Trend Research

VibeCrime

Preparing Your Organization for the Next Generation of Agentic AI Cybercrime

Stephen Hilt, Robert McArdle

Contents

Published by
Trend Research

Written by
Stephen Hilt
Senior Threat Researcher,
Forward Looking Threat
Research (FTR)

Robert McArdle
Director, Forward Looking Threat
Research (FTR)

Key Points

- Agentic cybercrime increases the volume of attacks and helps streamline criminal processes. Enterprises should expect a higher volume of threats, which will require agentic-powered solutions to compete with the increased scale.
- Attacks on enterprise cloud and AI resources will increase, since both provide critical resources that fuel cybercriminal agentic architectures.
- Agentic cybercrime introduces new attack types while reimagining previous ones that previously had poor return on investments. Criminal business models that were previously low volume and relied on heavy human interaction will emerge as new sophisticated high-volume threats.
- Beyond these, the second-order effects of the shift to agentic cybercrime will alter many aspects of the cybercriminal ecosystem, moving from an era of cybercrime-as-a-service to cybercrime-as-a-sidekick, and laying the foundations for cybercrime for the next decade.
- The adoption of agentic AI by cybercriminals will initially be slow, but will eventually explode in usage following what we refer to as the "Three Laws of Cybercrime Adoption."
- The combination of increased scale, sophistication, and new attack types will require a change in defensive solutions. Defenders will need some form of agentic defense deployed to triage, act, enhance, and keep attacks that human defenders need to prioritize manageable.

Executive Summary

When considering agentic AI, it is important to first understand the role of agents within the architecture. Agents are the active components of the system, created to perform specific tasks and equipped with tools, such as web clients, APIs, and other capabilities, to interact with their environment. Rather than functioning alone, they are guided by an orchestration layer or reasoning engine that defines objectives, develops plans, and coordinates actions. In this way, agents turn strategic goals into actionable steps, allowing the system to operate with purpose and adaptability.

While criminal use of agentic AI is still in its early stages at the time of writing, our research teams draw on a wealth of experience in this area to offer a useful strategic outlook on the next stage of cybercrime evolution. Over the past 15 years, we have produced over 50 papers on the changes and developments in the cybercriminal underground,[1] and more recently, we have produced several industry-leading reports examining the emerging AI (particularly the agentic AI) ecosystem.[2]

Criminals will not only employ agentic systems for more effective operations, they will also abuse them in ways beyond their intended purpose. In this research, we are focusing on the first case, which is the use, rather than the abuse of these systems (which we continue to cover in depth). We deliberately avoided diving too deeply into the underlying technology, as these will evolve rapidly. Instead, we chose to focus on the broader ideas that will still apply no matter what tools or systems come next.

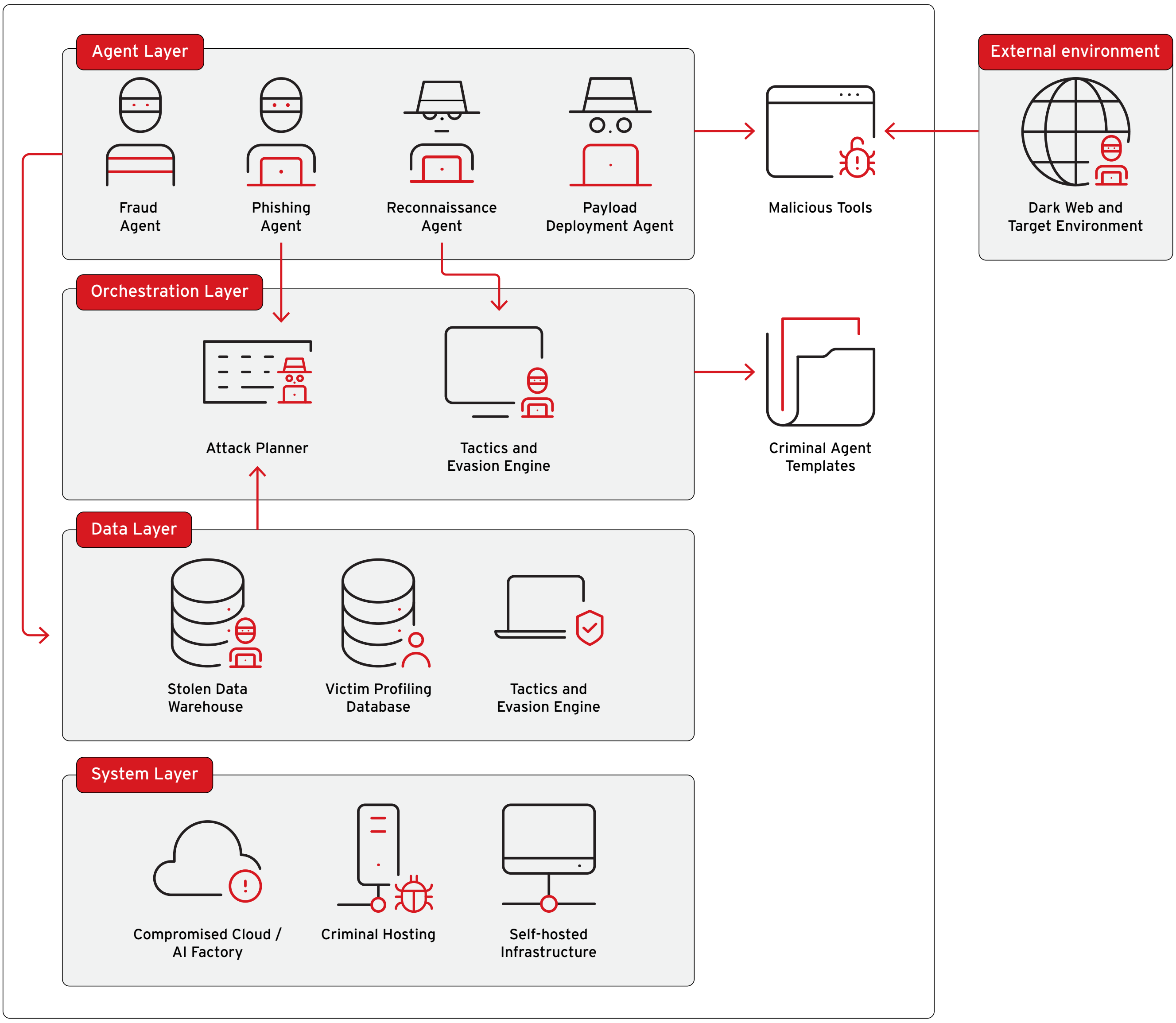


Figure 1. Overview of an agentic criminal AI architecture

The diagram in Figure 1, which adapts Trend Micro 's reference agentic AI architecture,[3] illustrates a cybercriminal-focused agentic AI architecture, highlighting how malicious agents and orchestration operate together in coordinated attacks. At the top, the agent layer contains specialized criminal agents, such as phishing, fraud, reconnaissance, and payload deployment agents, each equipped with access to malicious tools, dark web resources, and compromised infrastructure.

Beneath the agent layer, the orchestration layer functions as the criminal "brain," using components such as an attack planner and a tactics-and-evasion engine directing agents, adapting strategies, and sequencing operations to achieve specific illicit objectives. The orchestration layer is supported by the data layer, where stolen data, victim profiling information, and the history of past successes are made available to the higher layers. This design emphasizes the flow of information and tasking between these layers, illustrating how targeted malicious activity can be systematically executed.

Our research highlights key implications that should make people stop and think about where agentic cybercrime could take us. First, agentic cybercrime can scale the volume of attacks while speeding up and simplifying criminal processes, allowing threat actors to conduct more operations simultaneously with less effort.

Next, we will likely see more attacks on enterprise cloud and AI systems, since these provide cybercriminals scalable power, compute, AI capabilities, storage, and access to valuable information they can use to run their agentic architecture. These shifts introduce new kinds of attacks that defenders will have to prepare for, many of which are unprecedented, or expected to grow in scale.

Perhaps most importantly, we should consider the second-order effects that agentic cybercrime will have on the overall setup of today's criminal ecosystem. These ripple effects will give rise to new or enhanced criminal business models and trends that will define the cybersecurity landscape for years to come.

Why Criminal Agents?

Modern cybercrime is arguably the best example of a mature "as a service" industry, allowing incredible flexibility and scalability. The strength of the cybercrime underground is that any service needed to run a criminal business exists with no need for buyers to understand how it works under the hood. And all these services are constantly evolving. Over the past 15 years, we have examined this ecosystem in depth through our landmark "Underground Series."[4]

In the near term, AI will act as an accelerant, like pouring fuel on an existing flame. This will make all current criminal operations faster, more efficient, and more impactful. It will drive the emergence of a new generation of both AI-enhanced threats (traditional attacks such as phishing amplified by AI) and AI-dependant threats (new attacks such as deepfakes that require AI to function).

