

AI's Impact on Application Security From an Attacker's Perspective

and What to Do About it

Hey, I'm Chris



- Studied Psychology and Programming at UCLA
- Been in Sales for over 10 years, with the past 5 being in Cybersecurity
- Working as an Account Executive at Kodem Security, an AI Runtime AppSec company
- This is my 3rd AppSec company, having taken one to acquisition by GitLab
- I have the world's largest collection of vintage Star Wars comics

Hackers > Researchers > AppSec Practitioners



Aviv Mussinger
CEO & Co-Founder



Idan Bartura
Head of Engineering & Co-Founder



Pavel Furman
CTO & Co-Founder

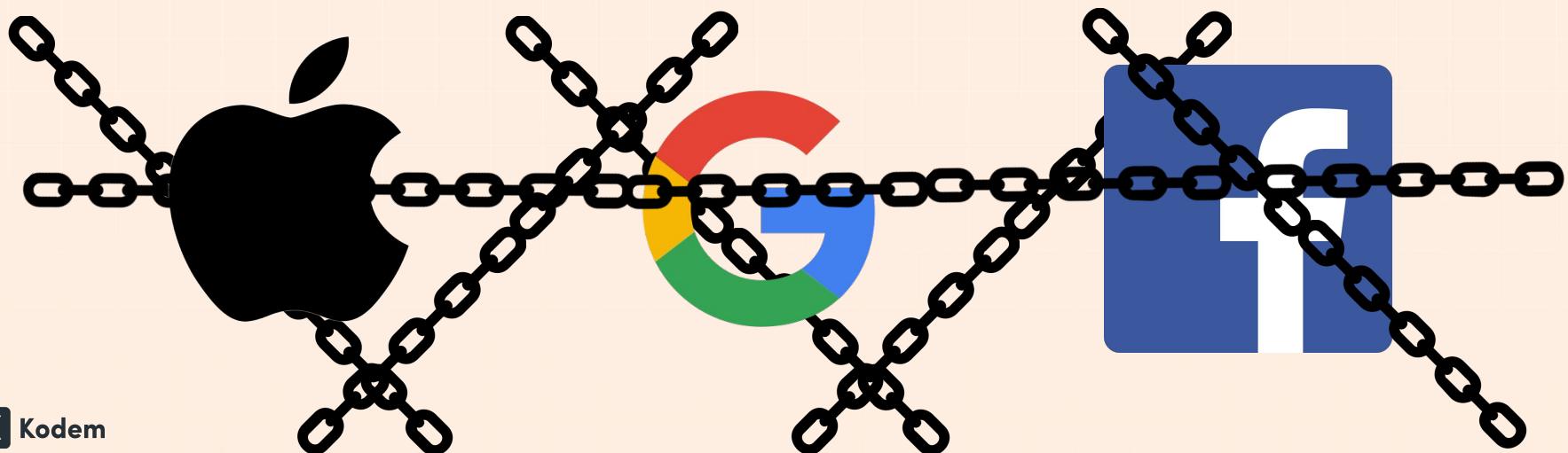
NSO Group vs. Apple, Google, and Facebook

How could a small team of researchers continually exploit and bypass the operating systems of the largest companies in the world?



NSO Group vs. Apple, Google, and Facebook

How could a small team of researchers continually exploit and bypass the operating systems of the largest companies in the world?

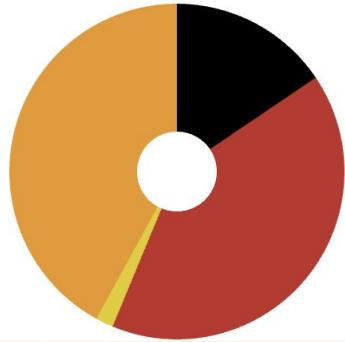


The Key to Maintaining Exploits Over Time: Low-priority Attack Chains

- By chaining together **Low and Medium severity vulnerabilities**, NSO Group was able to have functional exploits for longer periods of time and with more available permutations
- Because these issues were often less-prioritized or even ignored, it posed a powerful opportunity to bypass the cryptography of smartphones
- This approach laid the groundwork for future types of exploits that pose a significantly greater risk today

