, **우리는 아무 일도 하지 않는 것이 가장 안전한 길**일 것입니다. “Hello World 프로그램이 가장 보안적으로 안전하다”라는 우스운 말처럼, 아무것도 하지 않으면 공격받을 일도 없습니다. ... 그래서 저는 보안이 업무의 ‘우위’에 있는 것이 아니라, **업무와 항상 공동 1등**이 되어야 한다고 생각합니다. ...최근 빠르게 발전하고 있는 AI 혁명을 바라보면서, 우리나라의 **많은 기관과 기업들이 AI 도입을 주저하는 가장 큰 이유 중 하나가 ‘보안’이라는 사실이 아쉽습니다**. ... 우리는 **보안이라는 비용을 감수하면서도 업무 효율성을 극대화할 수 있는 AI 도입 방법에 대해 진지하게 고민**해야 합니다. ...

소프트웨어 공학에서 복잡성 (Complexity)를 다루는 방법: 반복 (Iteration)



- 변화하기 쉬운 구조 유지
  - 분할 (Decomposition)
  - 추상화 (Abstraction)
  - 모듈화 (Modularization)
- 반복(Iteration)적 개선
  - "작은 단계로 점진적으로 시스템을 완성해 나가는 개발 방식"
  - 초기 완벽을 추구하지 않고,
  - 점진적으로 개선, 수정, 확장
  - 리스크를 조기에 식별하고 관리
  - 복잡성을 단계별로 분해 및 대응

고급 해커들의 필요성 증대

고급 인력의 양성

KAIST Graduate School of Information Security

BEST OF THE BEST 차세대 보안리더 양성 프로그램

고급 인력들의 고용

https://www.shopify.com/stock-photos/photos/thank-you-for-your-service

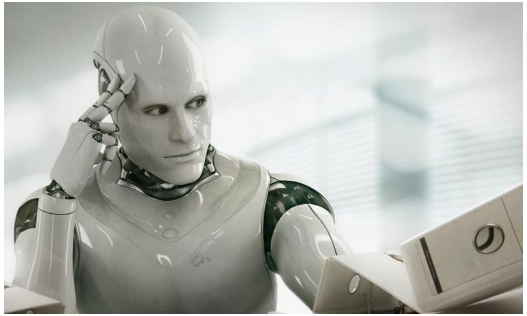
LLM에 대한 대응 (개인)

1000x Engineer

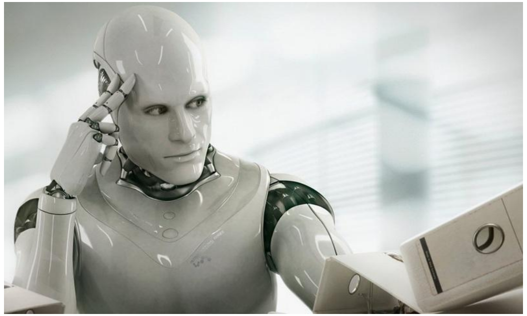
https://www.youtube.com/watch?v=hKfJZ3rWjik