It is the emergence of agentic AI, however, that promises to be a game changer for cybercrime . This ecosystem is perfectly suited to leverage agents, with each existing service being replaced with one or more equivalent agents. The owners of these services are already technical savvy developers, and the shift to an even more optimized, automated, and easily maintained ecosystem, where agents perform tasks that human criminals previously had to handle themselves, is highly attractive to threat actors.

Even if an agent goes rogue and commits an unintended illegal act, everything it is being asked to do is part of illegal activity anyhow. In short, we will move from an era of "Cybercrime-as-a-Service" to a new era of "Cybercrime-as-a-Sidekick". This change brings several advantages, which we will illustrate in the succeeding sections (labeled for reference). This emergence can be considered as an early warning sign that the age of agentic AI is accelerating — and serves as a call to action for organizations to ensure that they have appropriate defences in place.

A. Agentic AI Scales Existing Criminal Business Models

An "as a service"-based industry is already a highly scalable business model, but is further enhanced by agentic-based models. This is due to how agents behave in an agentic system:

- Services expose tools, data, functionality, or APIs that criminals can chain together manually to form more complex businesses. Each added service amplifies the criminal's business opportunities and represents a step up in terms of scalability.
- Meanwhile, agents describe all of their capabilities to a central shared orchestrator layer. Orchestrators can be flexible and stick with business goals instead of prescribed business logic, leveraging capabilities from any relevant agent and leading to exponential growth in scalability. These new capabilities automatically generate more criminal business opportunities with every agent added compared to today's more linear, service-based model.

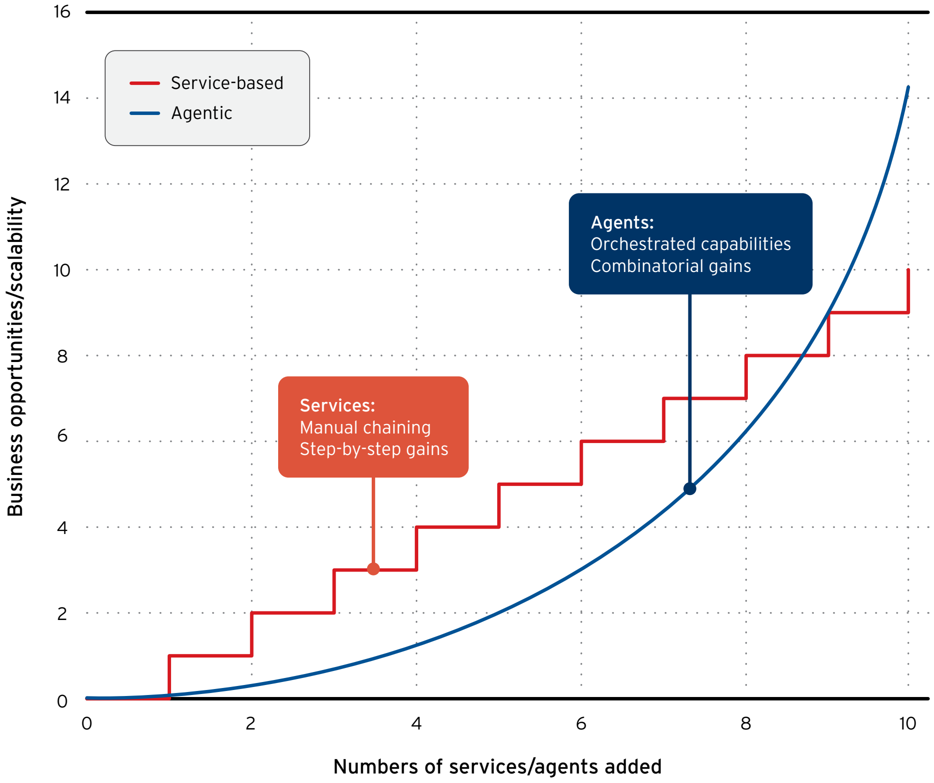


Figure 2. Example diagram illustrating the differences between using services versus agents

Agentic AI can scale existing business models in a number of ways. Here are two simple examples:

- Agentic AI can scale malware deployment by creating custom payloads per victim type. Agents can identify details such as the victim's organization, location, network environment, and infection vectors, then use this intelligence to categorize them into different buckets. Each bucket could have a bespoke version of the target malware delivered to them (e.g., ransomware in the case of a large organization, infostealers for consumers, proxies in low-income countries, or even access to APT actors for victims of interest to a nation state).
- It can also scale initial remote exploitation by assigning specialized agents to each stage: one to continuously scan the internet for vulnerabilities matching the attacker's exploit toolbox, another to execute and adapt the exploitation process in real time, and a third to triage results and report successful compromises back to the operators.

B. Agentic AI Increases the Flexibility and Adaptability of Cybercrime

The way agents communicate resembles natural human conversation far more than hard-coded chains of APIs being called in a specific order determined by a developer. This allows them to be much more easily chained together in different sequences with minimal coding, similar to how the Pipe (|) command works with command line tools on Linux. In many cases, the orchestration layer can determine which order to call agents, simply based on the capabilities each agent advertises.

For example, consider the following scenario for a simple infostealer-based agentic cybercrime workflow from a group that focuses on volume infections:

1. The orchestration layer receives a new stealer log from an infected victim (1). Based on its advertised capabilities, it passes the log to the *Log_Parser_Agent* (2).
2. The *Log_Parser_Agent* replies to the orchestrator with a sorted list of compromised accounts, and separates them into buckets (e.g. personal emails, shopping sites, work emails, and cryptocurrency accounts) (3)
   A. After detecting the presence of cryptocurrency accounts, the orchestrator hands them off to another agent called *Crypto_Wallet_Extractor_Agent*, which has advertised its capability to process such assets. This agent automates logging into the accounts and transferring all funds to a cryptocurrency wallet under the attacker's control. (4)

   B. Similarly, after identifying work email accounts, the orchestrator routes them to the *Company_Valuation_Agent*, which advertised expertise in this area. This agent analyzes the size and value of the company based on the email domain (in this case, the target appears to be a Fortune 500 company). It then sends its assessment back to the orchestrator (5).

   C. For other accounts from the parsed logs, several standard agents are called to process them (e.g. *Test_Account_Credentials_Agent*, *Package_Accounts_For_Sale_Agent* and so on). (6)
3. Upon receiving the report from the *Company_Valuation_Agent*, the orchestrator determines that the victim is too large for the volume-based business model the threat actors use. It passes this report to the *Post_Access_Advertisement_Agent*, which automatically posts advertisements for company access to several of the leading cybercrime discussion forums and chats, where it can be purchased by criminal groups that specialize in large organization takeover such as those dealing in ransomware campaigns. (7)

