

AIでMass Exploitationの 採算ラインが下がった

2026年秋の「漏洩まつり」を、攻撃能力より「攻撃コストの変化」から読む

GLM-5.3

N-day → N-hour

CVSS Medium

国内インシデント

20 min Lightning Talk

2026.10

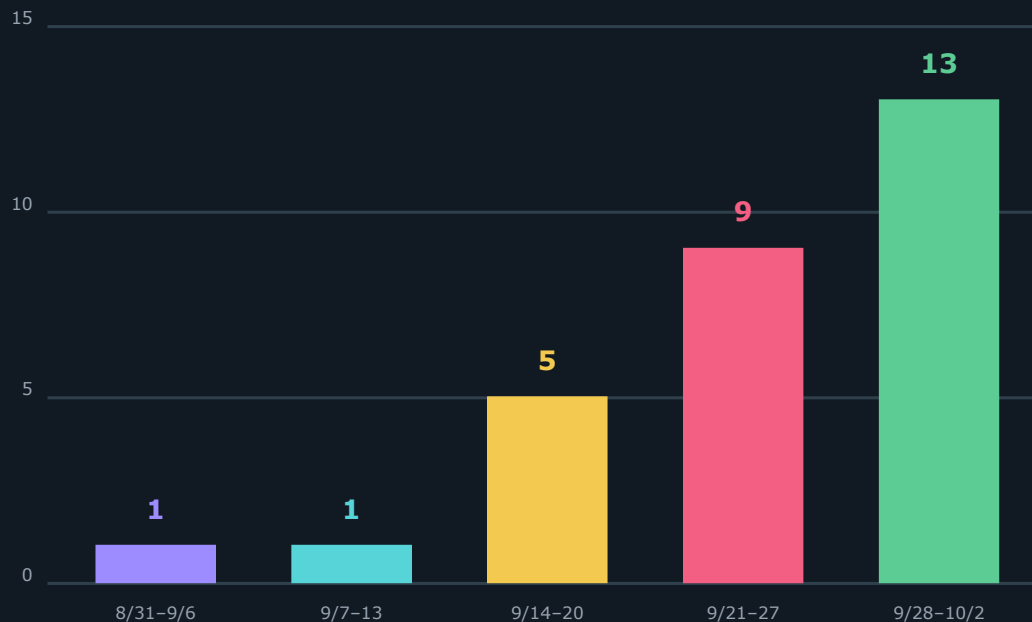
仮説：AIは「新しい攻撃」より先に、
これまで採算が合わず見送られていた攻
撃を成立させる。

