transmit security

Blinded by the Agent

How AI Agents are dismantling fraud detection as we know it

Summary

AI agents – autonomous software assistants capable of performing tasks on behalf of users – are rapidly becoming a reality in digital banking and commerce. OpenAI's newly released **ChatGPT Agent** exemplifies this trend: it can navigate websites, fill out forms, and execute transactions for users via a virtual browser. Such agents promise convenience, (e.g. automatically ordering groceries or making transfers) by using the same web interfaces humans do. However, they also raise urgent questions for fraud detection. Traditional anti-fraud controls such as device fingerprinting, behavioral biometrics, and bot detection assume a human end-user. An AI agent operating in place of a consumer or a fraudster can **bypass or confuse these controls**, challenging banks' and retailers' ability to distinguish legitimate user activity from fraud.

Unlike other technologies, consumer-operated AI agents cannot be fully controlled or restricted by application owners**. As a result, financial institutions and merchants are forced to rethink their fraud controls for an environment they no longer fully govern.**

The Rise of the Agent

- The majority (over 60%) of visitors to online shops and retailers are now bots, not humans
- The number of consumers who are already comfortable with AI agents shopping and browsing websites for them is greater than 50% and increasing sharply
- Generative AI-sourced traffic to U.S. retail and banking sites grew 2000% over the last 12 months and expediting

Key Takeaways

- The fraud stack assumes human input at every stage, but agents break that assumption. Nearly every fraud detection layer relies on signals that agentic AI either disrupts or mimics, making traditional signals less useful.
- Behavioral biometrics fraud detection technologies, used by many banks, break down in environments where users delegate actions to AI agents. Interaction patterns from agents look nothing like those of a human, even when the activity is legitimate.
- Device fingerprinting technologies used by both banks and online merchants are rapidly losing reliability as a user verification signal. AI agents operate from cloud infrastructure, meaning device identity no longer maps cleanly to a known customer.
- Bot detection systems used by many online merchants are entering an identity crisis. They are increasingly forced to choose between blocking useful AI agents or letting in sophisticated fraud that leverages legitimate AI agents.

- Fraudsters are likely to shift much of their fraudulent activity to legitimate AI agents. By doing so, they effectively blind the core detection layers used by applications today.
- Fraud losses are projected to rise by up to 500% in the coming years, driven by the sharp decline in fraud detection effectiveness.
- Financial institutions and online merchants lack Predictive AI capabilities that can adapt to these new patterns in real time, leaving them unable to fight fire with fire.

OpenAI CEO Sam Altman says he is nervous about an **imminent fraud crisis,** warning that **bad actors using AI to gain access to consumer accounts is coming "very, very soon."**

July 2025

Current Fraud Detection Techniques in Banking and Retail

Modern fraud detection at banks and online merchants relies on layered defenses to verify user identity and detect imposters or bots. Some of the core technologies include:

- **Device Identification & Fingerprinting:** Banks and merchants identify devices through browser fingerprints, OS attributes, cookies, and device IDs. A user's device profile helps recognize returning customers and flag unknown devices.
    - For example, a bank might treat a login from a **new or unrecognized device** as higher risk, triggering step-up authentication (one-time passcodes or security questions). Device intelligence solutions, (e.g. Fingerprint, ThreatMetrix), collect dozens of browser and network signals to create a unique device ID, used to link sessions and assess consistency.
    - In e-commerce, device fingerprinting can spot when multiple accounts or orders originate from the same device, indicating possible fraud. Overall, device identification helps establish a trusted device history and catch anomalies (like a fraudster using a device ID tied to known fraud or suddenly switching devices).

