# Let's get real about Agentic Commerce

## A practical Guide to

## Adoption, Fraud Risks, and the future

TrustML Studio

AI Focused Risk Consulting

# Introduction

Agentic commerce refers to AI-powered agents acting on behalf of users to shop and transact online autonomously[1]. Instead of a human manually browsing and checking out, an AI agent can search for products, compare options, and even complete purchases based on user-defined goals. This emerging paradigm promises unparalleled convenience – imagine getting what you need **without even visiting the store's website**[2]. However, it also introduces **fundamental shifts in risk and control** for merchants and platforms.

When AI "concierges" handle transactions, traditional fraud-prevention cues and user journey touchpoints are bypassed[3]. Merchants lose direct visibility into checkout flows and must trust that the agent is acting in the customer's interest and within safe parameters.

Meanwhile, fraudsters are equally eager to exploit these autonomous channels, whether by manipulating AI agents or deploying malicious bots that masquerade as legitimate shoppers[4]. **But lets be real,** this is scammer 101, and bit driven fraud has been a staple for a long time.

# Adoption Trends of Agentic Commerce

# Rapid Emergence but Nascent Adoption

Within the past year, agentic commerce has moved from concept to early reality. By early 2024, about **10% of U.S. consumers had used an AI tool to purchase a product** – typically for small, routine items. Trust remains a barrier (only 24% felt comfortable letting AI complete purchases), but importantly **64% of consumers are open to trying it** and nearly three-quarters would use AI for product research or price comparison [5][6].

This suggests a large pool of "fence sitters" who could adopt agentic shopping as confidence grows.

Industry projections underscore the anticipated surge: Gartner forecasts that **33% of e-commerce enterprises will integrate agentic AI by 2028 (up from <1% in 2024)**[7], and one analysis projects AI agents driving **$1.7 trillion in e-commerce transactions by 2030**[8].

In other words, while only a minority of transactions today, autonomous shopping agents are expected to become a mainstream channel within a few short years.

# Early Use Cases – Groceries & Household Items

A Bain & Co. survey found most people who have tried AI buying did so for groceries and everyday goods[9]. This makes intuitive sense: having an AI assistant restock toilet paper or reorder milk is convenient and carries little downside if the wrong brand arrives. Other retail categories likely to see early agentic traction include **replenishable apparel** (basic t-shirts, socks, workout clothes) and **consumer electronics accessories**, where specs are straightforward[10][11].

Current adoption skews toward **low-risk, repeat purchases.**

**Travel bookings** may also be an early adopter segment – think frequent flyers letting an agent auto-book flights within set price and schedule parameters[12]. In contrast, **high-consideration purchases** (expensive electronics, fashion for special occasions, furniture) will lag until trust in AI decision-making improves[11][13]. Geographically, North America is leading with Big Tech-driven initiatives, while Europe is cautiously exploring agentic commerce amid stricter regulations (e.g. strong customer authentication requirements, discussed later)

**Notable Developments (Last 6–9 Months):** Several announcements illustrate the momentum in late 2024 and 2025:

- **Amazon's "Buy for Me" Beta (April 2025):** Amazon began testing an AI shopping assistant that can purchase from **third-party retailers entirely within the Amazon app**[14]. The agent fills the cart, applies payment, and checks out on the user's behalf. This offers a glimpse of a future where retailers' own websites might not even be visited – the transaction becomes an "off-stage" process orchestrated by a platform agent[15]. Notably, Amazon's approach keeps the user experience within its ecosystem, potentially improving trust but also **"writing the retailer out" of their direct customer interaction**[16].
- **Perplexity.ai and Search Integrations (2024):** Perplexity, a conversational search AI, integrated one-click shopping in search results[17], allowing users to move from query to purchase in a seamless, AI-guided flow. Likewise, Microsoft's Bing Chat and Google's Bard have experimented with shopping assistant capabilities, blurring lines between search, recommendation, and transaction.
- **Agent Platforms & API Access:** Several startups and initiatives (e.g. Salesforce's AI shopping agent tools[18], Agent Commerce Kit (ACK) standards[19]) are emerging to connect agents with merchant systems. Payment providers are also jumping in – **Visa and Mastercard** have outlined visions for agent-mediated payments (Visa's "Agentic Commerce" guidance[20], and Mastercard's crypto-token approach discussed below).
- **Consumer Attitudes Shifting:** Consumers are warming to the idea of AI assistance in shopping. By late 2024, **71% of consumers globally expressed interest in AI agents for customer service or product Q&A** (early steps in the shopping journey)[21]. This comfort with AI "front-ends" may naturally extend to the checkout stage as positive experiences accumulate.

# Momentum vs. Resistance

The **North American market** (led by U.S. Big Tech and retail giants) is driving many agentic commerce pilots. In these regions, convenience culture and platform ecosystems (Amazon, Google, Apple) provide fertile ground for adoption.

**Europe**, on the other hand, has intense focus on data privacy, security, and regulated payments – which creates *both* friction and potential trust advantages. EU consumers are used to authentication challenges (e.g. widespread use of two-factor payment approval under PSD2), which could actually make them more amenable to agent-mediated payments *if* those agents comply with security standards[22].