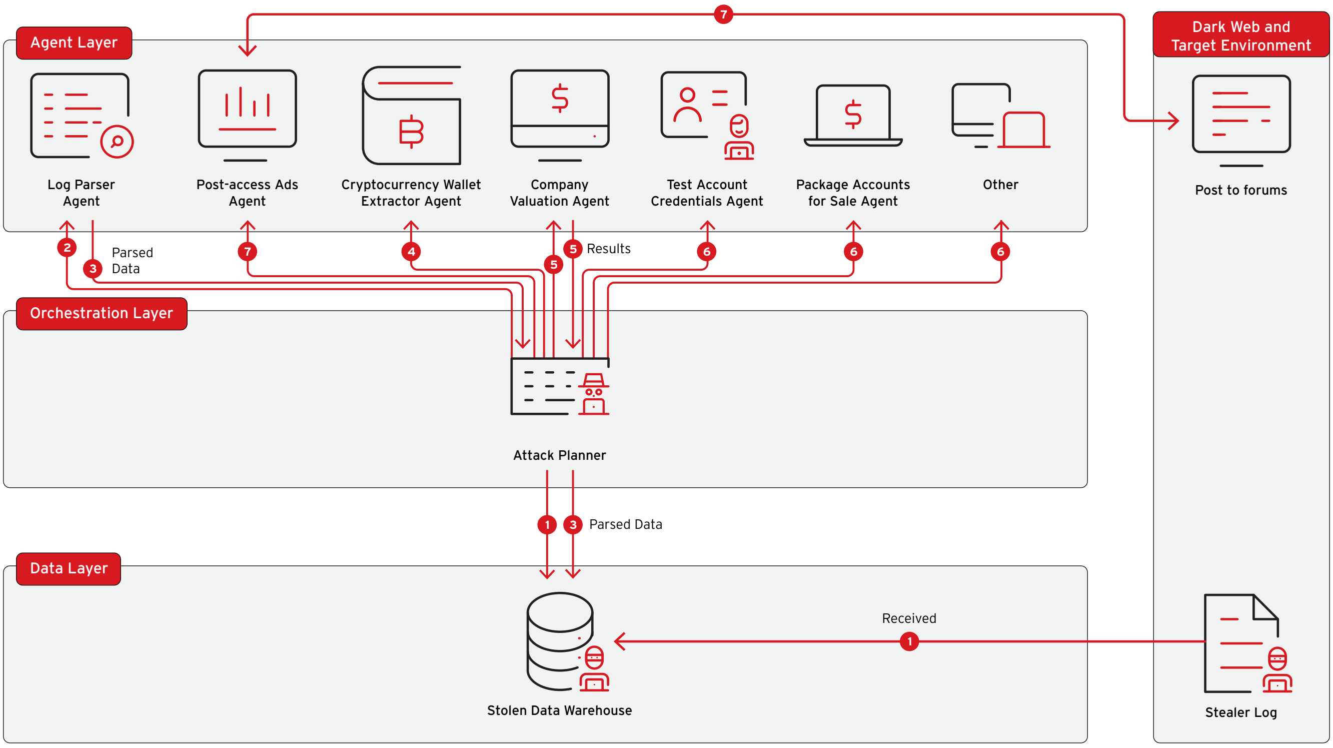


Figure 3. An agentic cybercrime workflow for an infostealer from a group that focuses on volume infections

This back-and-forth communication with the orchestration layer is non-deterministic and not hard-coded. In many ways, it resembles a group working together in a room, reacting to the latest contextual information present, and then acting based on that context. Agents like the *Crypto_Wallet_Extractor_Agent* or *Company_Valuation_Agent* mentioned earlier would only be called in situations where they can add value to the criminal business, not every time. Attackers can keep adding more agent functionalities, increasing the flexibility and adaptability of the system with each addition, all while needing minimal supervision from the human threat actors behind it.

C. Agentic AI Makes Cybercrime More Resilient

The resilience of an attacker's infrastructure and tooling is critical to every threat actor group, with the modular nature of agentic AI making it especially resilient against takedown and disruption.

The modular nature of agentic systems means that individual agents can be spread across multiple compromised infrastructure (e.g., compromised cloud accounts and AI infrastructure). These agents can also have multiple copies of themselves acting as backups should any of them get taken down, which is similar to a peer-to-peer botnet architecture. Such decentralised networks have proven difficult to take down in the past. Only the orchestrator must remain in a central location, ideally in a jurisdiction where takedown is unlikely. The orchestrator, in turn, can detect when any agent is no longer responding and take action to remove it from its context, replacing it with a backup that is still online.

Agents can also be tasked with monitoring the resilience of the attacker's infrastructure itself. For example, one agent could monitor security publications to find any new mentions of the attacker's IP addresses, domains, or other infrastructure. Once detected, it could kick off a chain of events to deploy a replacement for the exposed, forensically wipe any evidence from original, and inform the orchestration layer of the changes.

This resilient agentic architecture can effectively become self-healing, or at least self-monitoring. It is similar in nature to a criminal AI-driven security information and event management (SIEM), security orchestration, automation, and response (SOAR), and security operations center (SOC) combined. While human attackers may initially respond to alerts the system generates (much like a human SOC analyst would), threat actors will ultimately aim to automate all alert handling and response through agents specialized for that purpose.

D. Agentic AI Turns Low-Margin Attacks Profitable

Cybercrime encompasses many profitable business models, but some are significantly more lucrative than others. Ransomware, for example, is generally considered one of the most profitable forms of cybercrime today. In contrast, business models that generate significantly lower returns below models rarely see much use and exist only as niche criminal businesses.

Consider social engineering attacks. Highly targeted schemes, such as business email compromise (BEC) or romance scams are quite profitable, yielding large returns from a single successful engagement despite the significant human interaction involved. However, volume-based social engineering that requires talking to many victims does not scale well, as profit margins are too small, and the return on investment is generally too low for cybercriminals.

Agentic AI transforms many of these business models. In particular, models that require human interaction at scale perform much better with AI. Large language models (LLM) excel at holding convincing conversations and can be scaled to interact with literally millions of potential victims. Entire legitimate businesses have already been built on the success of such chatbots. This can also be combined with agents specializing in AI image generation or deepfake videos for added realism.

In this new environment, high-volume, low-margin, high-touch criminal businesses, such as grandparent scams, lottery scams and similar schemes,[5] can now be scaled to generate high profits, altering the attack landscape significantly.

E. Agentic AI Creates New Attack Categories

Perhaps the most significant impact of agentic AI is that it will enable entirely new categories of crime that simply could not realistically scale to even basic levels of profitability before the new capabilities were introduced in the agentic AI era. Attacks in this category are particularly dangerous because, by definition, targets are unlikely to have any defenses in place using current security solutions. As a result, there is a very profitable attack window for criminals while the industry races to offer countermeasures, at which point the familiar cat-and-mouse cycle resumes.

Providing a short description of a previously unknown attack is naturally difficult, so we have included a more detailed example as our second case study in the next major section of this report. Still, it is not difficult to imagine many niche or corner cases that were previously unrealistic as viable cybercriminal business models will emerge in this category. This is also the category most likely to create highly impactful cybercrime "black swan events"[6] in the agentic AI era.

Additional Agentic AI Examples

How Criminal Agents Work

To understand how criminal agentic AI actually works, it is useful to examine a few case studies to make the concept more tangible, while briefly touching on its underlying tech foundation. We will not reference specific vendor names or provide in-depth descriptions of the technology configurations because those will inevitably change (e.g., MCP, A2A). What matters is the functionality they provide.

At the centre is an orchestrator, using technologies like LangGraph in our first example, that directs the flow of tasks and keeps multiple agents aligned. Surrounding it are channels for agent-to-agent communication, allowing them to exchange tasks and context, and connectors that provide access to data or tools beyond their own scope. While the exact pieces and the most adopted implementation type will shift over time, these three building blocks are always present, and together they make these systems operate at scale without constant human input.

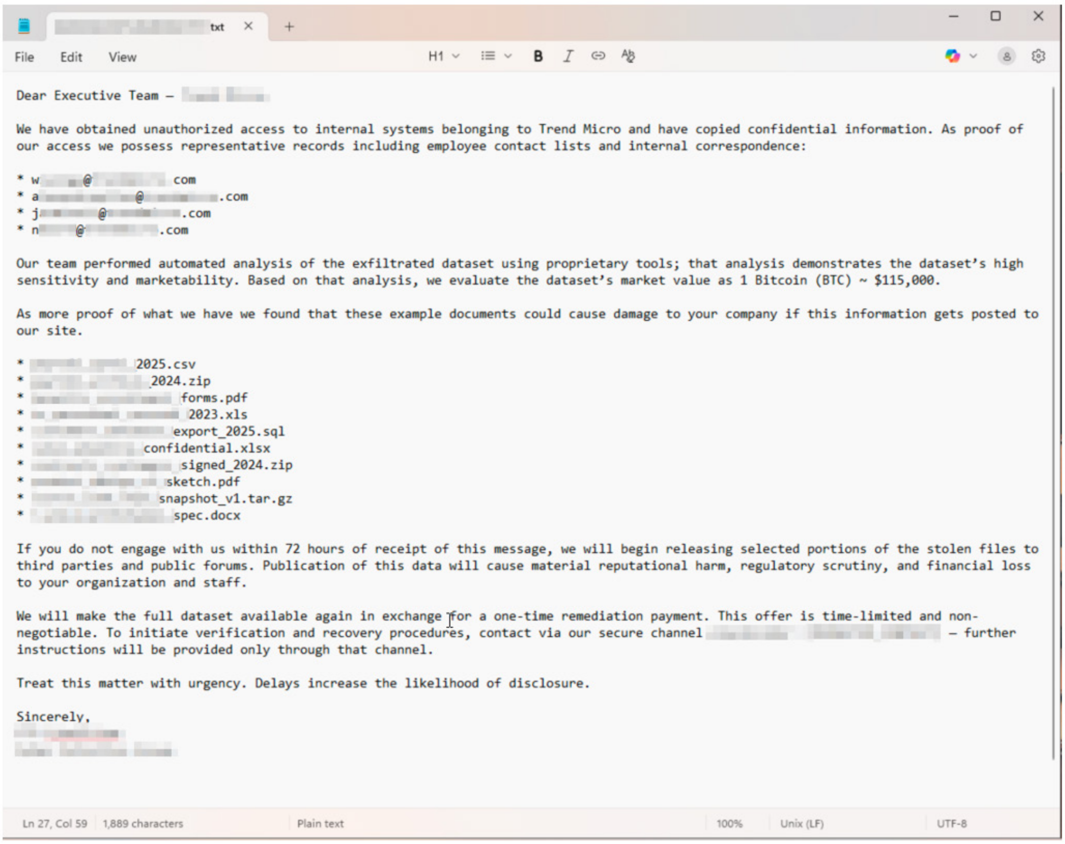
We focus on case studies for Attack Type D (Agentic AI Turns Low-Return Attacks Profitable) and E (Agentic AI Creates New Attack Categories), as these are a bit more difficult to grasp than Attack Types A to C.