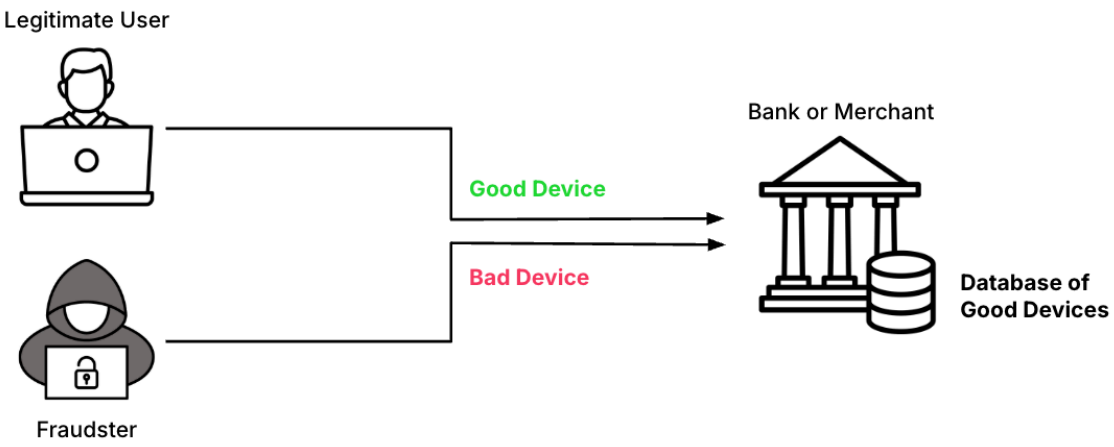

Figure 1: Device Identification and Fingerprinting

- **User Behavioral Analytics/Biometrics:** Financial institutions increasingly deploy **behavioral biometrics**, analyzing how users interact, as a passive fraud filter. This includes measuring keystroke dynamics, mouse movements, typing cadence, touchscreen pressure, and navigation habits.
    - Each user has distinctive "behavioral fingerprints" that are difficult for impostors to mimic. For example, a banking fraud system might monitor if the customer's typing speed and mouse patterns during login or transaction entry match their usual profile. Sudden deviations, (hesitation, copy-pasting password, perfectly steady cursor movements), can signal that it's not the genuine user but possibly a bot or coerced victim.
    - Behavioral analytics can also differentiate between human and non-human interaction in real time. In the **retail** sector, while full biometric profiling is less common, anomaly detection algorithms do watch for unusual user journeys, (for example, a customer rushing through checkout in seconds or making erratic page jumps).
    - Behavioral signals thus serve as an "inherence" factor – something about *how* the user behaves – to complement passwords or device IDs.

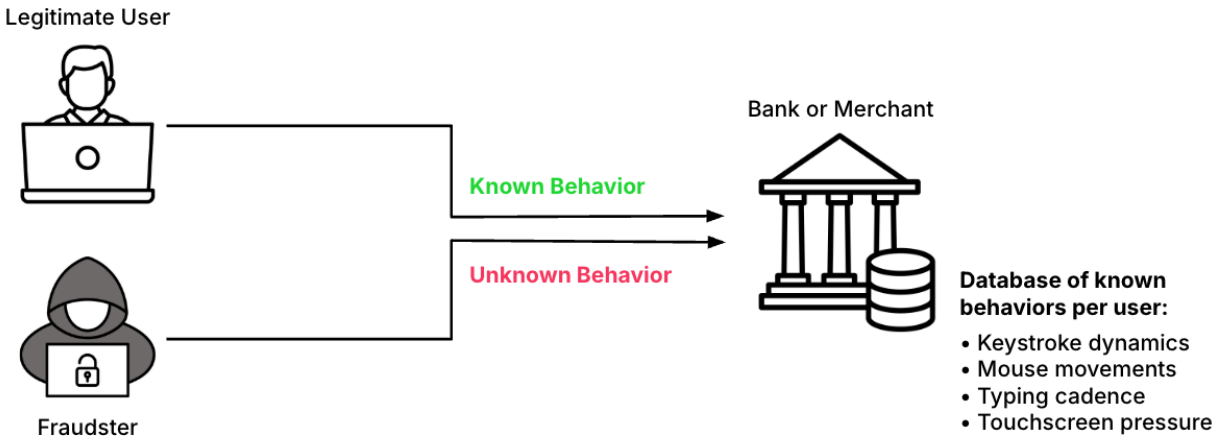

Figure 2: Behavioral Biometrics

- **Bot Detection and Automation Defenses:** Both banks and retailers employ tools to detect **non-human traffic**, since bots are often behind fraud attacks, (credential stuffing, card testing, fake account creation, etc.)
  - Anti-bot systems, (like reCAPTCHA, Akamai, Cloudflare, etc.), look for telltale signs of automation: missing or inconsistent browser properties, script-like navigation, (very rapid clicks or form fills), absence of mouse movement, or known bot user-agent strings.
  - Challenges like CAPTCHAs are deployed to distinguish humans from bots by forcing tasks that machines traditionally struggle with.
  - Advanced fraud prevention solutions use **device and network signals** to identify automation; for instance, flagging known headless browser frameworks or identifying traffic from **virtual machines and proxies** often used by botnets.
  - The goal is to **block or verify automated sessions** before they can abuse systems, without obstructing legitimate users. In summary, bot detection creates a baseline expectation of human-like randomness – any session too "machine-like" risks getting flagged.