「増えた」を語る前に、公表日と侵害日を分ける

左はこのLTで拾った国内29件の公表日。右は米国ITRCのdata compromise統計。

国内：主要公表事案の週別件数

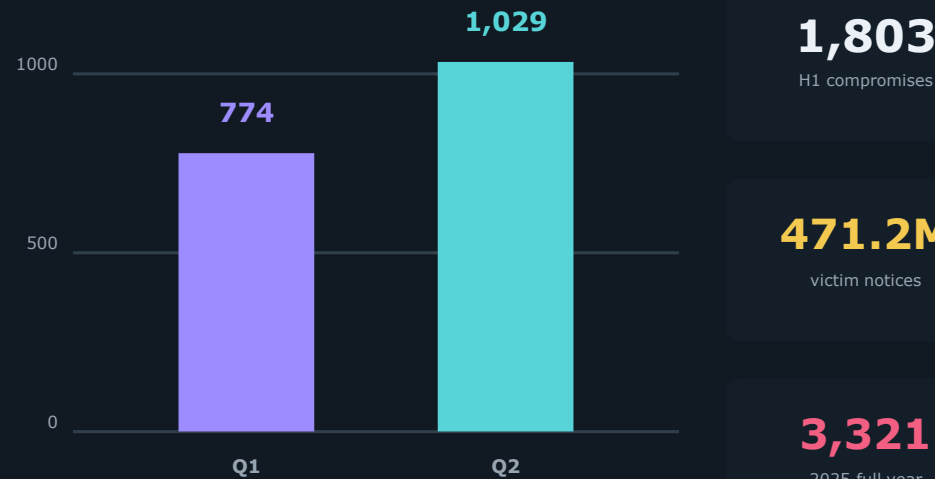
このLTで拾った29件。網羅統計ではない



22 / 29 件が最後の2週に公表

米国：ITRC Data Compromises

2026 H1 — reported compromises



→ 9月に侵入が急増したとは言えない。見えているのは公表の集中で、侵害時期は別に追う。

AIが漏洩を起こしたわけではない。変わったのは攻撃コストだ

Anthropicが実攻撃で見たのは、新手法より既存攻撃の単価低下だった。

$$\text{Attacker ROI} = \text{Payoff} - (\text{Recon} + \text{Exploit Dev} + \text{Adaptation} + \text{Data})$$