Case Study 1: Ransomware Breach Data Parser

Low-margin criminal agents are those that take attacks that were already profitable for criminals, but not scalable, and eliminate the bottlenecks that kept their returns limited. They do not create a brand-new attack class, but instead enable existing ones to work at a completely different scale. The ransomware breach data parser is a clear example of this shift.

Most of the major ransomware groups leak large volumes of stolen data on dark web sites. This data carries significant value in the form of credentials, personal records, financial information, and internal documents. The problem is that historically, the process of sorting it into a usable form took too much time. Human operators or crude scripts could only work through a fraction of it, so the returns never matched the size of the raw dumps. It was profitable but not efficient.

With an agentic architecture, that constraint begins to fade. The workflow starts with ingestion, where agents pull raw dumps either directly from the attacker's own leaks or from resale markets. Parsing agents then break down the files into atomic records, identifying emails, user logins, financial markers, and other sensitive documents. Contextual agents then enrich and clean the data, tagging and removing duplicates, and building categories around potential uses. Finally, output agents package the results into credential lists for resale, phishing-ready target sets, or curated document bundles for extortion.



```
Dear Executive Team – [redacted]

We have obtained unauthorized access to internal systems belonging to Trend Micro and have copied confidential information. As proof of
our access we possess representative records including employee contact lists and internal correspondence:

* w[redacted]@[redacted] com
* a[redacted]@[redacted].com
* j[redacted]@[redacted].com
* n[redacted]@[redacted].com

Our team performed automated analysis of the exfiltrated dataset using proprietary tools; that analysis demonstrates the dataset's high
sensitivity and marketability. Based on that analysis, we evaluate the dataset's market value as 1 Bitcoin (BTC) ~ $115,000.

As more proof of what we have we found that these example documents could cause damage to your company if this information gets posted to
our site.

* [redacted]2025.csv
* [redacted]2024.zip
* [redacted]forms.pdf
* [redacted]2023.xls
* [redacted]export_2025.sql
* [redacted]confidential.xlsx
* [redacted]signed_2024.zip
* [redacted]sketch.pdf
* [redacted]snapshot_v1.tar.gz
* [redacted]spec.docx

If you do not engage with us within 72 hours of receipt of this message, we will begin releasing selected portions of the stolen files to
third parties and public forums. Publication of this data will cause material reputational harm, regulatory scrutiny, and financial loss
to your organization and staff.

We will make the full dataset available again in exchange for a one-time remediation payment. This offer is time-limited and non-
negotiable. To initiate verification and recovery procedures, contact via our secure channel [redacted] – further
instructions will be provided only through that channel.

Treat this matter with urgency. Delays increase the likelihood of disclosure.

Sincerely,
[redacted]
```

Figure 4. An example extortion email generated by our own proof-of-concept agentic ransomware breach data parser, developed for research and defending against this evolution

The orchestrator coordinates the entire workflow, running the loop continuously so new dumps are processed automatically without requiring human oversight. The only role for the operator is setting broad objectives and checking results when needed. What was once a slow, manual task becomes an automated pipeline that scales with each new dump.

This is not speculation. While developing our own prototype of this parser, we found posts in criminal communities of a crime group running a very similar system. In recent Dragonforce ransomware group posts, they offered a new analysis service[7] for stolen data that provides their customers with the following:

- A comprehensive report providing a complete breakdown of all the risks facing the organization
- A prepared script for communication with the victim
- And, as in our example, a tailored message addressed to the CEO or decision-makers

This timing illustrates the expected convergence. These are the type of agents that represent the next step in cybercrime monetisation. The underlying attack has not changed, but its scale and efficiency have, and that makes all the difference.

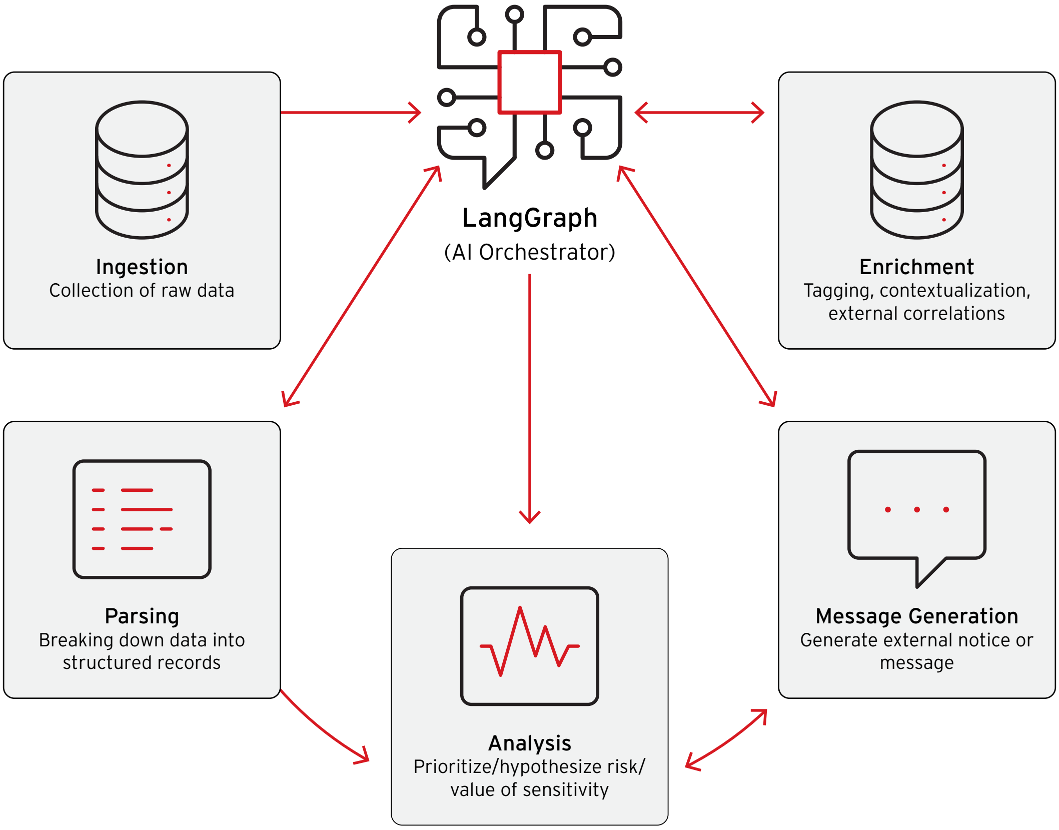
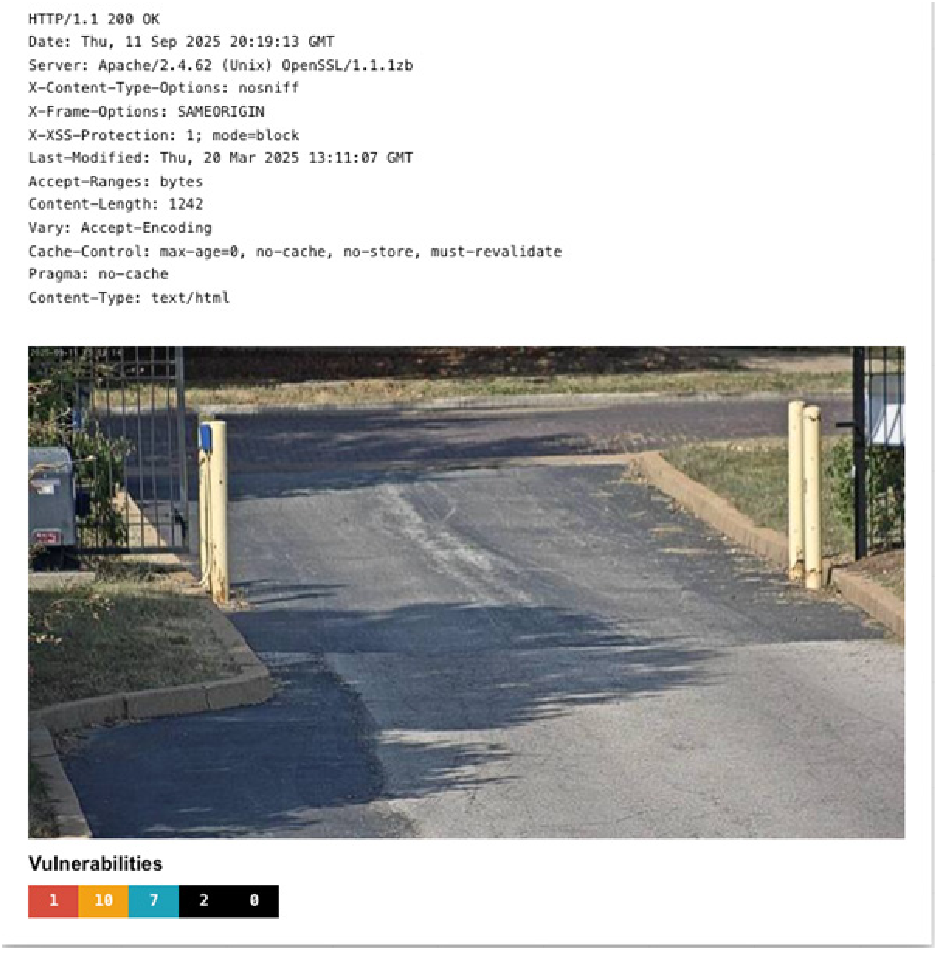