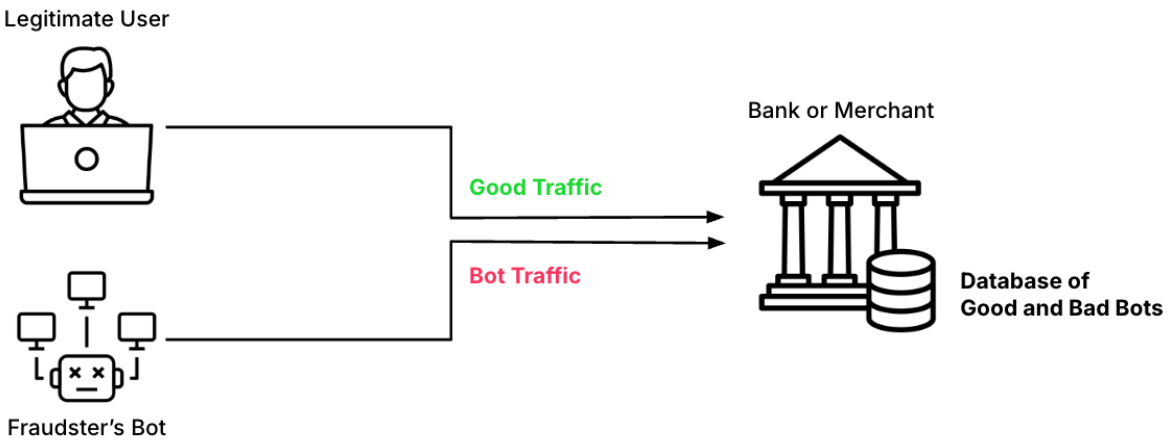

Figure 3: Bot Detection

- **Anomaly & Transaction Monitoring:** Beyond front-end device/behavioral checks, banks rely on back-end analytics to spot suspicious patterns in account activity or purchases. Rules might cover velocity, (e.g. too many transactions in a short time), amount anomalies, new payees, or geolocation changes.
    - For example, if a credit card suddenly makes dozens of purchases at different merchants within minutes, the system will likely decline some transactions or alert fraud teams.
    - Similarly, if an online banking session initiates a large transfer to a new beneficiary at 3 AM from an IP in another country, it may trigger a security challenge. These systems learn typical customer behavior and raise flags when out-of-pattern events occur.
- **User and Entity Behavior Analytics (UEBA)** tools extend this by profiling normal usage (login times, transaction habits) and detecting deviations that could indicate account takeover.
    - In retail e-commerce, fraud engines score orders based on multiple data points (device, IP reputation, past history, item being bought, etc.) and may auto-reject or hold orders that look high-risk (for instance, mismatched shipping address, or buying high-value electronics in bulk). The common thread is an assumption of **human limitations** – for example, a human shopper cannot realistically place dozens of orders across many sites in one minute, so such behavior appears inherently suspicious.

These technologies collectively create a layered defense. **However, AI agents acting on behalf of consumers fundamentally alter the assumptions behind these controls.** Below, we analyze how each fraud detection mechanism is impacted when the "user" is an AI agent rather than a human.

Impact of AI Agents on Device Identification

AI agents like ChatGPT Agent do not run on the consumer's personal device; instead, they operate from cloud servers or virtual machines (a "remote browser" controlled by the AI). This decoupling of user and device poses a challenge for device identification systems:

- **Unknown Device & Fingerprint Mismatch:** When a fraud system sees a login or transaction request initiated by an AI agent, it will typically register as a **new device environment** – different IP address (often a data-center IP), different device fingerprint, and lacking the cookies or tokens that mark a returning customer.
    - From the bank's or retailer's perspective, it's as if the customer suddenly switched to a brand-new computer in an unusual location. This often triggers high risk scores and additional verification.
    - For example, a user instructing an AI agent to log into their bank might encounter out-of-band authentication (one-time codes) because the device is unrecognized. Frequent use of agents could thus lead to **continual "new device" flags**, adding friction for legitimate users.
- **Bypassing Trusted Device Intel:** Many applications leverage "remembered" device profiles to streamline legit customers' access, (avoiding constant OTP challenges for a known device).
    - AI agents can **bypass this trusted device mechanism** simply by not being that known device. Unless the agent somehow inherits a user's device fingerprint, it appears as distinct every time.
    - In effect, device-based location and ID become less useful signals for confirming identity. Fraudsters might exploit this by using AI agents to mask their true device identity – spawning fresh virtual browsers per attempt, they can evade device fingerprint blacklists or reputation systems. Indeed, **fraud toolkits can randomize fingerprints** to appear as new devices continually, a tactic now potentially supercharged by AI automation.
- **Concentration of Agent Traffic:** On the other hand, if many consumers start delegating tasks to a popular AI agent service, this could cause **clusters of traffic from that service's infrastructure**.
    - For example, dozens of customers may inadvertently log in via the same pool of cloud IP addresses or with identical technical fingerprints of the operator's browser.
    - This anomaly might resemble a botnet or aggregator service to fraud systems. It could lead to legitimate agent-driven sessions being blocked or challenged en masse due to risk rules (for instance, "multiple customer logins from same IP range" alerts). *Case in point:* early financial aggregators (like screen-scraping fintech apps) faced this issue – banks saw a single server IP hitting many accounts and often treated it as an attack.

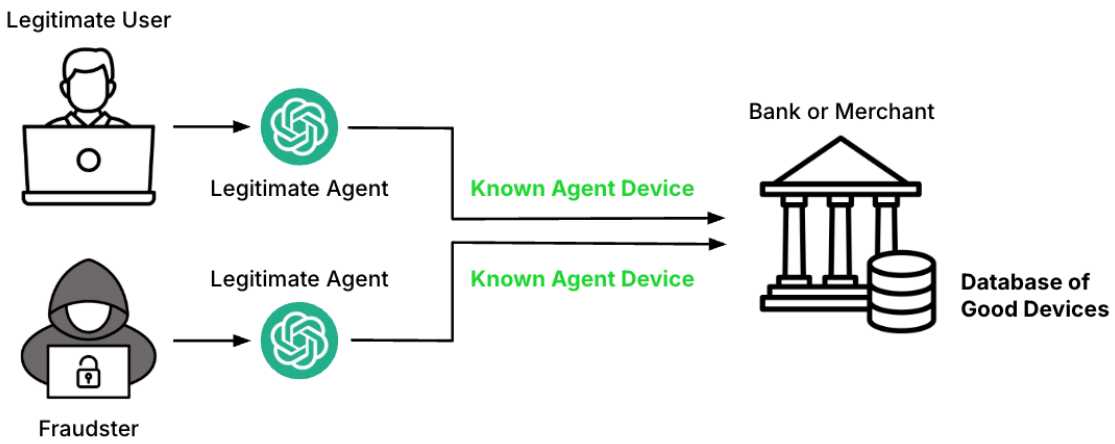

Figure 4: Impact of AI Agents on Device Identification

Bottom line: Consumer AI agents erode the reliability of device identity as an authentication factor. Banks and retailers can no longer assume that a consistent device equals a human customer, or that a new device is malicious – it might be an authorized AI helper. Heavy reliance on device fingerprinting will either **generate more false alarms,** (if every agent session looks like a "new device login"), or could be skirted by attackers rotating through AI-generated device environments.