Vulnerability Knowledge Gaps Widen

CVSS V3 Score Distribution



Severity	Number of Vulns	
CRITICAL	26073	15.54%
HIGH	68332	40.73%
MEDIUM	70504	42.72%
LOW	2861	1.71%

Source: <https://nvd.nist.gov/ageneral/nvd-dashboard>

CVE Status Count

Total	295666
Received	235
Awaiting Analysis	24959
Undergoing Analysis	13404
Modified	134932
Deferred	94591
Rejected	15199

NVD Contains

CVE Vulnerabilities	295666
Checklists	832
US-CERT Alerts	249
US-CERT Vuln Notes	4486
OVAL Queries	0
CPE Names	1414521

~56% of scored issues

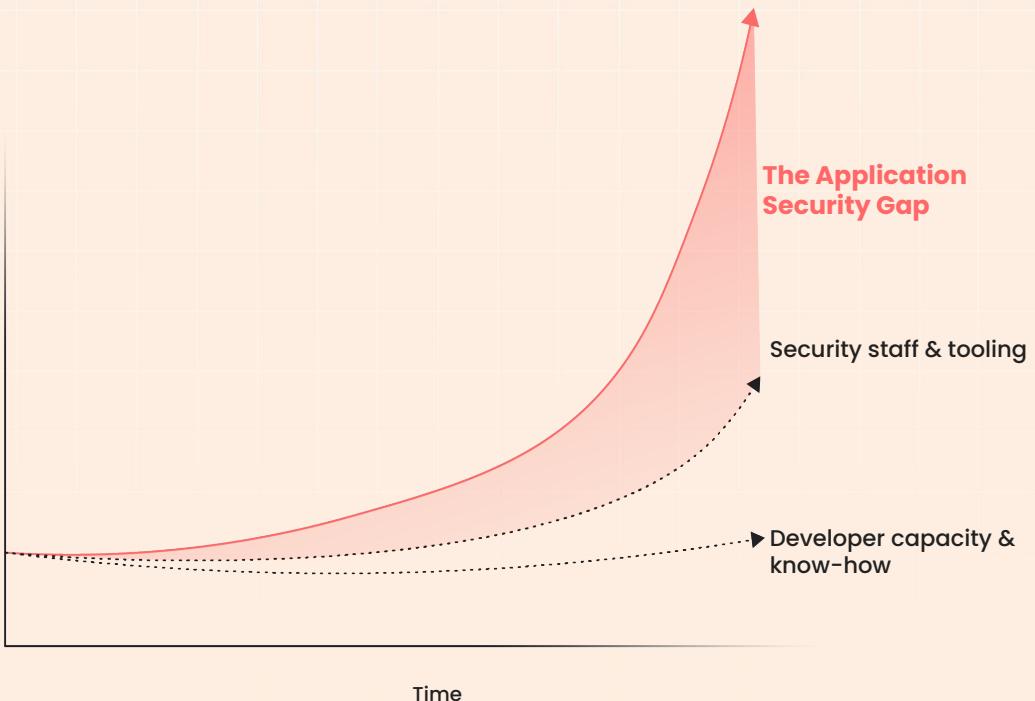
~44% of scored issues

167,770 scored / 295,666 total CVEs = 56.74% of known CVEs in the NVD are scored; this percentage is decreasing over time

Growing backlog of CVEs awaiting analysis; NIST proposes using ML to accelerate review process