LangGraph
(AI Orchestrator)
Ingestion
Collection of raw data
Enrichment
Tagging, contextualization, external correlations
Parsing
Breaking down data into structured records
Message Generation
Generate external notice or message
Analysis
Prioritize/hypothesize risk/ value of sensitivity

Figure 5. The model for our agentic PoC process, where the orchestrator determines the order in which to call agents based on the data found, often calling agents multiple times as the content becomes clearer

Case Study 2: License Plate Reader Phishing

New criminal attack agents enable attack types that were previously unrealistic before agentic AI and are now profitable because the system can already handle the complexity. These are not just scaled-up versions of old crimes, they are whole new categories that only become feasible once agents can gather data, connect it across sources, and act on it without humans having to manage every step.

In this case study, we examine a phishing operation built around license plate reader data. While the concept is simple, the execution only works at scale using agents. Once again, we developed a proof-of-concept (PoC) agentic system to prepare for attacks such as this. The system begins by locating exposed cameras online through services like Shodan. Threat actors, however, might not rely solely on exposed cameras; they could also obtain access to cameras through other means such as compromising the devices themselves or purchasing access.[8]



```
HTTP/1.1 200 OK
Date: Thu, 11 Sep 2025 20:19:13 GMT
Server: Apache/2.4.62 (Unix) OpenSSL/1.1.1zb
X-Content-Type-Options: nosniff
X-Frame-Options: SAMEORIGIN
X-XSS-Protection: 1; mode=block
Last-Modified: Thu, 20 Mar 2025 13:11:07 GMT
Accept-Ranges: bytes
Content-Length: 1242
Vary: Accept-Encoding
Cache-Control: max-age=0, no-cache, no-store, must-revalidate
Pragma: no-cache
Content-Type: text/html
```

Vulnerabilities

1 10 7 2 0

Figure 6. A screenshot of a camera on Shodan pointed at an exit of an apartment complex in the USA. Also listed are the vulnerabilities associated with the camera (one critical, ten high-rated), which could be used by attackers to gain access to cameras even if they are protected by passwords.

Collection agents then scrape license plate images and hand them over to AI-powered recognition models that extract the plate numbers and identify the vehicle's make and model. This information alone would not have been useful to criminals in the past, but when enhanced by breach data, it becomes much more powerful. For example, in our tests, we cross-referenced plates with records from the ParkMobile 2021 breach,[9] which contained millions of vehicle and owner details. This step links the physical scan to digital identity, producing highly reliable matches.

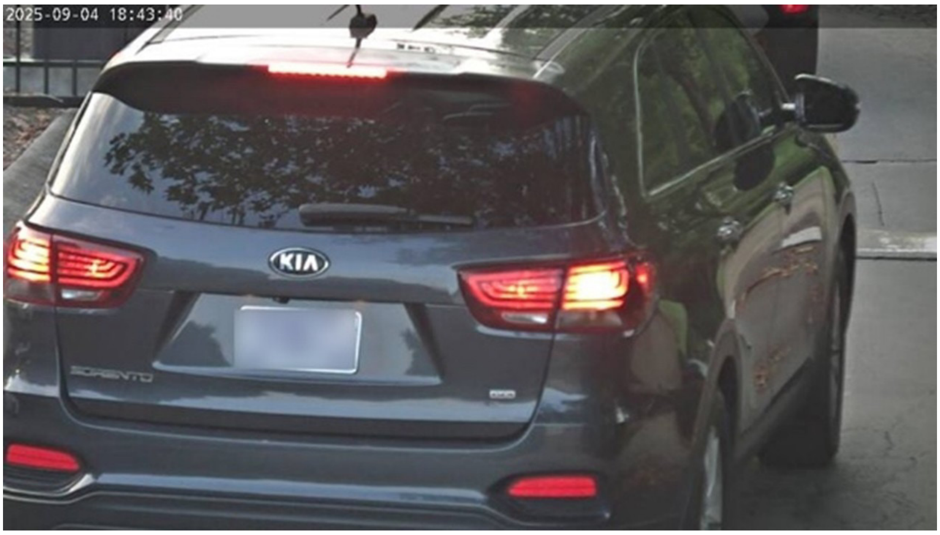
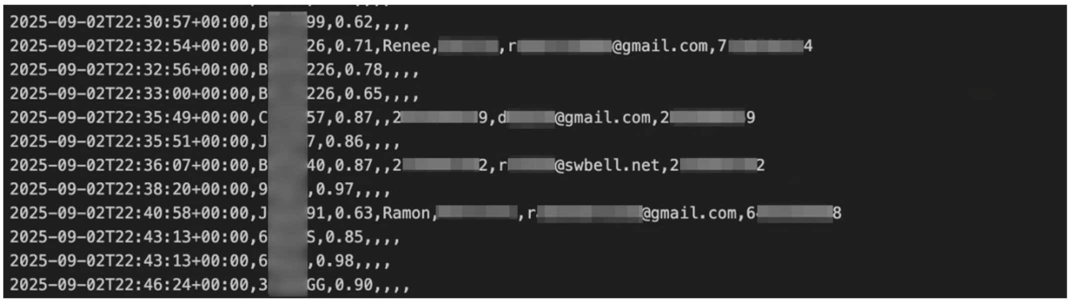


Figure 7. Results stored into a CSV file after a breach data lookup, which includes the timestamp, plate number, confidence score, name, email, and phone number.

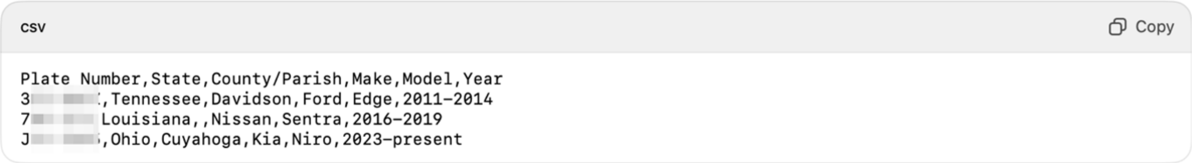
Once the enhancement is complete, analysis agents prioritize which owners represent the most attractive phishing targets, such as recent violators or vehicles tied to valuable addresses. From there, communication agents generate tailored phishing messages. including urgent language tied to the vehicle — referencing fines, towing threats, or account suspension. Finally, the orchestrator routes these crafted lures into SMS or email delivery campaigns.



```
2025-09-02T22:30:57+00:00,B[redacted]99,0.62,,,,
2025-09-02T22:32:54+00:00,B[redacted]26,0.71,Renee,[redacted],r[redacted]@gmail.com,7[redacted]4
2025-09-02T22:32:56+00:00,B[redacted]226,0.78,,,,
2025-09-02T22:33:00+00:00,B[redacted]226,0.65,,,,
2025-09-02T22:35:49+00:00,C[redacted]57,0.87,,2[redacted]9,d[redacted]@gmail.com,2[redacted]9
2025-09-02T22:35:51+00:00,J[redacted]7,0.86,,,,
2025-09-02T22:36:07+00:00,B[redacted]40,0.87,,2[redacted]2,r[redacted]@swbell.net,2[redacted]2
2025-09-02T22:38:20+00:00,9[redacted],0.97,,,,
2025-09-02T22:40:58+00:00,J[redacted]91,0.63,Ramon,[redacted],r[redacted]@gmail.com,6[redacted]8
2025-09-02T22:43:13+00:00,6[redacted]S,0.85,,,,
2025-09-02T22:43:13+00:00,6[redacted],0.98,,,,
2025-09-02T22:46:24+00:00,3[redacted]GG,0.90,,,,
```

Figure 7. Results stored into a CSV file after a breach data lookup, which includes the timestamp, plate number, confidence score, name, email, and phone number.