Recon

探索を並列化
→ 1件あたりコスト↓

Exploit Dev

PoC作成を短縮
→ 1件あたりコスト↓

Adaptation

環境差分へ自動対応
→ 1件あたりコスト↓

Data Processing

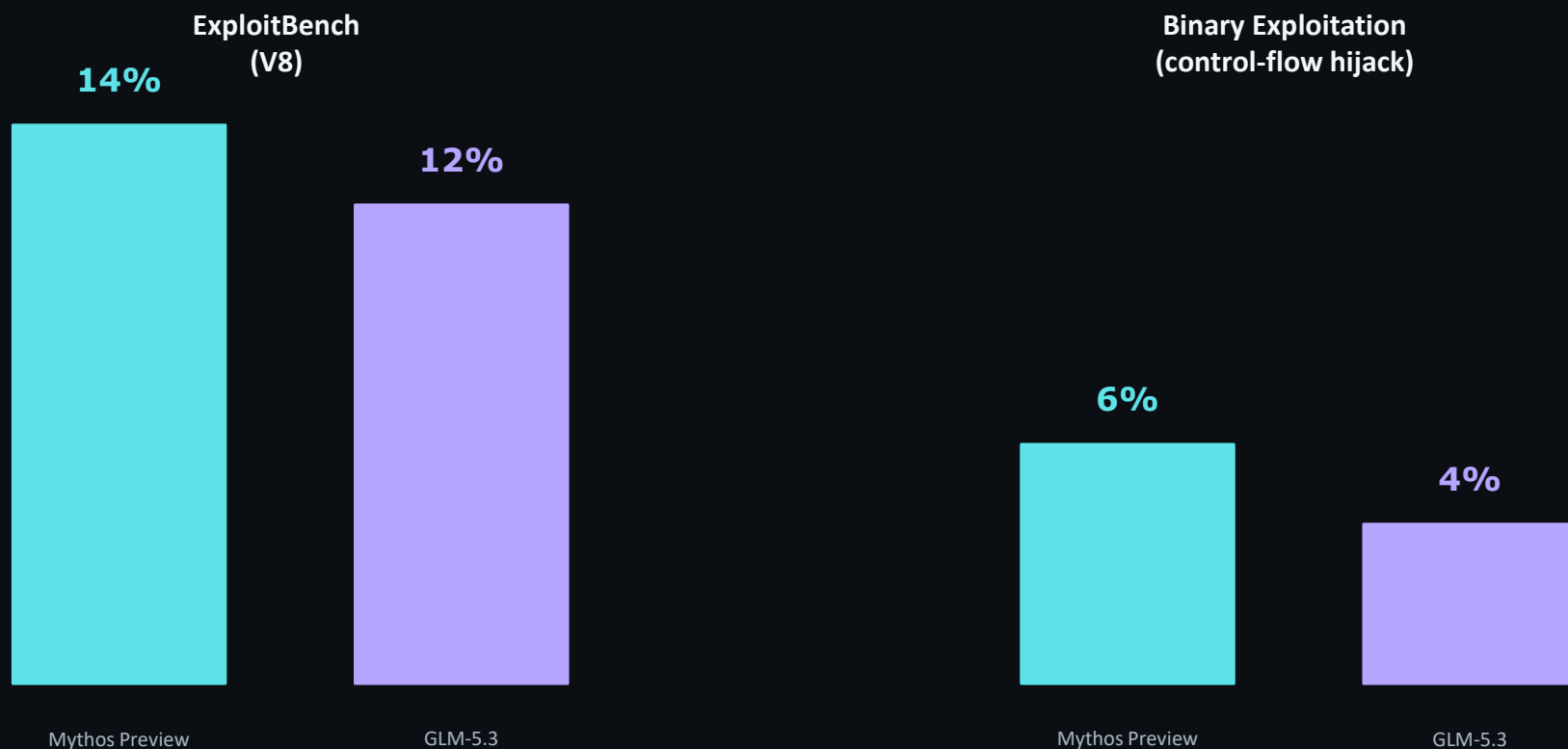
窃取データを自動選別
→ 1件あたりコスト↓



これまで採算が合わなかった“marginal target”まで攻撃候補に入る

GLM-5.3で、Mythos Preview級のexploit構築能力がopen-weightになった

能力の絶対値より、第三者の利用制限なしで持ち出せることが効く。



KNOWN-VULN CHAIN

20分

人間の注力

+ モデル8時間

\$20.40

ARM64 exploit chain
PAC bypass

以前のモデルでは、この水準のend-to-end exploitはほぼ0%だった

20.40ドルは既知欠陥連鎖のAPI料金

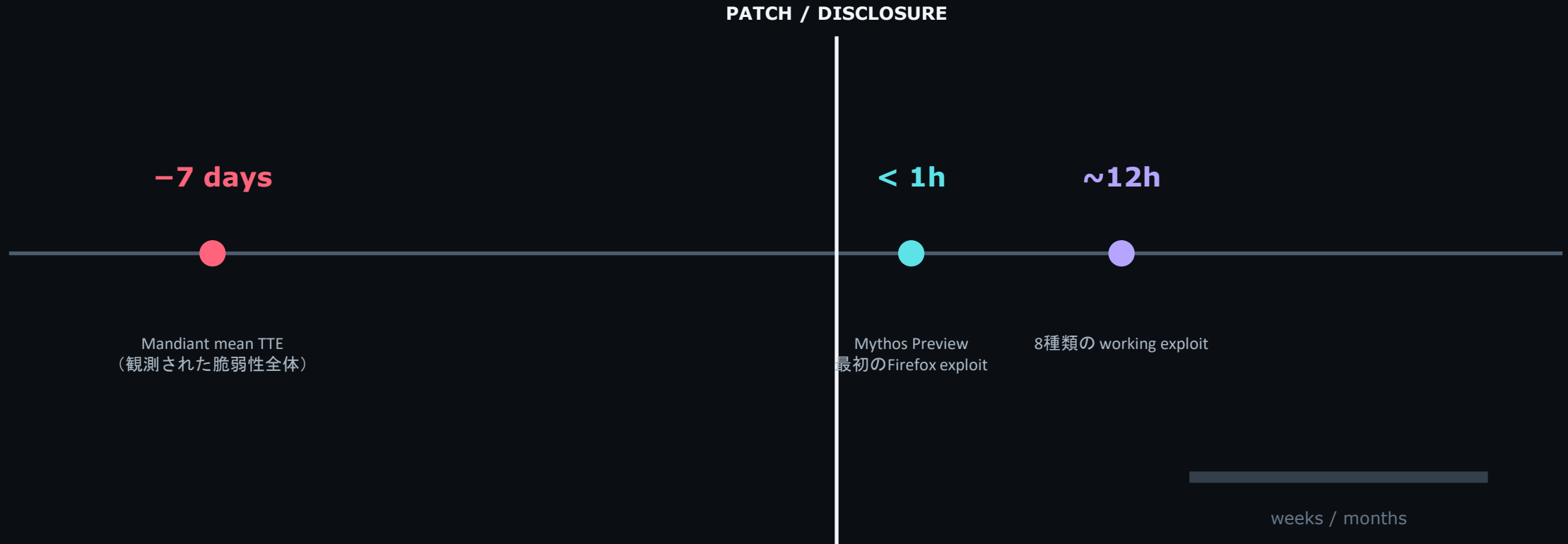
GLM-5.3-Flashを隔離環境のARM64で評価。人は約20分、モデルは8時間作業。