Source: 3/19/25 General Update: <https://www.nist.gov/itl/nvd>

Remediation Isn't Keeping Pace with Exploits



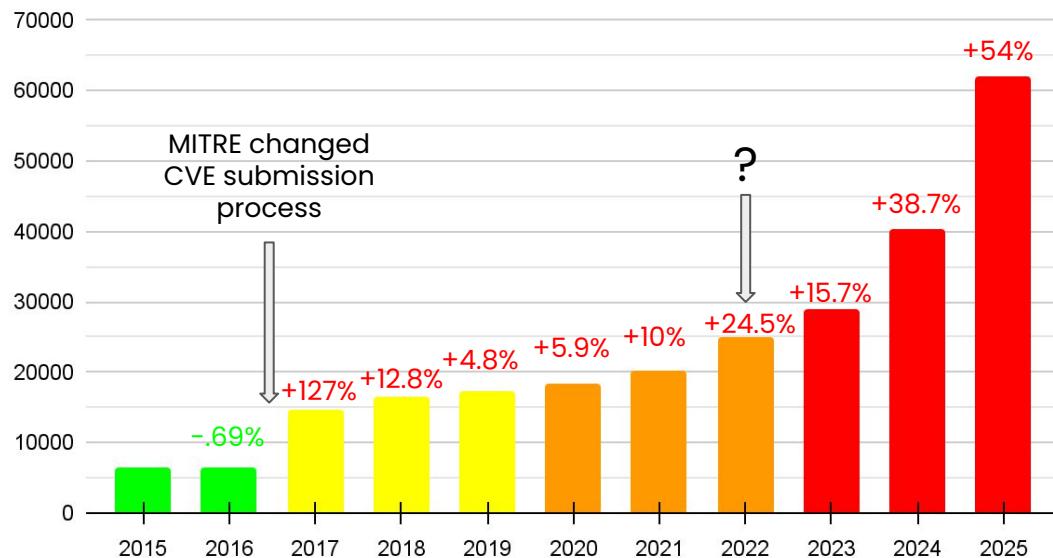
- ~80% of time is spent focused on Highs and Criticals vs. ~20% focused on Lows and Mediums
- **MTTR of Criticals:** 32 days from *first detection*; only 54% fully remediated within the year
- **Avg. Time to Exploit by Threat Actors:** 5 days from *known issue*

Source: [2025 Verizon Data Breach Investigations Report](#)

Why the Vulnerability Count is Widening

and Why Applications are Being Exploited More Often

CVEs Published Per Year (with 2025 Estimate)

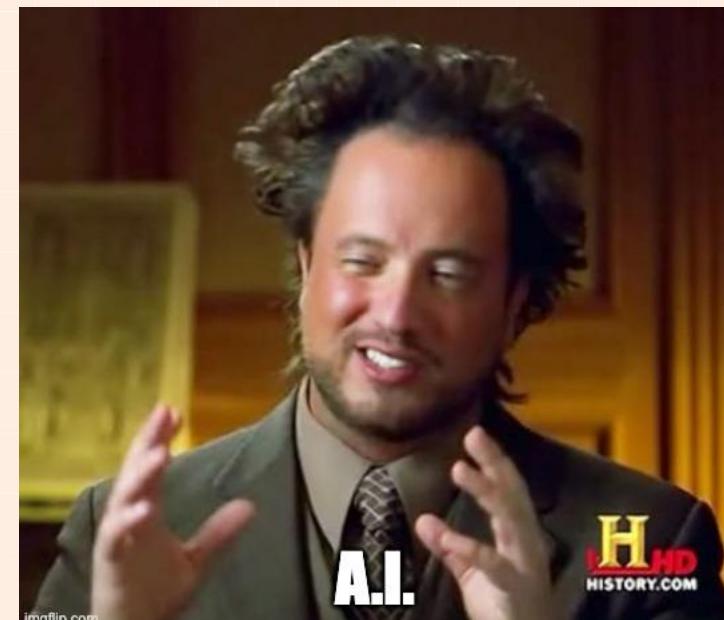
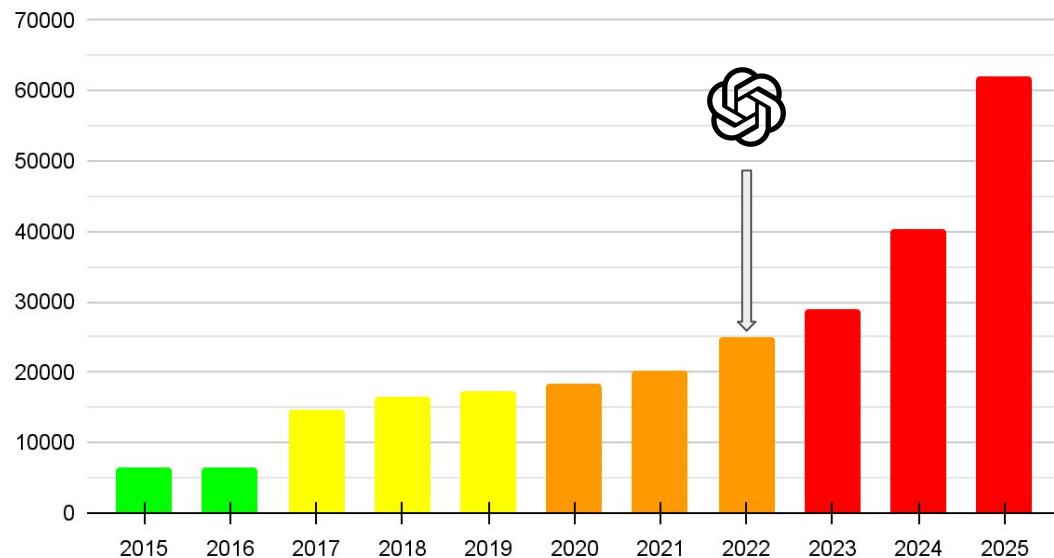