Impact of AI Agents on Behavioral Biometrics

AI agents executing transactions on behalf of users also upset the expectations of user behavior analytics. Fraud detection systems that rely on the subtle cues of human interaction may find those signals either missing or artificially reproduced:

- **Loss of Human Interaction Signals:** By design, an autonomous agent does not exhibit the natural hesitations, errors, and variability of a human user. Tasks that might take a person minutes of clicking and typing, an AI agent might complete in seconds with machine precision.
    - **Continuous behavioral monitoring** in banking, for example, measuring how a customer types their password or navigates between fields – could be effectively blinded during an agent-driven session. If an AI agent autofills a form almost instantly or navigates perfectly, the behavioral biometrics module may register anomalies: **too steady or too fast to be human**.
    - For example, no human can click through a multi-page flow in one second intervals with zero cursor deviation, or type a 16-digit card number without pausing – but an agent can. A bank's system might interpret that as a likely bot (and ordinarily, it would be right).

- Behavioral biometrics firms note that the arrival of agentic AI will increase the difficulty of distinguishing human vs. non-human users – especially if one can no longer rely on behavioral "tells" as before. In other words, when a **legitimate** user *intentionally* employs a bot, the lack of human biometrics doesn't necessarily equate to fraud, making traditional behavioral flags less definitive.

- **Mismatch with User's Normal Profile:** Even if some behavior is present, it likely won't match the genuine customer's habitual pattern. Each user has idiosyncratic rhythms (e.g. how they toggle between input fields or their typing speed).
  - An AI agent acting for the user will have a **different interaction signature** – or might even simulate a generic "average" user behavior. Behavioral fraud systems that perform user-specific profiling would then flag these sessions as high-risk because **deviations are detected**.
  - For instance, a retail bank's behavioral engine might detect that the current login, while successful with the password, shows none of the user's typical typing cadence and uses mouse movements that are completely uncharacteristic. This could trigger an alert or a step-up challenge (like asking additional security questions).
  - Thus, even authorized agent use can resemble account takeover from the viewpoint of behavioral biometrics. Banks may face an increase in false positives if they don't adjust the models or explicitly account for agent-driven interactions.

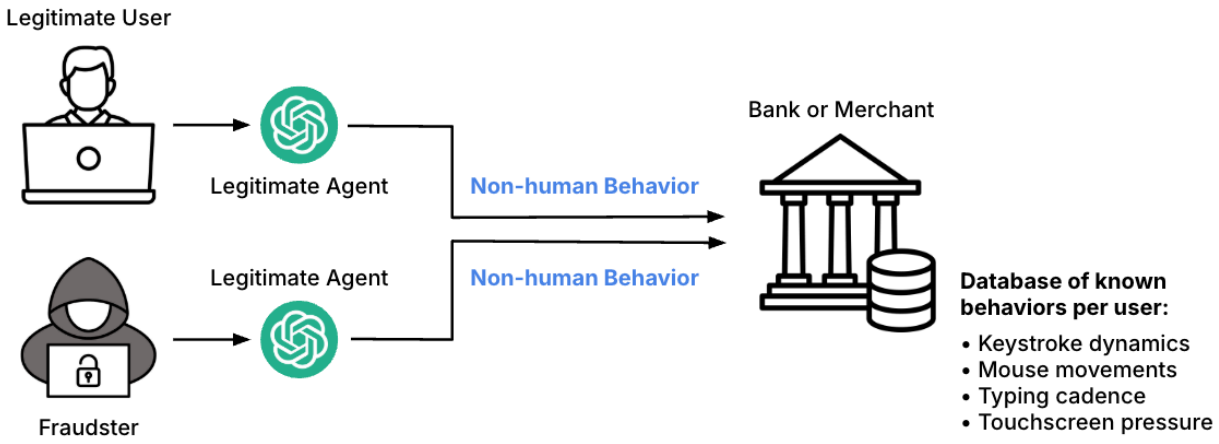

Figure 5: Impact of AI Agents on Behavioral Biometrics

- **AI Mimicking Human Behavior:** On the flip side, as AI advances, agents and fraud bots may attempt to **mimic human behavior** to evade detection. Modern AI agents could be instructed to insert random delays, move the mouse cursor in human-like ways, or even generate synthetic keystroke patterns.
  - In fact, some AI systems can **mimic human behavior with high precision**, potentially fooling simplistic behavioral checks. A well-designed agent might purposely introduce a bit of "noise" –

pausing briefly as a human would, scrolling the page in a non-linear fashion, or typing at a plausible speed – all to avoid looking robotic.