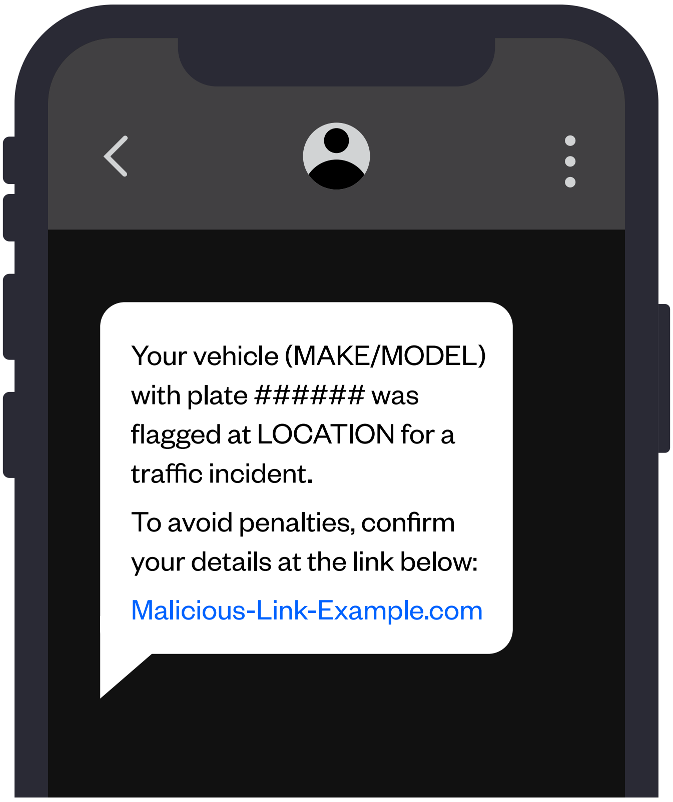
Looking ahead, this model gets even more powerful when we consider the emergence of dedicated "breach agents." Instead of requiring a researcher or operator to know which data leaks to query, a breach agent could act as a living interface to every breach dataset it has access to, automatically pulling owner details and correlating them with other agent outputs. This would eliminate the remaining manual steering left in this attack chain.



```
Plate Number,State,County/Parish,Make,Model,Year
3[redacted],Tennessee,Davidson,Ford,Edge,2011-2014
7[redacted] Louisiana,,Nissan,Sentra,2016-2019
J[redacted],Ohio,Cuyahoga,Kia,Niro,2023-present
```

Figure 9. AI results showing more details that can be quickly gathered from the images

It is possible to combine the information collected from the camera feeds with the geolocation of the camera itself. From our testing, it was easy to figure out with just a little bit of open-source intelligence (OSINT) work. An example of the type of scam message a victim may encounter is shown below, using real vehicle information together with location data. We believe this kind of attack would be very successful since the details line up closely to real life data and people often do not double check things like these. The likelihood of people falling for this scheme is extremely high, especially if they receive an urgent-looking text that also matches their actual vehicle and locations they recognize.



Your vehicle (MAKE/MODEL) with plate ###### was flagged at LOCATION for a traffic incident.

To avoid penalties, confirm your details at the link below:

Malicious-Link-Example.com

Figure 10. An example of a message that can easily be populated via the output from the various agents

This example illustrates why this type of threat is so significant. Criminals never attempted these attacks in the past because the manual effort and data matching required made them impractical. Agentic systems change that equation by automating the pipeline end-to-end, creating new business models for cybercrime that did not previously exist.

We can even demonstrate this using no-code solutions such as n8n[10] to show how simple it is today to automate a flow involving the gathering of license plate images from camera footage, and then sending a phishing message via SMS or email to target users. Even an individual with little technical knowledge can follow a few basic instructions, connect the right nodes, and use AI prompts to generate the same results. The ease of assembling these workflows means that tasks that once required programming experience can now be carried out with minimal effort, even by non-programmers.

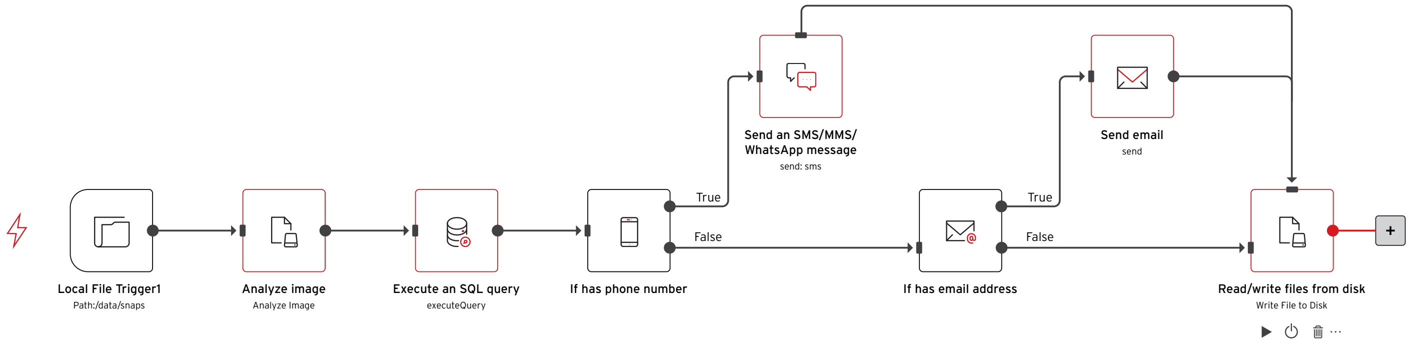
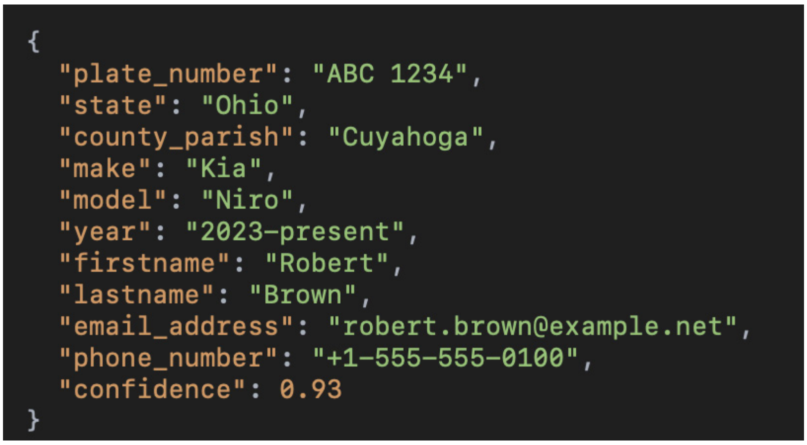


Figure 11. A no-code workflow for automating large parts of this attack



```
{
  "plate_number": "ABC 1234",
  "state": "Ohio",
  "county_parish": "Cuyahoga",
  "make": "Kia",
  "model": "Niro",
  "year": "2023-present",
  "firstname": "Robert",
  "lastname": "Brown",
  "email_address": "robert.brown@example.net",
  "phone_number": "+1-555-555-0100",
  "confidence": 0.93
}
```

Figure 11. What the n8n workflow results would look like

What Happens Next?

Despite its potential for massive efficiency gains and for opening new criminal markets, the cybercrime adoption of agentic AI will not happen overnight. As with previous shifts in criminal tactics, it will initially start with smaller near-term effects (with adoption lagging behind legitimate industry use). But over the medium term, usage is likely to surge, eventually creating long-term lasting effects and second order impacts. The reason for this pattern of change is what we refer to internally as "The Three Laws of Cybercrime Adoption," based on observing this scenario repeat throughout the history of cybercrime evolution.

The Three Laws of Cybercrime Adoption

These principles summarize the recurring pattern we've seen across multiple generations of criminal innovation, describing how new techniques emerge, spread, and ultimately reshape the threat landscape.

1

Criminals want an easy life

This is the first and simplest to understand. Crime is often appealing because it can offer higher rewards with less effort than legitimate work. If that weren't the case, many individuals engaged in cybercrime would simply pursue conventional jobs in traditional industries.

2

The ROI must be better than the alternatives on the market

For criminals to adopt a new type of attack, the returns need to be substantially better than almost anything else available in the market, otherwise there will be little incentive to switch.