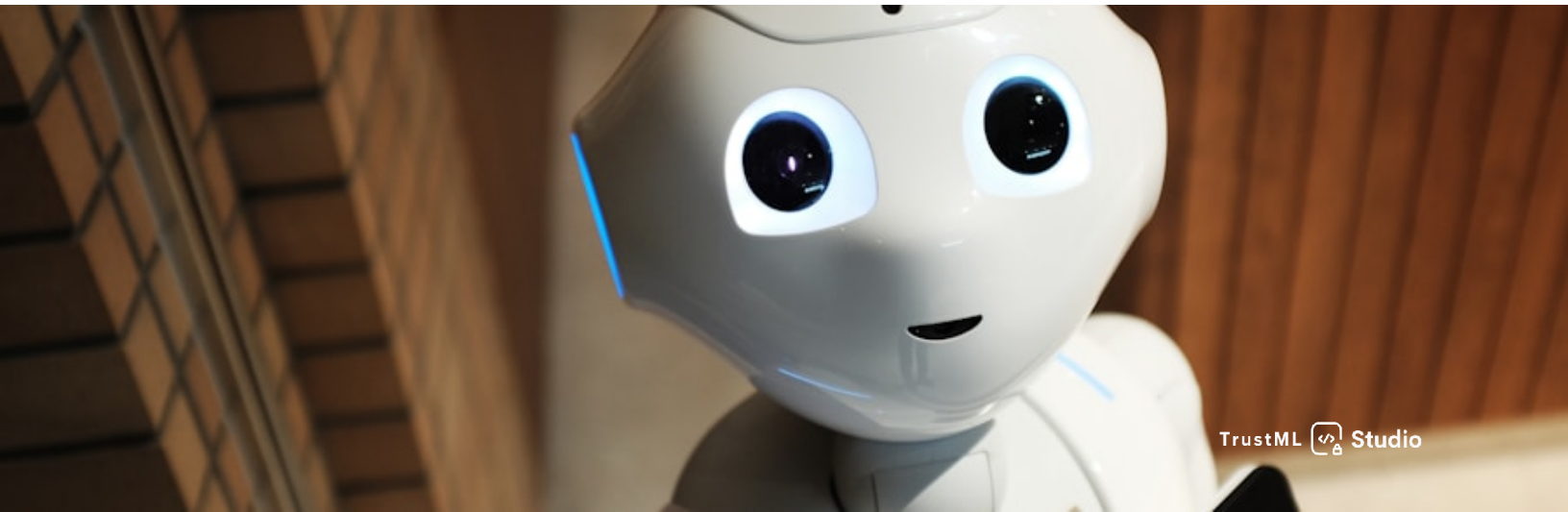
That said, European regulators are scrutinizing AI in consumer contexts; clarity on liability and transparency will be prerequisites for broad uptake in the EU. Overall, the last 6–9 months show **rapid innovation** and growing consumer curiosity, but also highlight that **trust is the linchpin** – adoption will only accelerate if these agent systems can prove they are secure and reliable in handling people's money.

# Anti-Bot Technologies: Impact on Users and Revenues

Online businesses have long battled “bots” that engage in undesirable behaviors (scraping data, snapping up limited inventory, credential stuffing, credit card fraud, etc.). In response, a mature **bot-detection and blocking industry** has evolved (reCAPTCHA challenges, device fingerprinting, web application firewalls, bot management services like Cloudflare, Akamai, HUMAN Security, etc.).

The advent of **legitimate** shopping agents forces a re-think of these defenses: **How do you welcome the “good bots” (user-authorized agents) while still repelling malicious bots?** This is a delicate balance.

Current anti-bot solutions, if left unchanged, could end up treating all automation as evil – causing **friction for genuine customers** and lost revenue. Let's get real and explore the current state and the trade-offs.



# Efficacy vs. User Friction

Traditional bot filters have never been perfect, and they often **snare real users in their net**. For example, visible CAPTCHAs are notorious for degrading the user experience – studies find CAPTCHAs can reduce conversion rates by around **3–5% on average**[23]. On mobile devices especially, solving a CAPTCHA (squinting at blurry text or clicking tiny checkboxes) adds significant friction, leading a portion of users to simply abandon their cart.

Even a modest 3% conversion drop can equate to thousands of lost sales on a high-traffic site[24]. More damning, a Stanford University study found form completions can plummet by up to 40% when a CAPTCHA is present[25]. This illustrates the **revenue trade-off of aggressive bot gating**: every suspicious traffic blocked or challenged could include some frustrated would-be customers who take their business elsewhere.

# Beyond CAPTCHAs, other bot mitigation tactics can also backfire.

False declines are an expensive, under-reported problem: in 2023, U.S. ecommerce merchants lost an estimated **$157 billion to false declines** – far exceeding the $48 billion lost to actual card fraud[[26]] Globally, false declines may have cost **over $443 billion** in 2023[26]. These staggering figures include transactions blocked by rule-based filters or manual review errors. Each false decline is not only a **sale lost** but potentially a **customer lost for good** – 41% of consumers say they will never shop with a brand again after being wrongly declined[27].