  - This cat-and-mouse game means behavioral biometrics solutions must look for subtler inconsistencies. For example, even if an agent tries to behave like a human, it might be *too consistent* or lack the genuine cognitive "micro-errors" that real users have (like hesitating before an unfamiliar step, or correcting a typo).

- **No Biometric Feedback Loop:** Another impact is on **continuous authentication**. Some banks deploy systems that continuously verify the user in the background (for instance, monitoring behavior during a session and logging out or re-challenging if the behavior no longer matches the authenticated user).

  - If an AI agent takes over after the user logs in, the behavioral change could prompt the system to terminate the session or require re-authentication.

  - Likewise, if the agent starts the session from scratch, continuous authorization can't even initialize with a known human baseline. In summary, these systems either become ineffective, (if tuned down to accommodate agents), or overly sensitive, (if left unadjusted, they will constantly intervene during agent use).

In summary, consumer AI agents undermine the reliability of behavioral biometrics as a fraud signal. Legitimate agent use can appear indistinguishable from a fraudster in terms of interaction patterns, leading to **false alarms and user friction**. Conversely, malicious actors can use AI to better conceal their non-human behavior, leading to **missed detection**. Fraud teams may need to develop new behavioral indicators specifically for AI-driven sessions. Until agent-specific fraud detection capabilities mature, banks and retailers must be prepared for **either relaxing some behavioral rules** when an agent is suspected (to avoid false declines) or conversely, treating any absence of human behavior as high risk and then **requiring alternate verification** to ensure the agent is legitimately authorized by the user.

Impact of AI Agents on Bot Detection Measures

By their very nature, AI agents are bots, albeit *authorized* bots acting for users. This blurs the line for traditional bot detection and anti-automation defenses, which were built to **keep out all bots**. The rise of consumer AI agents raises key issues:

- **Legitimate Bots Triggering Bot Defenses:** An AI agent using a retailer's website or a bank's online banking will likely be flagged by existing bot detection systems. These systems can detect common automation frameworks, and indeed many are already being updated to recognize **"AI agent" toolkits like OpenAI's ChatGPT Agent** via distinct technical signals.

- For example, the agent's browser might present a unique user-agent string or fail certain browser integrity tests that real user browsers pass. The immediate result is that the AI agent may be challenged with a CAPTCHA or blocked outright as a suspected bot.
- OpenAI's ChatGPT Agent itself acknowledges this. It is designed to **hand control back to the user for CAPTCHAs and logins**, precisely because these hurdles are meant to stop automated access.
- While this is a security positive (the agent isn't surreptitiously bypassing CAPTCHAs without user involvement), it also illustrates the **usability friction**: a user hoping for seamless automation might instead be interrupted by frequent "Are you human?" tests because the website can't tell a friendly agent apart from a malicious script.
- If banks and merchants do nothing, the default outcome is **legitimate AI agents will be treated as malicious bots** by their security filters – undermining the utility of agents (they get stuck at the front door) and annoying customers who then have to intervene.

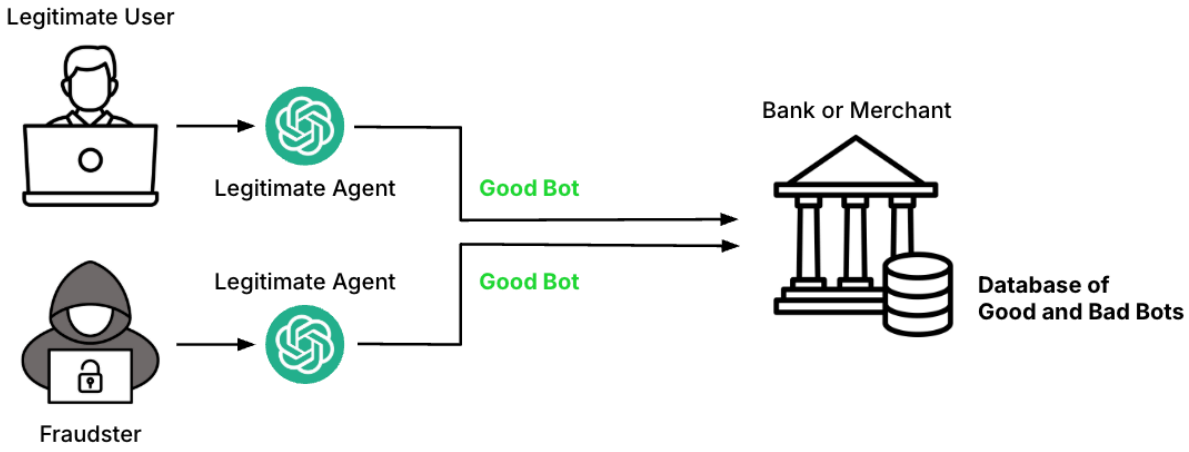

Figure 6: Impact of AI Agents on Bot Detection

- **Malicious Bots Becoming More Human-like:** While good bots face hurdles, bad bots get smarter. Fraudsters can integrate AI capabilities into their botnets to better evade detection. For instance, AI vision models can solve CAPTCHA's or even generate deepfake clicks that fool visual CAPTCHA checks.
- **Advanced AI can analyze and predict solutions to security challenges,** (like guessing the answers to common knowledge-based authentication, or defeating basic facial recognition checks with deepfakes). This means some anti-bot challenges may lose effectiveness over time as AI-augmented bots solve them at scale.