인간 vs AI

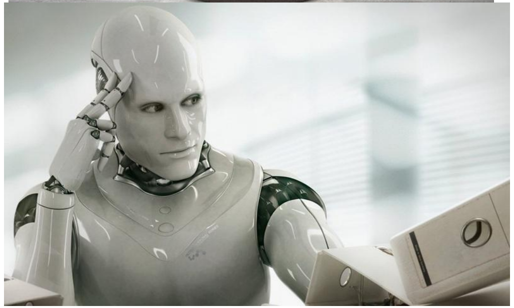
<=



인간 vs AI

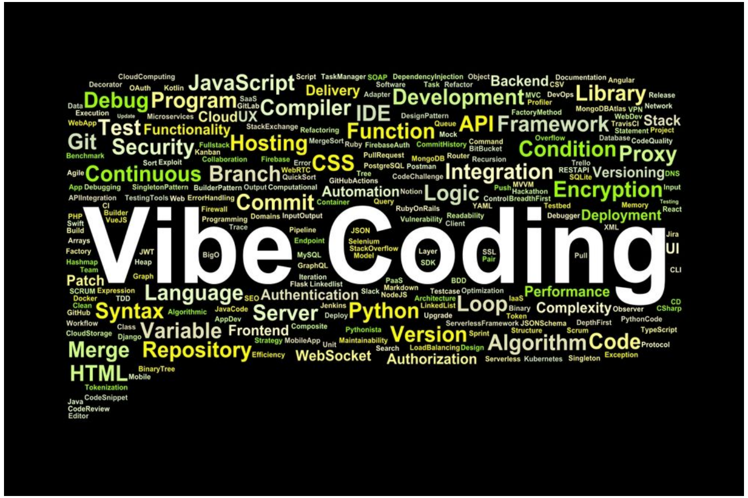


<



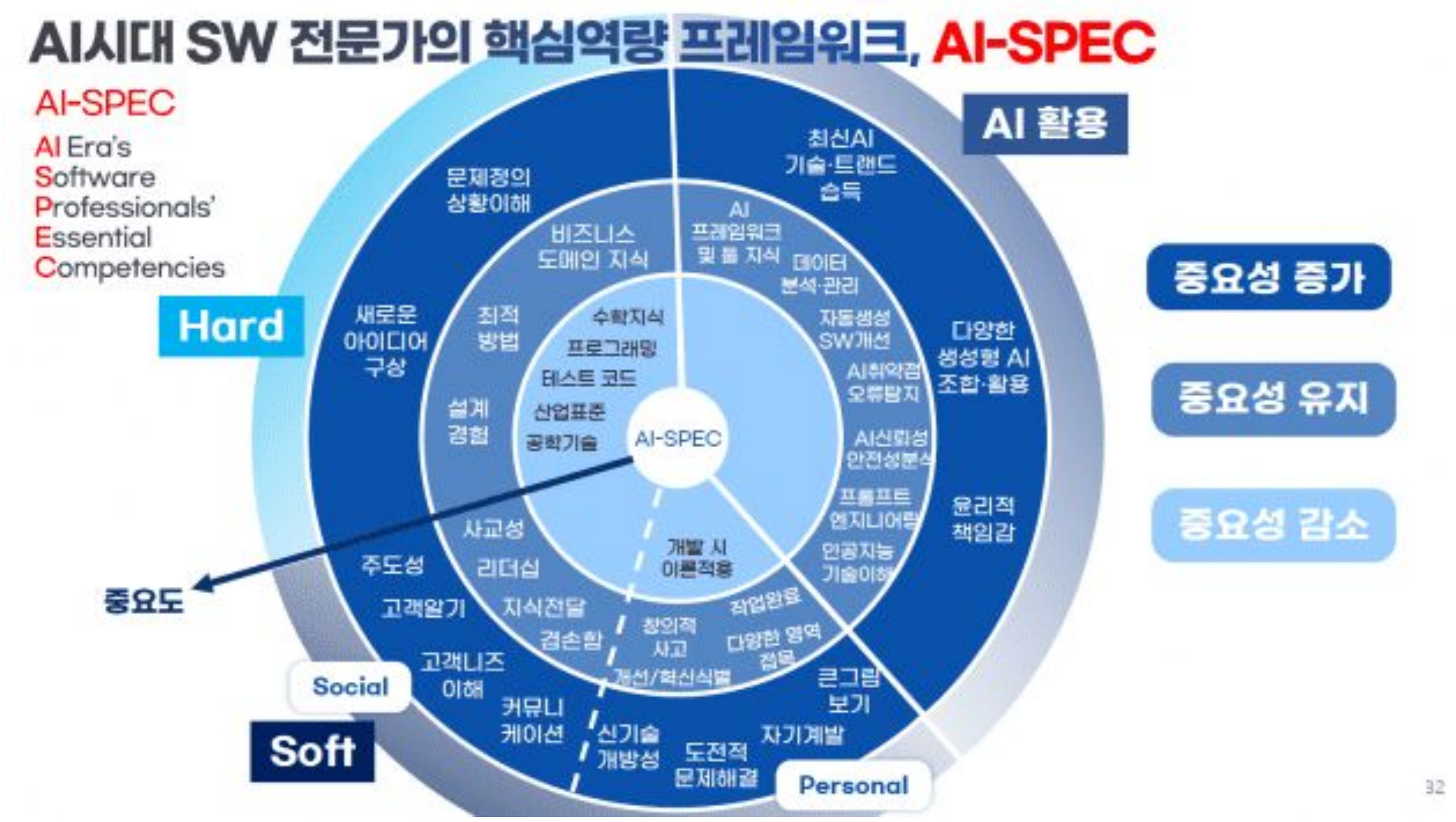

프롬프트
엔지니어링
챗GPT, 바드, 빙, 하이퍼클로바X까지
한 권으로 끝내기
반병현 지음
LLM 실무 필독서
생능북스


Cursor AI 사용법
NEXT.JS
CURSOR
아이디어에서 제품까지 15분...


Vibe Coding
JavaScript
Debug
Program
Compiler
IDE
Development
Library
Framework
API
Test
Functionality
Function
Hosting
Security
Git
Condition
Proxy
Continuous
Branch
CSS
Integration
Versioning
Logic
Encryption
Commit
Automation
Deployment
Patch
Language
Authentication
Performance
Syntax
Server
Python
Loop
Complexity
Variable
Frontend
Version
Algorithm
Code
Merge
Repository
WebSocket
Authorization
HTML
Backend
Delivery
CloudUX

AI 시대에도 화이트햇 해커가 필요한 이유 (by 엔키 화이트햇)

비교 항목	AI	화이트햇 해커
공격 방식	기존 알려진 공격 패턴만 학습하여 활용	새로운 공격 벡터를 창의적으로 탐색
데이터 활용	과거 데이터에 의존하여 패턴 기반 탐지	과거 데이터뿐만 아니라 실시간 분석과 직관 활용
적응력	학습된 환경에서만 효과적이며 새로운 환경에서는 어려움	새로운 환경에서도 창의적 접근을 통해 적응
창의적 사고	창의적인 사고 없이 데이터 패턴을 기반으로 분석	발상의 전환과 직관적 사고를 활용하여 보안 취약점 탐색
새로운 공격 기법 개발	새로운 공격 기법을 스스로 창출하지 못함	기존에 없던 새로운 공격 방법 개발 가능
예측 불가능한 공격 대응	예측 불가능한 공격에 대해 즉각적인 대응 어려움	실시간으로 예상치 못한 공격에도 즉각 대응 가능
자동화 가능성	단순 반복 작업 및 대량 분석에는 강점	자동화가 어려운 창의적 전략 수립 가능



https://zdnet.co.kr/view/?no=20250609103900

결 론

- LLM의 발전

- LLM의 영향
  - LLM의 긍정적인 면: 업무 자동화
  - LLM의 부정적인 면: 프롬프트 인젝션

- LLM에 대한 대응 (조직)
- LLM에 대한 대응 (개인)

감사합니다