**Device fingerprinting** or IP reputation rules might misidentify a legitimate user as “automated” – for instance, a fast shopper using autofill scripts, or multiple family members on the same IP appearing as high-frequency activity. When anti-bot or fraud systems err on the side of caution, the result is a **false decline** of a good order.

In sum, anti-bot and legacy fraud controls that are too heavy-handed can inflict more financial damage than the fraud they prevent.

# Current Anti-Bot Systems and AI Agents

Now consider how these systems might react to an AI shopping agent. By design, **automated agents don't behave like human users** – they might navigate a website in split-seconds or make API calls directly, without the telltale "human" patterns of mouse movements or keystrokes.

To an untrained bot detector, a surge of such activity looks identical to a malicious bot attack. Indeed, **fraud models have long treated "bot-like" behavior as high risk by default**[28].

Without adaptation, **legacy systems will flag AI agents as intruders**, resulting in a sharp rise in false positives (good orders blocked) as agentic commerce grows[29]. Ravelin (a fraud prevention firm) cautions that if merchants don't update their strategies, they will "see a substantial rise in blocked false positives" because **static systems can't distinguish benign AI shoppers from bad bots**[28].

The immediate implication is **lost revenue** for merchants who unknowingly turn away orders made by customers' new AI assistants. As one expert put it, *"more and more consumers will rely on a shopping method you are blocking"*[30] – a scenario no retailer wants to face.

On the flip side, **effective bot management can protect revenue** by keeping criminal bots out. Modern bot mitigation vendors tout behavioral analysis and machine learning that reduce false negatives and false positives.

For instance, solutions that analyze intent and context claim they can achieve near-zero false positives[31]. There is also evidence that **bad bot traffic imposes significant costs** if left unchecked: Imperva's 2024 Bad Bot Report found that scraping bots (often gathering data to fuel AI models) had grown to nearly **40% of total web traffic**, driving up infrastructure costs and polluting analytics[32][33].

In a recent 2025 initiative, Brightspot observed that advanced bots can evade many perimeter defenses (CDNs, WAFs) and that more **context-aware detection** (integrated at the application layer) was needed to reduce false positives[34]. This suggests that **next-gen anti-bot tools** – those leveraging richer data on user journeys – might handle the nuance of agentic traffic better than legacy, rules-based systems.

**Striking the Balance:** Going forward, merchants will need to recalibrate their anti-automation controls to differentiate **"declared" AI agents vs. "undeclared" bots** (more on this framework below). Several best practices are emerging:

- **Invisible or Low-Friction Challenges:** Instead of blatant CAPTCHAs, use invisible risk assessments (e.g. reCAPTCHA v3 scoring) that work behind the scenes[35]. Legitimate human or agent traffic isn't interrupted unless risk is high. This maintains user experience for the vast majority.
- **Behavioral Fingerprinting & Allowlisting:** Develop profiles for known *good* automation. For example, if an AI agent from a trusted platform is shopping (and perhaps *declares itself* via an API token), its behavior can be allowlisted or fast-tracked. Spec, a fraud solution provider, suggests maintaining **allowlists for verified beneficial bots** that support your business[36]. Early collaboration between merchants and agent providers can establish these trusted fingerprints.
- **Dynamic Friction:** Use a layered approach where suspected unknown bots are given a chance to prove legitimacy rather than being instantly blocked. For instance, present a stepped-up authentication if an agent-like pattern is detected but not recognized – this could be a 2FA challenge to the user's phone or an email confirmation. If the agent truly operates on the user's behalf, the user could confirm the transaction out-of-band. If not, a fraudster or undeclared bot will fail the challenge. This approach **preserves good orders** by adding friction only when needed, and only to the extent needed.
- **Monitoring and Analytics:** Invest in tools that provide visibility into automated traffic. Vendors are releasing **dashboards to monitor AI-originating orders**[37] so that fraud teams can see what percentage of traffic and sales are coming from agents, and how risk metrics compare to human transactions. Early data from Riskified showed that in some industries **AI-referred traffic carried 1.8–2.3× higher fraud risk than ordinary traffic**[38], highlighting that agents *can* be leveraged for abuse (e.g. automated scalping or reselling schemes[39]). Real-time analytics let teams adapt risk rules on the fly and spot emerging attack patterns.

# Anti-bot technology is at a crossroads

it must evolve from a blunt **“block all bots”** stance to a more nuanced **“identify and serve the right bots”** model. While this is somewhat true already, separating good bots (like Google) from bad, this is not true for bots attempting to log in or transact.

The **revenue imperative is clear** – merchants that overly restrict automated traffic risk alienating new customers and losing sales to competitors who welcome AI-assisted buyers[40]. (In fact, one could view this opportunistically: every time a rival’s site frustrates an AI agent with unnecessary friction, *that agent may take its user’s money elsewhere*, potentially to you, if you’ve enabled smoother agent access[40].)