- Similarly, an AI-driven bot can dynamically adjust its behavior each run to avoid patterns, one of the strengths of AI is adaptability and "creativity," which can be used to randomize attack patterns.

- Traditional bots might hammer a login form with identical timing and payloads, but an AI-enhanced bot could **vary its approach each attempt**, making detection by static rules harder. For example, one attempt might slowly type the password, another might paste it, one might navigate via keyboard, another via mouse – all orchestrated by an AI to appear as diverse as real users.

- In short, AI allows malicious automation to better impersonate human web browsing, which directly targets the effectiveness of bot detection algorithms that rely on predictable bot signatures.

- **Mass-Scale Automation (Superhuman Speed):** AI agents can operate at speeds and scales far beyond human ability. A single agent could open numerous browser tabs or sequentially perform actions 24/7 without fatigue.

  - If a user (or attacker) employs an agent to perform tasks in bulk, it may trip rate limiting and anti-bot thresholds. For example, a shopping agent might attempt to purchase limited stock items the moment they drop – making requests in milliseconds. Websites might interpret this as a bot (it is) and block the activity to give human customers a fair chance. But if that agent was legitimately deployed by a customer, from their perspective the **fraud system becomes an obstacle to normal service**.

    > Historically, essentially all non-human traffic was unwelcome on secure banking or shopping platforms

  - Similarly in banking, if an agent tries password attempts too quickly (perhaps as part of a password manager function), it might get the account locked out for "suspicious activity."

  - Thus, existing rate controls and bot filters don't distinguish motive – they simply react to the behavior. **Agentic activity can resemble attack traffic** if purely judged on volume or speed.

- **Distinguishing Good Agents from Bad Bots:** The critical need is developing ways to **differentiate "beneficial automation" from malicious automation**. This is a new paradigm: historically, essentially all non-human traffic was unwelcome on secure banking or shopping platforms.

  - Now, companies must accommodate *user-approved* agents (for convenience) while still repelling criminal bots. Some measures being explored include: requiring AI agents to **identify themselves** (perhaps via a specific HTTP header or token), so that a site knows this is Agent X acting for User Y. With identification, a policy decision can be made: e.g., allow this known agent but perhaps sandbox its actions or limit its rate.

- In the absence of explicit identifiers, detection has to rely on behavior and fingerprint. As noted, device intelligence systems now offer **bot/AI agent classification signals** that perform "intelligent classification on each request to determine whether a bot or agent is **legitimate or malicious**".

- They achieve this by cross-referencing with databases of known **automation frameworks and AI assistants** and by analyzing environment details (for example, detecting if the session is running in a **virtual machine** or using a **residential proxy** favored by fraudsters).

- If an agent is recognized as one of the "verified good" (say, an AI helper from a payment provider), it might be allowed to proceed with fewer hurdles. On the other hand, if it's an unknown automation or matches patterns seen in abuse, it can be blocked or flagged.

- **Policy and Liability Changes:** On a broader scale, the entry of agents is forcing a rethink of **security protocols and liability** for automated transactions.

  - For instance, who is liable if an agent makes an unauthorized purchase – the user, the agent provider, or the merchant? Some payment networks (Visa, MasterCard) are considering frameworks where **only designated (trusted) AI agents are allowed to transact on a user's behalf**, with new rules and protections in place.

  - This suggests merchants and issuers may eventually get a feed of information indicating the presence of an agent, and could reject transactions from unrecognized agents.

  - We are in the early stages, but companies like Visa have explicitly stated that merchants *"need to know when a transaction is initiated by an agent"* and follow extra protocols in those cases.

  - Those protocols might include stronger authentication or different fraud checks. In effect, agent-driven transactions could become a separate category in fraud systems, with their own validation flows.

To summarize, **the rise of AI agents breaks the traditional "block all bots" model**. Enterprises are now beginning to **intentionally allow legitimate bots,** such as ChatGPT Agents, to serve their customers. But this shift opens the door to a dangerous loophole: **fraudsters can exploit the same trusted agents by feeding them stolen credentials or malicious instructions**. From the application's point of view, the bot appears perfectly legitimate, **but the intent behind its actions is fraudulent**. This means that once trusted AI agents are whitelisted, **bot detection systems are effectively blinded**. They can no longer tell if the agent is acting on behalf of a real customer or a fraudster. The very tools designed to block malicious automation become useless when **the automation itself is approved**. The challenge now isn't detecting automation—it's detecting intent. And **fraud prevention teams need to shift their mindset**: from blocking bots to **understanding behavior and predicting outcomes**, even in trusted, machine-driven environments.

Impact on Fraud Rules and Anomaly Detection

AI agents also challenge fraud detection rules and anomaly detection algorithms in both banking and retail, as these often assume human limitations and behavior patterns. Rule-based and anomaly detection systems lack context for "why" a pattern is occurring; they only see the outcome.