- Better detection and verification methods
- Increasing complexity of codebases
- Expanding digital infrastructure
- ??? HINT: November 2022

Why the Vulnerability Gap is Widening

and Why Applications are Being Exploited More Often

CVEs Published Per Year (with 2025 Estimate)



Web Apps are Being Exploited at an Increasing Rate

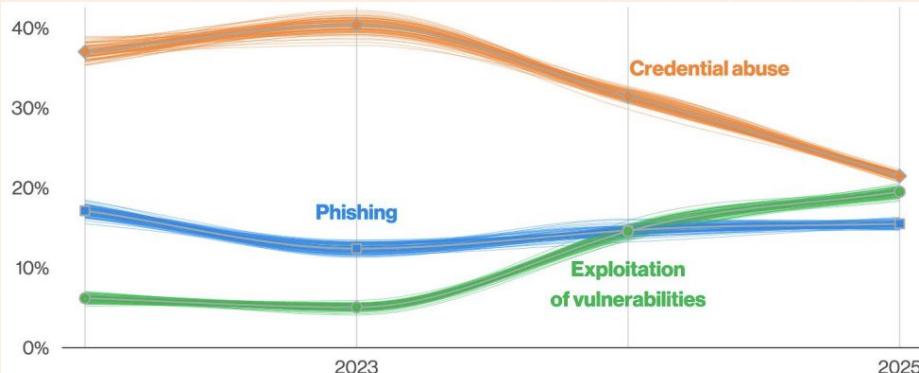


Figure 16. Known initial access vectors over time in non-Error, non-Misuse breaches (n in 2025 dataset=9,891)

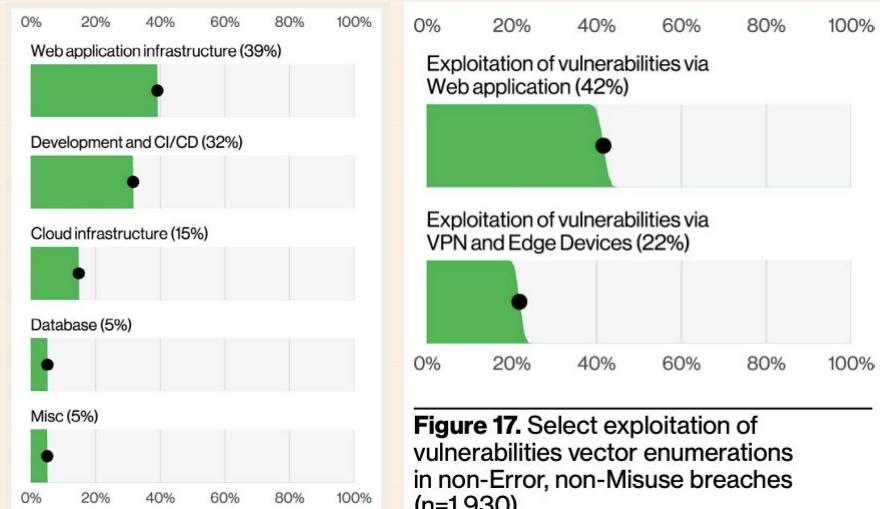


Figure 12. Top categories of exposed secrets in public git repos (n=441,780)