Next we will delve deeper into the fraud risks introduced by agentic commerce and how to build a governance framework that allows “good automation” to flourish while keeping bad actors at bay.

# Fraud and Risk in the Age of AI Shopping Agents

Agentic commerce doesn't just change *who* (or what) is shopping; it changes the **fraud risk landscape** across identity, transaction monitoring, and post-purchase dispute management. Here we identify the main fraud vectors and challenges introduced when AI agents operate in commerce, and examine how issues like chargebacks and evidence need to be rethought.

# **Identity Ambiguity and Accountability:** A core question arises: *How do we know an AI agent truly represents a legitimate customer and not a fraudster?*

In traditional e-commerce, the assumption is a human user with their device is behind each action. Agentic commerce breaks that link – the **user might not be "in session" at the moment of the purchase**[41]. An AI agent can initiate transactions 24/7 without the human present. This introduces an identity challenge: the merchant sees an incoming order, but **who is actually behind it?** Is it a trusted agent acting for John Doe, or is it a malicious bot imitating an agent? Many classic fraud signals – device ID, IP geolocation, typing cadence – lose meaning here or can be spoofed by AI[42]. Fraudsters can attempt to **"wear a mask,"** making a bot look like a benign shopping agent. This ambiguity calls for a robust **"Know Your Agent"** approach (akin to KYC for customers). The concept of KYA ("Know Your Agent") has emerged to ensure there is an **identity and trust layer for AI agents** interacting with commerce[43]. Without verification, merchants risk situations where a fraudster could deploy an AI agent that *appears* to be a customer's automated assistant but is actually making unauthorized purchases.

Accountability ties into identity – if something goes wrong (e.g. fraudulent transactions or errors), **who is liable?**

If the human claims "my AI did it without my knowledge," is the merchant stuck with the loss? These questions are still largely untested, but in the interim, **clear agreements and disclosures are critical**. For example, Amazon's "Buy for Me" likely requires users to consent that Amazon (or the AI) is acting as an agent on their behalf, implicitly shifting some responsibility to Amazon's platform for securing the transaction.

In the payments world, liability is often assigned to whoever authenticated the user – if an agent platform processes the payment (as merchant of record or via wallet token), they may assume liability by design[44][45]. One expert argues there must be *"a clear assumption of liability by the platform"* in agent-mediated flows; otherwise, it's ambiguous who should perform Strong Customer Authentication (SCA) and who eats the loss on a failed or fraudulent transaction[44]

We may see **liability consolidating in fewer hands** – the agent platform or network – because they will control authentication in these scenarios[46]. Until such frameworks solidify, merchants should operate with caution: treat agent-initiated orders with an understanding that **the usual proofs of identity may not apply**, and push for shared-liability arrangements with any third-party agent platforms.

# Higher Return and Dispute Rates
## Agents make decisions differently than humans, and they may misinterpret preferences.

An AI could order an item the consumer finds unsatisfactory, or from a merchant the consumer wouldn't normally trust. This can lead to **increased returns or chargebacks due to "miscommunication"**[52]. For example, a shopper might see a credit card charge from an unfamiliar retailer and dispute it, not realizing their AI chose an alternate vendor for a better price[52]. Or, the AI might slightly err in product specs, resulting in a return (especially if info it relied on was inaccurate[53]). Signifyd notes that these factors *"could mean more chargebacks,"* as statements might list the agent platform or an unexpected merchant, confusing customers[52].

Merchant policies will need to account for this confusion – e.g. proactively communicate when orders are placed via agents and make it easy for customers to reconcile those charges with their expectations.

# **Chargebacks and Dispute Challenges**
# Chargeback handling becomes more complex when AI agents are in the mix.

Normally, to fight a fraudulent chargeback, a merchant gathers evidence: device/IP data, login timestamps, shipping address consistency, etc., to prove the legitimate cardholder likely made the purchase. With agent-mediated transactions, two things happen: (a) some evidence is obscured or anonymized (the “device” might be a cloud server or the agent’s system, not the user’s phone or PC), and (b) new parties are involved (the agent platform, possibly an API intermediary). Proving that *“the cardholder was indeed the one who approved this transaction”* can be tricky if the cardholder wasn’t hands-on.

As noted, liability frameworks may shift some burden to the agent provider – for instance, if an agent uses a wallet like Apple Pay, the biometric authentication at purchase provides strong evidence of user approval (and carries liability shift). Using **payment methods that require payer authentication for each transaction is a key mitigation**[54]. Apple Pay, Google Pay, or any tokenized wallet that prompts fingerprint/FaceID or a PIN ensures the human is in the loop at the moment of purchase and generates proof (cryptographic token with user auth).

By contrast, an agent using raw card details with no 3-D Secure challenge is risky – the bank might later say SCA wasn't performed and side with the customer by default in Europe. **Merchants are advised to favor agent-mediated flows that support liability-shifted payments** (3-D Secure, wallets, etc.), so that if a dispute arises, the issuer's logs show proper auth was done[22]. In regions like NA where SCA isn't mandated, this is harder – requiring a 3-D Secure challenge on every agent purchase could kill conversion.