Some notable impacts include:

- **False Declines and Customer Friction:** Merchants and banks worry about **false declines**, legitimate purchases blocked by fraud systems, because they directly translate to lost sales and upset customers.
    - Agent-based transactions raise the risk of false declines if the fraud systems aren't updated as **agents don't 'look' like a trusted customer.** The agent might not carry forward the usual identifiers of the customer (device, behavior, location), causing the fraud model to misjudge the identity or trust level. An AI agent can perform dozens of actions across multiple platforms nearly simultaneously – something a single human would never attempt. This creates **transaction patterns that traditional fraud models deem "impossible" or highly suspicious.**

      For example, consider a scenario in which an AI agent is given a long shopping list and tasked with making the purchase while optimizing for price, customer feedback, and delivery time. The agent might then split this task into 50 separate orders at 11 different online vendors, and complete all those checkouts **in 90 seconds**.

      Each individual transaction might be fine, (legitimate merchant, within expected amount range), but the *aggregate pattern* – rapid-fire orders with no delay, is extremely abnormal for one account. Under conventional rules (velocity checks, multi-merchant burst flags), this legitimate use of an agent would almost certainly be throttled. In banking, similarly, if an AI personal finance agent initiated **a series of fund transfers or bill payments in quick succession**, it might trip anti-fraud systems that assume malware is scripting the account.
    - Retailers could see their **conversion rates drop** if, say, an AI shopping assistant attempts to buy on behalf of users and gets blocked frequently.
    - This in turn could slow adoption of agent-based commerce unless addressed – consumers won't use AI assistants that constantly trigger declines or require phone calls to the bank to clear things up.
- **Disabling Fraud Rules**: A growing risk is the fraudster's ability to **shift activity to trusted agents like ChatGPT Agent**, bypassing anomaly detection systems. Many fraud rules are designed to trigger when users behave abnormally such as making unusual purchases, logging in at odd hours, or accessing unfamiliar locations. However, **some organizations are beginning to treat agent activity as inherently "different" and exclude it from strict behavioral rules to avoid false positives**. This

opens the door for abuse: **fraudsters can operate through trusted agents, performing actions that would normally appear suspicious for a human user but now go unchecked because the origin is an AI agent**. In effect, anomaly detection systems are taught to "look away" when the activity comes from an agent, giving fraudsters a safe channel to operate in plain sight.

- **New Fraud Vectors via Agents:** On the flip side, fraudsters might use agents to **amplify certain abuses**. For example, an AI agent could be instructed to find and exploit every available promo code or loyalty offer across a series of merchants, something a human scammer would do manually in a limited way.
    - This could lead to **"AI-aided abuse"** like automated returns or promotion stacking at a scale that erodes merchant margins.
    - Fraud systems tuned for traditional abuse patterns might not catch a cleverly orchestrated agent that, say, orchestrates dozens of small refunds that fly under thresholds, or coordinates multi-platform arbitrage.
    - Additionally, agents might enable **scaling of synthetic identity fraud** – an AI can open many fake accounts and manage them concurrently, each behaving a bit like a real customer (making small transactions, building a "trust" pattern until a big cash-out). This could defeat rules that rely on seeing obvious red flags (because the agent-run fake accounts behave normally most of the time). **AI-generated synthetic identities** can establish convincing normal behavior patterns that evade detection.
- **Loss of Visibility and Audit Trail:** When interactions become AI-to-AI (agent to merchant system), some traditional signals are lost. Merchants might get less granular data on the user's actions leading to a purchase because the agent abstracts it away. **As interactions shift to machine-driven processes, they become opaque and rapid, making visibility increasingly difficult.** This makes it harder for risk teams or investigators to reconstruct events. If a dispute or fraud case arises from an agent transaction, deciphering what actually happened (which steps the agent took, whether the user authorized each step) can be challenging without proper logging.

  In banking, this could complicate things like **chargeback disputes or fraud claims** – a customer might say "I didn't do this, the AI agent did it on its own," shifting how liability is analyzed. As agents introduce an intermediary layer, fraud systems may need more sophisticated logging and analytics to maintain end-to-end visibility (e.g., knowing the chain: User X → Agent Y → Performed Action Z).

Predictions and Conclusion

> “ AI has “fully defeated” most of the ways that people authenticate who they are. ”
>
> Sam Altman, July 2025

Fraudsters are likely to shift much of their fraudulent activity to legitimate AI agents. By doing so, they effectively **blind the core detection layers** used by most applications today: device identification, behavioral biometrics, velocity rules, bot detection, and even session anomaly detection. These controls were built on the assumption of human interaction, and AI agents break that assumption completely. As a result, **most of the fraud currently caught by these layers will bypass detection** as attackers adopt agent-based automation. Application owners will be forced into a difficult choice: either **aggressively challenge their legitimate users**, introducing friction and abandonment, or **accept up to five times more fraud,** (assuming their existing controls were blocking at least ~80% of attacks). Without new detection models purpose-built for agentic activity, **the fraud prevention stack risks becoming obsolete.**

Without the right technology to understand and respond to AI agent traffic, **fraud teams will be overwhelmed**. As AI agents become a primary interface for both legitimate users and fraudsters, traditional detection systems will generate massive volumes of alerts, many of them false positives. Fraud analysts will be forced into a reactive mode: **chasing new rules, investigating unclear signals, and responding to surging customer complaints**. We estimate fraud teams will face **2–3 times more operational workload** over the next 12–18 months, just to maintain current levels of protection. The cost isn’t just internal strain, it’s customer dissatisfaction, missed threats, and an overall decline in fraud detection effectiveness.