INPUT	PROCESS	COST
公開情報 既知の2欠陥 • ChromeのCVE-2026-11645 • ほかの既知欠陥 • 研究者が両者の詳細を入力 未知欠陥の発見実験ではない 対象は隔離したARM64環境	Flashモデル 人の作業 約20分 モデルの作業 8時間 • 2欠陥を連鎖 • PACの保護を回避 • 動作する攻撃を構築 研究者の指示は限定的 結果は一つの実験	API単価換算 \$20.40 • ZhipuのAPI価格で試算 • 人件費などは含まない • 全攻撃の費用ではない 対象実験のモデル利用料 実支払額ではなく換算値

Source [Anthropic \(2026-09-29\)](#)

N-dayの猶予は週・月単位から時間単位へ縮んでいる

Patch diffが出れば、exploit作成の裾野が一気に広がる。



「パッチが出てから考える」運用では遅い。公開直後からexploit前提で動く必要がある。

攻撃者は、もうAIを実運用に入れている

Reconから窃取データの選別までを、並列・高速に回している。



ATTRIBUTED ACTOR

UNC6240 / ShinyHunters

GTIG観測：Claude Codeで難読化コードやperimeter bypassを補助。MCP連携では窃取ディレクトリを解析し、恐喝準備にも使った。

GTIGが観測した認証情報窃取は6時間未満で構築・実行

侵害済みクラウド上で、AIを組み込んだ自動走査と収集を実行した一事例。

ENTRY

攻撃の足場

侵害済みクラウド

- 先に組織の環境を侵害
- そこから攻撃を実行
- 正規IPが送信元になった

攻撃の開始点も実被害
AIが入口を突破したとは未確認

BUILD

AIの使い方

6時間未満

- AI coding chatbotを利用
- 指示ファイルを実行手順に
- 計画・構築・実行まで

6時間未満は一事例の観測
平均所要時間ではない

RESULT

観測された結果

数千件

- 第三者の認証情報
- 自動走査と収集
- 障害対応・IP切替も自動化

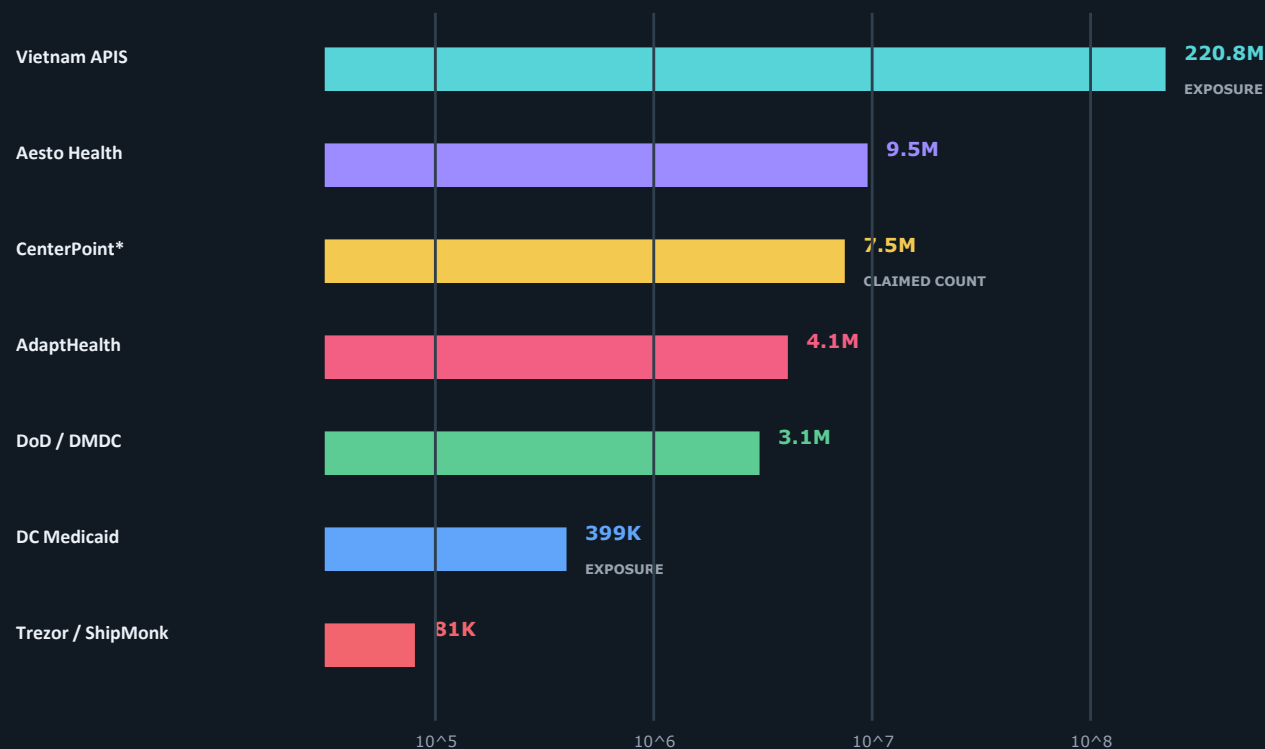
GTIGは実攻撃として報告
攻撃全体がAI単独とは言わない

海外事例は、件数より「入口」と「データ集約点」で見る

9月公表の代表例。人数・records・actor claimが混ざるため、単純合算せず攻撃経路で整理する。

Selected exposure scale

people / records — log10 scale



同じ「漏洩」でも、入口はかなり違う

VULN / EXPOSED SYSTEM

FBIJobs / PeopleSoft
DMDC file-sharing vuln

IDENTITY / SOCIAL

AdaptHealth contractor
Revolut fake gov request
DriveWealth social engineering

THIRD PARTY / DATA GRAVITY

Aesto healthcare archive
Trezor → ShipMonk

MISCONFIGURATION