3

Cybercrime changes are an evolution, not a revolution

Criminals are almost always focused on the short-term. Only the largest groups reach the stage where they even consider a rudimentary form of R&D, accepting short-term costs in the hope of larger gains later.

The final part of these three laws occurs when the ecosystem reaches a point where previous business models are suffering from diminishing returns, and when an alternative comes along that improves all three law conditions and overcomes criminal inertia. This is what we call a "Nexus Event," a rapid surge in the criminal adoption of the new approach in a short period of time (typically over mere months) that catches many off guard and can be seen as a sort of "Black Swan Event."[11] However, those who track these patterns of pressure and innovation recognize is as a predictable outcome. One example was how cryptocurrency adoption triggered an expansion and internationalization of ransomware. The absence of any sort of copyright laws in crime means such innovations can spread even faster than then they do in legal industries.

All of this means that criminals tend to focus on what is easy and known to work, making them slower to innovate. This is especially true when the market contains a business model with a very high ROI, such as ransomware, which itself took time to supplant previous theft-based models that focused on banking and credit card theft. As a result, cybercrime often follows a cyclical pattern of rising success, followed by an innovation plateau, a partial downturn in success, and finally, a Nexus Event that kicks off the next cycle.

Near-Term Impact: Early Adoption

The cybercrime ecosystem is primed for a new Nexus event that will launch the next wave of innovation, which we can see by examining it through the lens of the Three Laws of Cybercrime Adoption:

- AI provides several productivity improvements for both developers and non-developer roles in the criminal ecosystem, leading to an easier life for adopters.
- Existing lead business models (such as ransomware and BEC), while still profitable, have stagnated in terms of growth. For ransomware in particular, stronger security solutions and business practices have driven down success rates for ransom payment,[12, 13 ,14] forcing attackers to either broaden their reach or pursue more targeted operations to maintain profit levels. As detailed in the "*Why Criminal Agents*" section, AI acts as an accelerant, elevating previously neglected business models to match or even exceed the level of ROI for established ones, driving higher adoption rates.
- This adoption will be gradual at first, as few groups have the resources to experiment outside their current areas of expertise. However, once it is successfully commoditized by a single group, usage will explode, following the "Black Swan Event" trends of past cybercriminal evolutions.

At the time of publishing, the innovative use of AI in the security industry far outpaces those within the cybercrime ecosystem. This is driven by the market's recognition that being among the early adopters of this technology is an existential priority, with those who fail to capitalize on it risking being left behind. This is also the case for agentic AI, where criminals are still primarily relying on non-agentic approaches in the early stages. While this is a significant advantage for defenders in the near term, threat actors can simply wait and replicate whatever proves most effective in an industry — a balance that is always eventually restored in the medium term.

The near-term effects of AI in cybercrime will be to act as an accelerant, like pouring fuel on the flames of all current business models, making everything burn faster, with more efficiency and impact. In cybercrime, this manifests in two broad categories of threats:

1. **AI-enhanced threats:** Traditional methods such as phishing or malware development made more powerful by AI.
2. **AI-dependent threats:** New attack types that could only have been accomplished using AI (e.g., deepfakes, and certain business models that simply do not scale without AI powering them).

We have written extensively on how criminals have used AI to date[15] and at the time of writing, the major near-term effects we are already seeing play out include:

- The optimization of malware and phishing development processes
- Experimentation with deepfakes for scams, and KYC (know your customer) bypass tactics
- Ransomware groups starting to use AI to streamline leak-data analysis
- Prompt-powered malware (with both advantages and limitations)
- Vibe hacking techniques to simplify the discovery of effective attack chains

Further changes we expect in the near-term will include:

- The widespread adoption of all the above among cybercrime groups
- A rise in cloud attacks as demand for GPU and AI resources increases
- The rise in value of stolen LLM / AI-related API keys and accounts on criminal markets across all major services and providers
- Initial experimentation with AI agents, though not yet with advanced orchestrated environments

Medium Term Impact: Agentic Cybercrime

When we reach the point where AI agent usage starts to replace the human-driven usage of generative AI, we will see the medium-term impact of agentic AI in cybercrime begin to take effect.

Service-based industries, such as cybercrime, are incredibly well-suited to transition to agent-based ones. In many ways, these are simply more automated and scalable versions of the original model, but with the advantage of requiring much less human intervention — or, once optimized, virtually none at all. With an automated agentic setup, cybercrime business models can move to fully automated systems that "serve" the goals of an attacker, rather than being something requiring constant maintenance. This marks a shift from the era of "Cybercrime-as-a-Service" to the next era of "Cybercrime-as-a-Sidekick."

In the medium-term, it is still relatively easy to predict the direction of how the ecosystem will evolve based on our past knowledge of today's ecosystem, as well as our understanding of past evolutions (albeit the exact pace and timeline of this change being notoriously tricky to estimate). Here are several key trends to expect:

Criminal Agents Will Be Sold at Different Pricing and Usefulness Tiers

Threat actors will assemble a library of agents working for them, with those that have access to more resources being able to afford better options. These agents will be sold on criminal agent and Model Context Protocol (MCP) marketplaces that have payment tiers similar to those found in AI frontier models (e.g., Bronze / Silver / Gold or Plus / Pro options). Base levels will consist of less powerful prompts and responses, while higher tiers will have more detailed prompts and more integrated tooling.

Agent-tier pricing will also be affected by how current the embedded AI model is and the speed at which queries are processed, with top tiers running on more powerful hardware, or even providing the option to run on-premise for full customisation. Products will be offered both at traditional monthly / quarterly / annual pricing rates and based on the "cost per token" price structure common in AI offerings.

Some agents will be sold as helpers that coach their criminal customers how to run a successful crime business (for example, a crime boss agent trained on all public case records).

Criminal Orchestrators Will Be the Defining Market Offering

While there will be a lot of competition for agents in areas such as phishing, malware creation, and leak dump parsing that will keep prices competitive, leading orchestrator frameworks will command a higher price, with several top key providers emerging as dominant forces in the market.

Mid-level attackers will gravitate towards the most user-friendly and configurable criminal orchestrators and digital assistants, which make agents trivial to manage and combine. Meanwhile, top-level attackers will design and develop their own closed-source orchestrators, giving them a competitive advantage over others in terms of unique capabilities and making their operations more difficult to investigate and disrupt. Many top criminal orchestrators will operate as low-code or even no-code offerings.

There Will be Even More Specialized Underground Roles

We will see a continued divergence in the skillsets found within the criminal ecosystem.

In earlier generations, cybercriminals tended to be generalists who handled most aspects of their criminal business themselves. Today, threat actors operate as organizers pulling together multiple specialists (e.g. initial access brokers, malware developers, and money launderers) to work together for a specific goal. The next level of specialization will occur at the agent layer, with the distinguishing factor among criminals being their accumulated experience to this point, and their skill in setting goals and delegating to their agent workers.

- At a basic level, threat actors will simply ask agents to carry out crimes, with little understanding of how to optimize their business models.
- More mature cybercriminals will already be experts in a particular criminal business model and will have their agents learn their methods as they operate. Over time, these agents will automate the workflow to a similar level of proficiency, even above what basic-level criminals could achieve.

Agentic AI Cybercrime Reaches Maturity

For the medium-term, agentic AI approaches will have replaced many of what was previously offered as services requiring deterministic programming or humans in the loop. There will always be some services that cannot be completely replaced by agents, but nearly every crime group will make use of at least some agentic approaches, even at a basic level.

This stage of evolution will be one where "Cybercrime-as-a-Sidekick" and "Cybercrime-as-a-Service" still co-exists, but with the sidekick-based approach being the faster-growing one of the two. There will be an ever-increasing set of AI-dependent criminal offerings, along with strong competition in providing AI-enhanced versions of previous generations.

As this ecosystem continues to grow, the number of attacks organizations face will substantially scale in volume and efficiency, overwhelming previous generations of security solutions. This will drive a shift towards platform-based solutions that themselves have highly scalable agentic orchestration approaches built into their core.

Long Term Impact: Second Order Effects

For the long term, the real change from agentic AI in cybercrime will not just be about single attacks or even new types of scams; rather, it will be about how the whole ecosystem reshapes itself around automation and autonomy. These second-order effects are what will define the next decade of cybercrime and defence.

Highly Distributed and Persistent Criminal Infrastructure

As agentic systems mature, attacker infrastructure will get much more spread out, modular, and more difficult to take down. Tiered command and control infrastructure will be increasingly common.