> **Financial-services industry fraud losses in the U.S. could reach $40 billion by 2027, up from $12.3 billion in 2023, because of the impact of GenAI.**
>
> Deloitte’s Center for Financial Services

The solution lies in more advanced, predictive AI fraud detection, (fighting fire with fire), that can adapt to these new patterns in real-time, tied to strong authentication capabilities. An agent-aware, well trained predictive model could recognize that, say, all 50 rapid transactions ultimately tie back to an authenticated, trusted corporate account, reasonable aggregated activity, and an approved agent, and thus approve them despite the unusual pattern, (perhaps with a single consolidated risk check). In contrast, a fraudster using an agent might have subtle inconsistencies in identity linkage that the model can catch.

It’s important to note that we are still in the **early days** of agentic AI in consumer applications. However, agentic AI technologies progress extremely fast and the adoption of AI agents is expected to skyrocket sooner than we think. The **absence of a long history of incidents** is precisely why fraud professionals need to be proactive – **by the time agent-driven fraud patterns are common, it may be too late to retrofit systems.**

Evidence

AI Browsers & Autonomous Agents

AI companies are rapidly advancing web-native autonomous agents and browsers.

1. **Radware Finds 57% of Online Shopping Traffic Now Bots, Not Buyers**
   New 2025 E-commerce Bot Threat Report details rise in bot attacks, emerging threat vectors, and shifting defense strategies
   https://www.globenewswire.com/news-release/2025/04/23/3066593/8980/en/Radware-Finds-57-of-Online-Shopping-Traffic-Now-Bots-Not-Buyers.html

2. **OpenAI's Sam Altman warns of AI voice fraud crisis in banking**
   https://apnews.com/article/openai-ceo-sam-altman-fed-ad87262a4c1e71a0695ff6d06a2586f2

3. **OpenAI** launched *ChatGPT Agent* and *Operator*, tools enabling AI to perform tasks inside virtual environments and the live web
   https://openai.com/index/introducing-chatgpt-agent/ https://openai.com/index/introducing-operator/

4. **Perplexity AI** unveiled *Comet*, an AI-native browser designed to automate and enhance web experiences
   https://indianexpress.com/article/technology/artificial-intelligence/can-comet-replace-google-chrome-perplexity-ai-browser-closer-look-10140421/

Industry Sentiment & Readiness

5. **96% of IT professionals view AI agents as a growing security threat**, even as 98% plan to expand their use within the next year.
   pub1.ey.com+20TechRadar+20Axios+20

6. A recent Accenture survey revealed that **80% of 600 bank cybersecurity leaders believe generative AI equips hackers at a faster pace than financial institutions can defend**.
   https://www.businessinsider.com/sam-altman-federal-reserve-financial-institutions-voice-prompt-authentication-2025-7?utm_source=chatgpt.com

7. **90% of Americans fear the rise in fraud and the impact of AI on successful fraud attempts**, according to a survey by Abrigo.
   https://www.abrigo.com/news/abrigo-fraud-fear-survey/?utm_source=chatgpt.com

Technology Gaps & Legacy Challenges

8. Despite increased investment in headcount, **nearly half of financial institutions report a lack of adequate resources and technology to fight financial crime**.

https://financialit.net/news/compliance/nasdaq-verafin-announces-launch-its-agentic-ai-workforce-delivering-step-change-aml?utm_source=chatgpt.com

9. A study by EY indicates that **institutions mostly rely on traditional AI/machine learning and enhanced legacy controls with limited or early adoption of GenAI-augmented tools**, which falls short against sophisticated AI-driven fraud attacks.
https://pub1.ey.com/content/dam/eysitesprogram/ey-csg/c1/documents/thought-leadership/EY-Forward-thinking-GenAI-and-Agentic-AI-Approach.pdf?utm_source=chatgpt.com

Emerging Strategies & Adoption

10. By 2025, **25% of companies using generative AI will have launched agentic AI pilots**, with projections reaching 50% by 2027.
https://www.linkedin.com/pulse/from-sops-smart-ops-agentic-ai-revolution-fraud-disputes-sharma-w3ntc?utm_source=chatgpt.com

11. OpenAI CEO Sam Altman has warned that the use of voice authentication in banking is particularly troubling, as AI now has the ability to mimic human voices with near-perfect accuracy, predicting a significant fraud crisis stemming from this capability.
https://apnews.com/article/ad87262a4c1e71a0695ff6d06a2586f2?utm_source=chatgpt.com

About Transmit Security

Transmit Security delivers future-proof customer identity experiences in a world where AI is accelerating change across both fraud and user access. We do this by fusing customer identity, fraud prevention, and identity verification into a single, unified system, eliminating silos and enabling rapid adaptation.

At the core is our Predictive AI, continuously learning from real-time signals to detect intent, uncover emerging fraud patterns, and make accurate decisions before damage is done. This fusion-first approach allows leading enterprises to stay ahead of evolving threats while delivering seamless, secure experiences to their customers.

Request a demo | Transmit Security

CONTACT US