Vietnam APIS / DC Medicaid
→ malicious access未確認でも影響大

帰属は別レイヤで持つ。PeopleSoftはMandiantがUNC6240 / ShinyHuntersに帰属。Berlin / Rhysidaなどはactor claim。全件をAI起因とは扱わない。

GSSでは、既知・CVSS Medium・対応中でもパッチ前に悪用された

Internet-facing edgeでは、Severity順の修正キューだけでは間に合わない。



漏えい可能性：約24.6万件

Source デジタル庁：VPN機器の既知脆弱性、当初CVSS Medium。対応中だったが修正適用前に悪用

GSSはCVSS Mediumの既知VPN欠陥をパッチ前に悪用された

2026年6月25日に異常を検知。通常より早い対応を進めたが間に合わなかった。

TIMELINE

検知と公表

6/25 検知

- 7/9 VPN侵入を確認
- 7/15 個情委へ報告
- 9/11 事案を公表

侵入日自体は非公表
公表日は攻撃日ではない

ENTRY

悪用された入口

CVSS Medium

- 攻撃前に公表済み
- 一般的な基準より早く対応
- 修正適用前に悪用

具体的な欠陥は非公表
AI利用も確認されていない

IMPACT

漏えいのおそれ

約24.6万件

- 職員等の個人情報
- 氏名・業務用連絡先など
- 対象を広めに数えた件数

漏えい確定件数ではない
デジタル庁は調査を継続

漏洩件数より、「次の侵害に使える情報か」を見る

PIIに加え、運用資格情報・セッション・OCR・予約情報が次の攻撃につながる。

OPERATIONS

Shopserve

8.85M records

- 管理画面 ID/PW
- メール ID/PW
- FTP ID/PW

→ downstream compromise
BEC / EC改ざん / credential reuse

DERIVED DATA

Gyazo

**23.62M users
+ 490M metadata**

- session ID
- OCR text / EXIF
- image ID / source URL

→ 画像本体がなくても
内部情報を再構成できる

CONTEXT

TEMAIRAZU

予約データ

- 氏名 / 日程
- 宿泊施設
- 予約内容

→ 実在の予約に見える
決済再入力・標的型詐欺

Gyazoで削除済み画像のメタデータ約1.74億件が流出

2026年9月25日時点の公表。画像ファイル自体の流出件数ではない。

DELETED	OTHER DATA	USERS
削除済み画像 約1.74億件 - 画像に関するメタデータ - 主に2023年2月以前に削除 - 画像ファイルは含まれない 過去に消した画像の情報 第二報で新たに確認	画像メタデータ 約4.9億件 別区分 240万件 - 主に2019年1月以前に投稿 - 総画像数の約14.4% 240万件は条件限定の取得 公表された区分ごとに読む 画像本体の流出数ではない	ユーザー情報 約2,362万件 - ユーザーに関するデータ - 約76%は匿名ユーザー - 全件がメール流出ではない 人の数として扱わない 登録有無で影響が異なる