It may be worth at least **flagging agent-initiated transactions at the payment level** (more on that in governance) so that issuers and networks know an AI was involved, which could in future lead to tailored dispute rules.

Another challenge: **evidence retention and past-behavior proof.** Suppose a customer claims “I didn’t order this, it was fraud.” If a human did it, the merchant might show that the device and IP matched the customer’s prior orders, and the shipping address was theirs – compelling evidence. But if an AI agent did it from a different IP or via server infrastructure, those data points won’t match the user’s historical profile. Merchants and their fraud vendors will need to leverage alternate signals.

This is where we need to get real and question the hype from some payment fraud vendors.

Many of these services are in a sky is falling moment. Market fear, and promise their solutions or consortium will be the answer. For the most part they are actually part of this problem, and likely don't actually have a solution yet. They are all hyper tuned to decline anything that looks like a bot.

But let's get real, some environmental data will be different, but **about 80% of fraud signals remain the same** even if an AI is involved – data like billing address, geo-location of delivery, card velocity, etc. still apply.

**Capture as much data as possible at each step** of the agent-driven journey. Having comprehensive logs – including any *agent identifiers, API keys, timestamps of user authorization*, and communications between the user and agent – will be vital if you need to defend a chargeback months later.

# fraud risks in agentic commerce are *amplifications* of existing e-commerce threats plus some novel twists.

Transactions happen faster and at greater scale, meaning a successful fraud can proliferate rapidly (as one report quipped, *"any scheme a fraudster can come up with can be executed with stunning speed and scale"* via automation[58]).

The **lack of human checkpoints** (no eyeballs on a screen to notice something phishy) places more onus on systems to detect anomalies. Nonetheless, it's not all doom and gloom: advanced fraud defenses can adapt. Machine learning models can be trained to recognize **"trusted agent" behavior vs. malicious automation** by analyzing metadata and cross-merchant patterns[59][60]. Many agents will still leave telltale signs that differentiate them from human fraudsters – consistent, predictable operation for genuine ones, versus erratic, multi-purpose probing by malicious bots[61].

And crucially, **merchants are not powerless**: by implementing new governance and verification mechanisms (as discussed next), they can insist on transparency and security from agent providers, thereby reducing ambiguity and aligning incentives for safe agent deployments.

# Declared vs. Undeclared Agents: A Framework for Governance

A key proposal in managing AI-driven commerce is to distinguish between **declared** and **undeclared** agent interactions. This concept provides a framework for governance and enforcement:

- **"Declared" agents** are **transparent, authorized, and accountable**. They identify themselves to the platform, and their usage is disclosed and consented to by users. A declared agent might use an official API or token, carry a digital signature, or be otherwise **recognized as an approved automation**. In practice, declared agents would operate under a trust framework – for example, an AI shopping assistant that has registered with the merchant (or a network) and has its identity verified, similar to how payment wallets or apps are registered.
- **"Undeclared" agents** are essentially **bots in disguise**. They attempt to mimic human shoppers or act without any upfront identification. This includes both malicious bots and even well-meaning scrapers or tools that haven't gone through any vetting. Undeclared agents do not announce their true nature, so from the merchant's perspective they are indistinguishable from either a regular user or a sophisticated bot attack.

# Why push for declared agents?

Because it creates a **clear pathway to trust and control.** If an AI agent declares itself, a merchant (or anti-fraud service) can apply **tailored policies**: allow it through with higher limits if it's known to be safe, or monitor it more closely if not yet vetted.

Declared status can come with requirements like cryptographic signing of requests, timestamps, standardized data sharing, etc., that provide auditability. In contrast, undeclared agents force merchants to rely on behavioral guesswork (which is error-prone and resource intensive).

**The goal is to incentivize AI agent providers and users to choose the declared path** through a mix of easier access, better performance, and policy compliance and to deter undeclared usage via detection and penalties.

# Mechanisms to Enable Declared Agents

What would it take to achieve a robust declared-agent ecosystem? Several measures are being explored:

- **Agent Registration & Verification:** Agent platforms (the companies providing AI shopping assistants) should implement **enhanced onboarding for users who want to enable purchasing**[62]. This might involve verifying the user's payment methods and identity in advance. It doesn't necessarily mean full government ID KYC for every shopper, but at least confirming that a valid card or bank account is on file and possibly that the issuing bank has vetted it (similar to how Apple Pay confirms a card with the issuer, generating a device token)[62]. When an agent has this backing, it can present a "verified agent" credential to merchants. The agent could be issued an **agent ID or certificate** asserting something like: *This request comes from ACME ShoppingBot v1.2 on behalf of user Jane Doe, who has been verified and payment source tokenized by Visa.* There are early moves in this direction – for example, **Human Security's *AgenticTrust*** initiative and Riskified's partnership aim to create a **unified trust framework** linking merchants, agents, and fraud signals[63][64].
- **Rigorous Agent Authentication:** Any agent with purchasing power should enforce strong auth for the user when it is activated or when it executes orders beyond a threshold. Essentially, treat **AI agents like fintech apps or wallets** – require users to log in to the agent with MFA (multi-factor auth) before it can spend[65]. If an agent is dormant for a while or the device context changes, re-authenticate. This prevents unauthorized parties from hijacking an agent without the user's knowledge. It also gives merchants more confidence that a real user is behind the agent's actions at the time of purchase. Klarna's evolution is a good analogy: it started with low-friction "buy now, pay later" but introduced more rigorous authentication as volumes grew[65]. Agents might start permissive but almost certainly will need to tighten security as they scale.