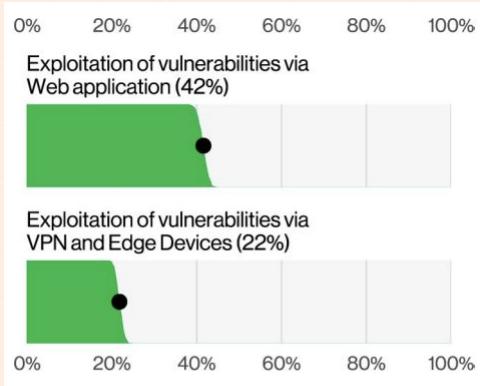


Figure 17. Select exploitation of vulnerabilities vector enumerations in non-Error, non-Misuse breaches (n=1,930)

To Summarize So Far:

- The 1st team to bypass the iPhone's security and cryptography realized they could keep attacks paths open by **chaining and substituting** Low and Medium severity vulnerabilities because they tended to be fixed much later, if at all
- There's a growing knowledge gap between known and unknown vulnerabilities, with new issues appearing more frequently than ever
- "AI" (ML Models or LLMs) has been a catalyst for developer productivity but also often leads to insecure code and web app risks
- Threat actors are focusing more on insecure web applications due to the ease of finding and using new exploits with slower response times from defenders to fix them

How to View These Issues From the Lens of an Attacker

With the MITRE ATT&CK Framework

- Provides examples of how attackers breach digital infrastructure
- Can be used to tell a story around the impact these attacks will have on applications and your organization



Hacker Math for Web Apps

CVSS 2.4 + 3.2 + 1.8 =
9.8?!

Don't use CVSS score as your sole determining factor

Chaining Lows and Mediums are how you end up getting exploited through a truly critical exploit



Validating Attack Chains Using an AI Red Team

And the MITRE ATT&CK Framework

Credential Exfiltration

Current Issue



Initial Access



Current Issue



Execution



Current Issue



Credential Access



Discovery



Collection



...

Details: Attacker sends a specially crafted request to exploit buffer overfmedium in WASM module.

Prerequisites: The target system must be running a vulnerable version of Go with
GOARCH=wasm GOOS=js.

Buffer Overflow
CVE-2021-38297 on go stdlib



Has Fix

Was this helpful?



How Attacker's Can Use AI

1. LLMs and/or AI models being poisoned by prompt engineering
2. Agentic AI being exploited through similar prompt engineering/injection
3. Developers using AI to unintentionally write insecure code and/or non-obvious attack paths
4. Creating and injecting malicious payloads
5. Finding attack chains, especially if they have access to organizations' AI tools



Ken Huang, CISSP · 2nd

AI Book Author | Speaker | DistributedApps.AI | OWASP T...

6h • Edited •

✓ Following

...

An Agentic access control vulnerability in the GitHub MCP integration allows attackers to hijack a user's coding agent by creating a malicious GitHub Issue in a public repository; when the agent, such as one used with Claude Desktop, fetches and processes the issue, prompt injection tricks it into leaking sensitive data from private repositories—such as proprietary code—by autonomously creating a pull request in the public repo, making the confidential data accessible to the attacker, and this flaw is not a bug in the server code but a systemic issue requiring agent-level security controls and permission boundaries to prevent cross-repository data leaks. The call for a new Agentic IAM approach that we discussed in our recently published paper. See the link in the comments

[Link to post](#)

Where to Start in Securing Your Web Apps

1. Apply guardrails to your ML Models, LLMs, Agentic AI and all other forms of “AI” to harden your security posture and prevent leaking access controls or sensitive data; especially if it’s being used by your developers
2. Establishing an inventory (SBOM) of your code and packages, and understanding what’s currently in use and where your most vulnerable assets are
3. Ask what kind of mitigation factors do you already have in place and where do you have gaps? Use this in conjunction with your most critical assets to prioritize securing those areas first
4. Approach security from the perspective of an attacker; take into consideration attack chains rather than individual issues and how you can most quickly and easily break those chains and how many times certain vulns appear in those chains

Tools and Projects Worth Checking Out



<https://cyclonedx.org/>



[https://owasp.org/
www-project-vxdf/](https://owasp.org/www-project-vxdf/)



Let's Connect!



Thank
You

Learn About Kodem!