Source [Heliofeel 第二報 \(2026-09-25\)](#)

AIで、これまで割に合わなかった「長尾」が攻撃対象に入る

未知の脆弱性を魔法のように解くより、既知だが放置された対象を大量に処理する効果大きい。



Source Anthropic: “previously marginal targets viable” / unique & obscure configurations become exploitable

防御側もAIを使う。ただし、CVE要約だけでは速度差は埋まらない

脆弱性管理の一連の流れをmachine-speedに近づける。



CVSS順の修正キューをやめ、P0は数時間で切る

Internet-facing × 認証回避/RCE × 高価値data plane × weaponizableを最優先にする。

P0 / HOURS

Internet-facing edge / auth bypass / RCE / BI ・ 管理面 / actively exploited

隔離 ・ patch/mitigate ・ credential rotate ・ retro-hunt

P1 / < 1 DAY

Reachable internal / privilege escalation / high-value SaaS integration

修正 ・ 検知追加 ・ ログ保全 ・ blast radius確認

P2 / NORMAL

低到達性 / exploit chainが長い / compensating controlが強い

通常SLA。公開diff / PoC / exploitation観測が出たら昇格

優先度 = Severity × Exposure × Reachability × Business impact × Exploit intelligence × Machine weaponizability

明日から変えるなら、この3つでいい

全部をAI化するより、攻撃コストが下がった箇所から防御のSLAを変える。

01

Internet-facingは別SLAにする

VPN / Gateway / CMS / BI / admin plane。
CVSS Mediumでもexploit intelligenceとexposureを優先する。

02

patchで閉じない。必ずretro-huntする

公開・露出開始からpatch適用までを遡る。
ログがなければ「侵害なし」ではなく「判定不能」とする。

03

AIは防御の処理量上げるために使う

asset照合、patch diff解析、sandbox再現、
検知クエリ生成、証拠整理をagentic workflowに載せる。

「わかっているけど先延ばし」にしていた修正ほど、攻撃単価が下がった世界では危険になる。

AIは、新しい攻撃より先に「攻撃しなかった理由」を消す

防御側が変えるべきなのは、脆弱性の知識量より response economicsだ。

BEFORE

「攻撃するには手間がかかる。
だから、たぶん来ない」



NOW

その「手間」を
AIが引き受ける

AIが漏洩を起こしたと断定する必要はない。
「Mass Exploitationの採算ラインが下がった」と見るだけで、2026年秋の現象はかなり説明できる。

守る側も「見つける → 判断する → 直す → 遡る」を、攻撃側に負けない速度へ。

主要ソース

LT本編は一次情報を中心に構成。国内事案の帰属やAI利用は、確認できた範囲だけを書いている。

Anthropic — GLM-5.3 and the spread of advanced cyber capabilities
www.anthropic.com/research/glm-5-3-and-the-spread-of-advanced-cyber-capabilities

Anthropic — Measuring LLMs’ impact on N-day exploits
www.anthropic.com/research/n-days

Anthropic — Countering misuse of AI: September 2026
www.anthropic.com/threat-intelligence-report-september-2026

Google / Mandiant — M-Trends 2026
cloud.google.com/blog/topics/threat-intelligence/m-trends-2026

Google / Mandiant — AI-Assisted Vulnerability Management
cloud.google.com/blog/topics/threat-intelligence/ai-assisted-vulnerability-management

デジタル庁 — GSS不正アクセス Q&A
www.digital.go.jp/press/5fc99139-a4e2-4b7b-8b0c-d475e926143f