- **Transaction Signing and Tokens:** A declared agent system may use **cryptographic signatures** to validate transactions. For example, when an agent places an order on a merchant's site via API, it could sign the order payload with a private key that corresponds to its registered agent identity (or even the user's key). The merchant, or a network service, verifies this signature to ensure the request truly comes from a legit agent and hasn't been tampered with. MasterCard's experimental "Agent Pay" concept hints at this, using **cryptographic tokens (dynamic single-use identifiers)** for payments by agents[66]. By replacing static card data with tokens bound to each transaction, even if a token is intercepted, it can't be reused[66]. In a broader sense, cryptographic techniques can ensure that *"Agent A did place Order #123 and it was authorized by User U"* – providing non-repudiation. Standards may emerge (like the hypothetical x402 protocol Coinbase mentioned[67]) to handle real-time, machine-to-machine payments with built-in signing.
-

- **Activity Monitoring by Agent Providers:** Declared status isn't "set and forget." Agent platforms must self-police their agents' behavior. If a particular user's AI agent suddenly starts scraping hundreds of sites or making abnormally rapid transactions, the provider should detect that and throttle or investigate[68]. Essentially, agent companies need anti-fraud too – to ensure none of their agents (or underlying models) are exploited or go rogue. They should also offer **audit logs** to merchants or users on what actions the agent took and why, which is part of being accountable.
- **Payment-Level Indicators:** An important medium-term measure is for payment schemes (Visa, Mastercard, banks) to incorporate an **"agent indicator"** on transactions[69]. This could be a flag in the authorization message that denotes an AI agent initiated this payment on behalf of the cardholder. If such data is passed through, merchants and issuers can treat those transactions with appropriate checks. For example, an issuer seeing an "AI-agent purchase" flag might approve without an OTP if it recognizes the agent's device token and sees that biometrics were used, etc., or conversely might decline if the agent flag is present but their customer isn't enrolled for that service.

Crucially, if a dispute arises, this flag could help assign liability (e.g. if flagged and a fraud still happened, maybe the agent platform guarantees it). Visa and Mastercard are likely evaluating this as they don't want unchecked fraud via agents undermining trust in the payment system. Early forecasts from card networks show they see agentic commerce coming (Mastercard estimated up to 40% of e-commerce transactions could be AI-driven by 2025!)[70], so they have incentive to get ahead with standards.

# Controls for Undeclared Activity

No matter how much we encourage the "declared path," there will be undeclared bots – some malicious, some simply non-conforming. Platforms and merchants need a playbook to **detect and handle undeclared automation**

- **Advanced Behavioral Analytics:** Use AI to catch AI. Systems should analyze **journey context** – e.g., navigation sequence, timing, cross-session linkages – to spot when something non-human is likely in play. Spec's approach of looking at **sequence, cadence, and interaction patterns** to infer intent is instructive[71]. For instance, a bot (undeclared) might perform a perfect checkout in 2 seconds – far faster than any human. It might also ignore UI elements that a human would at least hover on. On the other hand, a *declared* agent might use an API that provides context explicitly. By comparing the two, detection models can flag likely undeclared bots for review or challenge. Importantly, this analysis should run **before login and through all stages**[72], since undeclared agents may not log in at all (guest checkout) or may create new accounts just for the attack.
- **Linking Related Events:** Undeclared bots often try to evade detection by using different accounts, IPs, or devices. Thus, **link analysis** is key – finding connections between activities that appear separate. As recommended, merchants can **persist identifiers across sessions** (via cookies, behavioral fingerprints, etc.) to recognize if the same actor is behind multiple attempts[73]. If an unknown automation hits a site repeatedly, it can be caught even if each attempt uses a fresh account. This helps isolate orchestrated bot campaigns.

# Intent-Based Scoring

Rather than a simple “bot or not” rule, leading practices suggest scoring automation by its likely purpose. An agent that, say, rapidly checks price API endpoints but doesn’t go to checkout might be a benign price comparison bot – possibly tolerable. But one that adds high-value items and tries many cards in succession is clearly malicious. By evaluating automation intent and consistency (does it behave like a price aggregator, a genuine shopper’s agent following a typical sequence, or like a card-testing script?), merchants can allow some flexibility.

**“Good”** automation tends to have consistent, repetitive patterns aligned with its role, whereas malicious bots will change tactics (adding random items, probing different site parts) as they search for vulnerabilities[61].

# Selective Friction and Enforcement

For suspected undeclared bots, deploy countermeasures at strategic points.