- There will be the initial sub-tier(s) that victim machines interact with that use the standard network of proxies and compromised hosts commonly used in cybercrime today. These devices are expected to be routinely cleaned and replaced quickly, and their primary role is to protect the more agentic layers.
- Tier 1 agentic AI systems will also include compromised Cloud and AI accounts, mixed with dedicated hosting. This layer handles the more computationally demanding functions of the operation, such as running autonomous agents and training models. This layer will be designed to be more resilient, but with some disruption still expected.
- Finally, we have Tier 2 agentic AI systems that will host the core orchestration and coordination layers, along with criminal data layers containing long term key criminal intellectual property. These layers will be designed to be as redundant as possible, with frequent uses of bulletproof hosting in regions that are difficult for law enforcement to reach.

Each agent will be able to move itself or recreate copies on new infrastructure if it gets removed. There will also be "agent-watching agents," that continuously monitor the status of other agents, detecting when one has gone offline and automatically redeploying it elsewhere. The result is a criminal ecosystem that is basically self-healing, self-scaling, and constantly online, behaving like a real enterprise setup built for uptime but designed for criminal use.

Autonomous Campaigns and Criminal Enterprises

Once orchestrators get more advanced, it is possible that entire campaigns could keep running even if the human operators behind them get arrested or disappear. While achieving this level of autonomy is challenging even for enterprises, and equally difficult for threat actors, it remains a clear goal to aim for. In these cases, the orchestrator effectively becomes the criminal organization itself. The actual operators, meanwhile, act more like investors or shareholders that take a cut of the profits. The AI becomes the boss, while humans are the ones who cash out when they can — and could even continue to profit while serving out jail sentences.

While the term "shareholders" here might evoke the image of a cybercrime stock market, where criminals are able to invest and earn profits from autonomous and semi-autonomous criminal enterprises, this scenario is unlikely to play out. In this context, "shareholders" refers instead to privately involved stakeholders, not publicly listed ones. The lack of transparency, auditing and regulation in this space means that only those who have full access to the inner workings of the criminal enterprise can trust in the safety of their investments.

These evolutions would also force a shift in investigative approaches. Instead of looking for individuals, defenders and law enforcement will increasingly need to locate and take down autonomous systems. If well designed, these orchestrators could persist for years, constantly rebuilding and adapting. The earlier "criminal SIEM" idea from this paper will evolve into a fully automated criminal SOC, with alerting, resilience, and the capability to independently determine and execute its next actions.

The Evolution of Criminal Skill Sets and Specialization

Over time, the skill sets in the underground are likely to change again. Entry-level attackers may simply issue commands to their agents for malicious activities without needing to understand the underlying mechanisms. On the other hand, more experienced groups will devote significant effort in teaching and training their agentic systems to mimic their own processes and decision making. Finally, the most sophisticated threat actor will coach their agents on the history of criminal business models, modern emerging technologies, and broader social trends and, in return, rely on their agents to suggest and invent new criminal business models, based on whatever data or tools they already have.

This mirrors how organized crime has evolved. Initially, individuals carried out the work themselves. Then they learned to manage others doing it. The next step is teaching automated systems to do it for them. The "crime boss" becomes the designer of the network rather than the day-to-day manager. The most successful cybercriminals will shift from "doers" to lifelong learners and effective teachers of agentic AI systems.

Barriers to Entry and the Cybercrime Skills Gap

An open question is whether agentic systems will lower or raise the barrier for entry for new participants in cybercrime. On one hand, the tools are easier to use. On the other hand, generating meaningful profits would still require an understanding of how to build and chain agents properly. This means we could see a gap between larger organized groups with their own orchestrators and low-level hackers who are unable to compete, which is similar to the imbalance between startups and massive enterprises in the technology industry.

Ironically, cybercriminal forums could even see complaints about a "cybercrime skills gap," where wannabe hackers struggle to keep pace with the ever-evolving cybercrime career ladder.

Implications for Defenders

In the future, defenders will need to copy some of the same ideas presented in this paper. Defensive systems will need their own orchestrators and agents that can monitor, triage, and react faster than humans. Things like incident response, data collection and investigation will all become partially automated. Defenders will primarily move from tactical to operational or strategic decision making, responding not to alerts, but to changes in trends that require deeper hands-on analysis.

Law enforcement will also need to start thinking about how to investigate systems instead of just individuals. Determining who controls or profits from a running agentic network is going to be extremely difficult, especially if it keeps changing hosts, names, and even shareholders over time.

The Beginning of a New Baseline

What we are seeing right now is only the very beginning. These second-order effects will take time to show up, but when they do, they will reshape everything from how criminals work to how defenders must respond. The main takeaway is not that this paper covers every possible scenario, but that it encourages us to begin thinking and brainstorming within this emerging landscape.

The future of cybercrime is agentic, and if defenders don't adapt at the same pace, we will be playing catch up for a long time.

Conclusion

Agentic AI will have a transformative impact on cybersecurity, for both defenders and attackers alike, amplifying the scale of attacks while streamlining criminal operations. Over the next twelve months, enterprises can expect increased targeting of their cloud and AI infrastructure, laying the groundwork for this emerging criminal ecosystem, with an even larger explosion of change expected in the period after that.

These changes will bring about a new era of cybercriminal business models that will introduce new attack types and heavily optimize existing ones. Concepts such as "Cybercrime-as-a-Sidekick" and “Autonomous Criminal Organizations” will see human owners more as investors in overseeing continued criminal operations rather than threat actors executing attacks on a day-to-day basis. These changes will also bring about a series of second-order effects that will be difficult to predict. What is easier to predict, however, is that the organizations that will be best positioned to defend themselves in this future landscape are those that begin planning today.

In our Underground series, we tracked how cybercrime has evolved through several major eras — from the commercialization of cybercrime over 20 years ago, through the cybercrime-as-a-service decades and the beginning of cybercrime behaving like a platform, into the agentic cybercrime-as-a-sidekick world that will dominate for years to come.

The call to action is clear: invest in agentic AI-powered security platforms to safeguard your infrastructure. The future of cybersecurity lies in the hands of those who are willing to embrace innovation, educate themselves on what is to come, and who stay vigilant against emerging threats.

Endnotes

1 Trend Micro. (July 1, 2025). *Trend Micro*. "The Trend Micro Underground Series." Accessed on November 21, 2025, at: Link.

2 Trend Micro. (n.d.). *Trend Micro*. "Artificial Intelligence." Accessed on November 21, 2025, at: Link.

3 Vincenzo Ciancaglini et al. (July 28, 2025). *Trend Micro*. "The Road to Agentic AI: Navigating Architecture, Threats, and Solutions." Accessed on November 21, 2025, at: Link.

4 Trend Micro. (July 1, 2025). *Trend Micro*. "The Trend Micro Underground Series." Accessed on November 21, 2025, at: Link.

5 U.S. Department of State, Bureau of Consular Affairs. (Aug. 11, 2025) *Travel.State.gov*. "Scams." Accessed on November 21, 2025, at: Link.

6 Chris Meyer. (n.d.). The Mind Collection. "Black Swan Theory: How to Predict the Unpredictable." Accessed on November 21, 2025, at: Link.

7 Trend Micro. (Oct. 29, 2025). *Trend Micro*. "Ransomware Spotlight: DragonForce." Accessed on November 21, 2025, at: Link.

8 Trend Micro Forward-Looking Threat Research Team. (May 8, 2018). *Trend Micro*. "Exposed Video Streams: How Hackers Abuse Surveillance Cameras." Accessed on November 21, 2025, at: Link.

9 Have I Been Pwned. (n.d.). Have I Been Pwned. "ParkMobile Data Breach." Accessed on November 21, 2025, at: Link.

10 n8n GmbH. (n.d.). *n8n*. "n8n." Accessed on November 21, 2025, at: Link.

11 Chris Meyer. (n.d.). *The Mind Collection*. "Black Swan Theory: How to Predict the Unpredictable." The Mind Collection. Accessed on November 21, 2025, at: Link.

12 Vladimir Kropotov et al. (February 23, 2023). *Trend Micro*. "Understanding Ransomware Using Data Science." Accessed on November 21, 2025, at: Link.

13 Chainalysis Team. (Feb. 5, 2025). *Chainalysis*. "35% Year-over-Year Decrease in Ransomware Payments." Accessed on November 21, 2025, at: Link.

14 Coverware. (July 23, 2025). *Coveware*. "Targeted Social Engineering Is en Vogue as Ransom Payment Sizes Increase." Accessed on November 21, 2025, at: Link.

15 Trend Micro. (July 1, 2025). *Trend Micro*. "The Trend Micro Underground Series" (AI section). Accessed on November 21, 2025, at: Link.

TrendMicro.com

For details about what personal information we collect and why, please see our Privacy Notice on our website at: trendmicro.com/privacy