Helpfeel — Gyazo 不正アクセス
corp.helpfeel.com/news/news-20260916-1

Eストアー — Shopserve 不正アクセス 第2報
estore.jp/press/20260802

タイムズカー — 不正アクセス 第3報
share.timescar.jp/news/2026/0929/1816.html

スマレジEC — 不正アクセス
ec.smaregi.jp/repeat/news/2026925-2

VOISING — BIツール不正アクセス 第4報
voising-official.com/news/1015

手間いらず — TEMAIRAZU 不正アクセス
www.temairazu.co.jp

Q&A

海外事例はAppendixに確認レベル付きで掲載

国内事例で、AIが効きやすい工程はどこか

AI使用を確認したという意味ではない。工程ごとに、AIで圧縮しやすい部分を示す。

事案	公表された入口	AIで圧縮されうる工程	圧縮余地
GSS	VPN既知脆弱性	asset discovery / exploit adaptation	High
Gyazo	upload server vuln → RCE	vuln research / exploit dev / data processing	High
スマレジEC	login機能の脆弱性	mass scan / exploit adaptation	High
VOISING	BI tool 既知脆弱性	N-day weaponization / mass scan	High
Shopserve	不正program実行 / exfil	data processing / follow-on credential abuse	Medium
TEMAIRAZU	侵入経路は調査中	stolen-context processing / phishing generation	Medium
タイムズカー	侵入経路は非公表	unknown — attribution不可	Low

※「圧縮余地」は、その工程をAIで効率化しやすいかの評価。個別事案でAI使用を確認したという意味ではない。

2026年秋、国内では「公表」が短期間に集中した

同一攻撃者・同一キャンペーンを示す証拠はない。国内主要例は公開情報上、ほぼ帰属未確認。



見えているのは、新技術で全部やられた世界ではない。既知脆弱性、公開Web、SaaS/BI、認証情報、設定不備が短期間に大量に刈られる世界だ。

Source デジタル庁 / Helpfeel / エストアール / タイムズカー / スマレジEC / VOISING 各社公表

Grokの一覧は、そのまま使わない。公表日・侵入日・確認範囲を分ける

「秋に公表された」と「秋に侵入された」は別。actor claimも別の確度で持つ。

CASE	LISTでの見え方	時系列の補正	確認できること	留保
Aesto Health	9/1頃に9.54Mが可視化	侵害は2025/12	件数はHHS等で確認	公表集中 ≠ 秋の新規侵入
Vietnam APIS	9月に報道	発見は2026/6	220.8M exposure	悪用・侵入は未確認
CenterPoint	9/14 SEC開示	9月にpostを認知	一部顧客情報取得を確認	7.5Mはactor側の主張
FBIJobs	9/22以降	PeopleSoft campaignと接続	侵害・機微dataは確認方向	2–3TBはactor側の主張
Microsoft / ExfilSquad	「9月」ではない	初期claimは7/25–26	一部sampleの真正性報道	130GB / 8Mは公式確認なし

海外事例は、確認済み・主張・露出を分けて並べる

CONFIRMED / CLAIM / EXPOSUREを混ぜない。件数がなくても、攻撃経路が重要なら残す。

DATE	ORG	SCALE	VECTOR / OBSERVATION	ACTOR / ATTRIB.	STATUS	SOURCE
09/01	Aesto Health	9.54M people	AWS環境への不正アクセス（入口非公表）	Unknown [U]	CONFIRMED	aestohealth.com / HHS
09/01	Novocure	1.4K IDs + <50 PII	一部情報システムに不正アクセス	ShinyHunters [B]	CONFIRMED	Novocure / SEC / DLS claim
09/01–10	Nutex Health	件数未確定	data theft + ransomware / extortion	The Gentlemen [B]	CONFIRMED	SEC / The Record
09/04–11	Florida DMV DAVID	~200K claim	stolen police credential / personal device	ShinyHunters [B]	VECTOR CONFIRMED	FLHSMV / DLS claim
09/03–04	Berlin State	5.7TB claim	ransomware / data theft	Rhysida [B]	INCIDENT CONFIRMED	Berlin / Reuters
09/07	Condé Nast	32.8M listed	database sale; root cause未確定	Unknown [U]	SAMPLE GENUINE?	sample verification