One strategy: **honeypots and tarpit responses** for bad bots[74]. For example, present a fake “discount code” field to only suspected bots – human users won’t see it, but a bot might fill it out, revealing itself. Or deliberately slow down response times (“shadow ban” the bot by sending it through long delays), which deters those that rely on speed.

Conversely, **do not slow down or interfere with traffic believed to be good or declared**[75][76]. The emphasis is on **surgical intervention** – mitigate bad actors at the last possible moment (e.g. at final submit) to gather maximum evidence first, and keep the path clear for normal customers and declared agents[74].

# Policy and Legal Deterrents

Platforms should update their **Terms of Service** to explicitly forbid undeclared automated access or bot scraping. While “TOS violations” alone won’t stop a hacker, it does give a legal basis to pursue serious offenders and sets expectations with legitimate users/developers.

Some merchants may choose to **block all agents by default** initially[77] (as Amazon reportedly is doing broadly, likely until their own agent solution matures). This is a short-term blunt approach and can’t hold forever, but in these early days it might be employed by risk-averse merchants. If so, they must be aware it’s an **“arms race of attrition”** – agent developers will continually tweak fingerprints to evade detection[78].

A better long-term strategy is creating a **path to compliance** (declared route) so that you can say “unauthorized bots aren’t allowed, but here’s how an agent can play by the rules.”

# Partnerships and Industry Collaboration

No single merchant can unilaterally enforce declared vs undeclared across the whole internet; this needs **industry standards and partnerships**. Encouraging signs include:

- The **Riskified + HUMAN Security alliance (2025)**, which is explicitly creating a "unified security framework" for agentic commerce[63]. Riskified brings fraud risk expertise and a merchant network; HUMAN brings bot detection and the new AgenticTrust solution. Together, they aim to give merchants **visibility and control over AI-driven interactions**, applying consistent trust decisions across human and AI traffic[79]. They've begun offering tools like *AI Agent Approve* (to facilitate safe merchant–agent communications via API) and *AI Agent Policy* (to enforce rules against things like excessive returns or reseller bots)[80][81].
- **Card Network and Bank Initiatives:** Mastercard's "Agent Pay" tokenization and Visa's explorations indicate the financial industry's recognition of agentic commerce. We can expect emerging guidelines from networks on how agents should authenticate users and identify themselves in transactions. Partnerships between agent providers and payment companies (e.g. Apple's Siri or Google's Assistant with Visa Checkout) could fast-track adoption of secure agent payments.
- E-commerce Platforms & AI Vendors: E-commerce software providers (Shopify, Magento, Mirakl for marketplaces[82], etc.) can build agent-friendly features into their platforms – for example, APIs or modules for approved agents to fetch product data and submit orders in a controlled manner.
- Simultaneously, AI vendors (OpenAI, Google, etc.) can incorporate safety: requiring agents to abide by robots.txt, to throttle interactions, and to disclose their identity via user-agent strings or tokens. There is even discussion of an HTTP header or similar that an AI agent could send like AI-Agent: ProviderName/AgentID as a declaration. While standards are nascent, proactive collaboration now will shape norms (the first movers can help set the rules to their advantage, by establishing trust marks and preferred integrations with merchants).

In essence, the declared vs undeclared framework is about **creating a "safe lane" for AI commerce** – analogous to car traffic having designated lanes or even separate roads for authorized vehicles (think HOV lanes), while disallowing or filtering out those that don't qualify. The *declared lane* is built on disclosure, verification, and data sharing; the *undeclared lane* is subject to checkpoint inspections and roadblocks. Over time, as declared agent usage becomes common and proven, we can expect undeclared bot traffic to be increasingly identified and isolated – much as email spam is largely filtered today by cooperative infrastructure.

# Practical Controls and Recommendations for Merchants & Platforms

TrustML Studio

How can online merchants and marketplaces **operationalize** these ideas? Here we outline a set of practical technical and policy controls to foster a secure "declared" agent ecosystem, while minimizing friction for legitimate users. We also discuss the operational implications – from fraud team processes to customer experience – of implementing these controls. These recommendations draw on industry reports and expert insights, tailored for North America and Europe contexts:

**1. Embrace Advanced Fraud Detection (AI vs. AI):** Traditional rules or legacy fraud systems will not suffice. Merchants should **upgrade to AI-driven fraud prevention that can adapt to new patterns**. Many vendors (Sift, Signifyd, Ravelin, Riskified, etc.) now highlight their machine learning models' ability to handle agentic commerce. For example, Ravelin notes that its **ML models will learn to identify the risks associated with AI commerce agents as they see more data**[83][84], differentiating trusted vs suspicious behavior. Merchants should actively query their solution providers: *What steps are you taking to enable good AI agents and stop bad ones?*[85]. If a provider doesn't have a clear roadmap, consider partnering with one that does or layering additional tools (like a specialized bot detection service). A concrete action item is to **capture more data points in the customer journey** (as many as privacy allows): even if an agent skips parts of the UI, instrument your APIs and site to log headers, interactions, and handoff points. This data feeds the ML systems to make accurate decisions.