海外事例は、確認済み・主張・露出を分けて並べる

CONFIRMED / CLAIM / EXPOSUREを混ぜない。件数がなくても、攻撃経路が重要なら残す。

DATE	ORG	SCALE	VECTOR / OBSERVATION	ACTOR / ATTRIB.	STATUS	SOURCE
09/18報道	Vietnam APIS	220.8M records	misconfig + default credentials	None / Unknown [U]	EXPOSURE	research reporting
09/14	CenterPoint Energy	7.5M claim	external-facing systemから顧客情報	Unknown [U]	PARTIAL CONFIRM	SEC / Reuters
09/09	AdaptHealth	4.116M people	contractor social engineering → cloud/EHR	ShinyHunters [B]	CONFIRMED	breach notice / reporting
09/12–16	Revolut	~680 customers	偽政府機関のdata request / process abuse	Unknown [U]	CONFIRMED	Revolut / TechCrunch
09/22–29	FBIJobs / PeopleSoft	2–3TB claim	PeopleSoft exploitation。scopeの一部はclaim	UNC6240 / ShinyHunters [A/B]	BREACH CONFIRMED	Mandiant / Reuters
09/24	DriveWealth	count非公表	social engineering → customer data	Unknown [U]	CONFIRMED	DriveWealth / partners

海外事例は、確認済み・主張・露出を分けて並べる

CONFIRMED / CLAIM / EXPOSUREを混ぜない。件数がなくても、攻撃経路が重要なら残す。

DATE	ORG	SCALE	VECTOR / OBSERVATION	ACTOR / ATTRIB.	STATUS	SOURCE
09/26	ATNS (South Africa)	count非公表	ransomware-linked malware / possible exfil	Unknown [U]	INVESTIGATING	ATNS / Dark Reading
09/18 notice	DoD DMDC	3.054M total	file-sharing vulnerability / unencrypted PII	Unknown [U]	CONFIRMED	DoD / Federal News
09/28	DC Medicaid / DHCF	399,086	公開reportのhidden-fields exposure	None / Unknown [U]	EXPOSURE	DHCF / HHS
09/30	FTAPI	count非公表	internal serverへのransomware	The Gentlemen [B]	CONFIRMED	FTAPI / Heise / DLS
09/04 update	Trezor via ShipMonk	80,689 customers	Metabase exploit + 過去データ 残存	ShinyHunters [B]	CONFIRMED	Trezor / extortion emails
07/25~26	Microsoft	130GB / ~8M claim	Power Pages/D365 exposure説。公式未確認	ExfilSquad [C]	CLAIM ONLY	actor claim / reporting

海外事例は、公表主体と規制当局を一次情報として追う

発表ではURLを短縮表記。更新があれば、公表主体・規制当局を優先する。

Aesto Health	aestohealth.com/notice-of-data-security-incident-12-18-25/
--------------	--

Berlin	berlin.de/rbmskzl/aktuelles/pressemitteilungen/2026/...
--------	---

CenterPoint	SEC Form 8-K, 2026-09-14
-------------	--------------------------

Revolut	TechCrunch 2026-09-12 / Euronews 2026-09-16
---------	---

DriveWealth	legal.drivewealth.com/cyber-response-us
-------------	---

Trezor	trezor.io/blog/news/recent-customer-data-exposed-in-shipping-provider-incident
--------	--

DoD / DMDC	Federal News Network 2026-09-28 / TechCrunch 2026-09-30
------------	---

FBI / PeopleSoft	Reuters 2026-09-23, 09-26, 09-29 / Google Mandiant
------------------	--

ITRC statistics	idtheftcenter.org — H1 2026 Data Breach Report
-----------------	--

Anthropic	anthropic.com/research/glm-5-3... / threat-intelligence-report-september-2026
-----------	---

攻撃者帰属は、侵害の確認とは別の確度で管理する

DLS掲載だけで帰属を確定せず、誰がどの根拠で帰属したかを残す。

A VENDOR / TIが帰属

PeopleSoft mass exploitation

UNC6240 / ShinyHunters

Mandiant / GTIGがcampaignを帰属

B 侵害確認 + ACTOR CLAIM

Novocure / Florida DMV / Berlin
Nutex / FTAPI / ShipMonk

ShinyHunters / Rhysida /
The Gentlemen

侵害は確認済み。actorはDLS・恐喝claim

C ACTOR CLAIMのみ

Microsoft

ExfilSquad

公式のbreach確認前。主張としてだけ保持

U 公開帰属なし

Aesto / CenterPoint / Revolut
DriveWealth / ATNS / DMDC 等

Unknown

無理にclusterへ寄せない

注意：FBIJobsの「PeopleSoft経由」はShinyHunters自身の主張。Mandiantが高確度で帰属しているのは、CVE-2026-35273を使ったPeopleSoft mass-exploitation campaign。