**2. Develop an Agent Policy and Communicate It:** Formulate a clear **"Agent Usage Policy"** for your platform. This should define what automated agents may or may not do on your site. For instance: *Agents must identify themselves via X method; only certain data can be scraped; purchasing agents must use an approved API or undergo verification.* Include this in your terms of service and post guidelines for third-party developers if applicable. While not all bad actors will abide, legitimate companies will. This also prepares you for conversations with partners: if Google's AI shopping wants to integrate, you have a ready stance on requirements (e.g. "we will allow you through our bot firewall if you attest to identity and sign requests"). On the consumer side, **disclosure is key**: ensure that if an order is placed through an agent, the customer is notified (e.g. "Your purchase at Store X was completed by your AI assistant on [date]"). This manages expectations and reduces confusion that leads to chargebacks[86]. It might be as simple as an email or app notification from the agent platform, but as a merchant you can also include messaging in the package or receipt, like "Thank you for your order (placed via [AgentName])."

**3. Leverage Known Secure Payment Methods:** Steer agent-driven transactions toward payment options that provide built-in fraud protection and authentication. The earlier discussion noted **Apple Pay / Google Pay** (tokenized wallets) as ideal for agent transactions because they require biometric or device-level auth for each payment and shift liability (in many cases) away from the merchant[54]. Encourage use of these in agent flows – for example, if your site supports Apple Pay, an AI agent on a user's iPhone can complete the purchase seamlessly with FaceID confirmation.

In Europe, where PSD2 SCA is law, consider taking advantage of exemptions wisely. An agent transaction might not technically meet "merchant-initiated" criteria, so it could require SCA; however, if it's a low-value or a whitelisted beneficiary scenario, use those exemptions to avoid unnecessary friction (provided you trust the agent's origin). **Monitor your chargeback rates on agent transactions** specifically – if you see higher disputes, tighten the payment methods (perhaps require 3-D Secure challenge for agents until they build more history).

**4. Implement Spending Caps and Velocity Limits for Agents:** Especially in the early stages of this paradigm, it's prudent to **limit the blast radius** of potential fraud or errors by AI agents. This could mean setting **velocity limits**: e.g. an agent can only place X orders per hour/day on your site, or only $Y value per order until further verification.

These limits can be gradually raised as confidence grows (for instance, if an agent has done 100 orders with no issues, allow higher value purchases). Mastercard's vision mentioned **biometric approvals and spending caps as ways to build trust**[70]. A practical example: you might require any single order above $500 via an agent to get a one-time passcode confirmation from the user before processing.

For subscription or repeat digital goods purchases by agents, ensure there are **ceilings and alerts** (prevent an agent from buying 100 gift cards in 5 minutes, for example).

**Communicate these rules:** users should know that their agent might be limited in what it can do automatically. Transparency can also protect you legally and encourage users to properly configure their agents (maybe they can explicitly raise limits with additional verification).

**5. Partner with Agent Providers (APIs and Data Sharing):** If you identify certain AI agents that your customers are using (e.g. a surge of orders coming from the IP ranges of a known service or with a common user-agent string identifying "AI Shopper 1.0"), **open a dialogue** with that provider. It is in both parties' interest to ensure smooth transactions. This could lead to a partnership where you offer an **official API or feed** for the agent to use, with the requisite security (API key, encryption, etc.), and in return the agent provider agrees to pass along rich info (like the original user's account ID, a risk score, or an "agent attestation").

These partnerships might also involve **evidence-sharing**: if a fraudulent incident occurs through an agent, the agent provider should supply logs to help investigate and resolve it.

An example in motion is **Google's approach**: their new "AI Shopping" feature adds items to cart and checks out on behalf of the user with Google Pay, meaning the merchant's normal fraud tools see a Google Pay transaction (with all its protections) [87]. Google effectively acts as an intermediary, and merchants just need to ensure they don't mistakenly decline those orders due to unfamiliar patterns. As a merchant, being part of such beta programs (with Google, Amazon, etc.) will give you valuable experience and perhaps a competitive edge in capturing early agent-driven sales.

**6. Enhance Customer Support & Evidence Retention:** Operationally, prepare your customer service and fraud teams for this new mode. **Train support reps** to handle inquiries like "I got charged via this AI service, what is it?" – they should be able to explain and reassure customers rather than defaulting to "sounds suspicious, file a chargeback." Have readily available logs of agent transactions that CS reps can reference. For instance, if a customer says "I didn't order this $200 item," the rep can check and see a note that it was via the customer's linked agent at a certain time and perhaps gently remind them or troubleshoot why their agent did that.

From the fraud operations perspective, create a **protocol for disputes involving agents**: since evidence is different, perhaps include an extra data sheet in chargeback responses explaining that the purchase was AI-initiated but authorized by the user's credentials (if you have such proof). In contentious cases, reaching out to the agent provider for a supporting letter or record could turn the tide in your favor with the bank. Also, **store data a bit longer than usual** for agent orders – chargeback windows can be 3–6 months; given the novelty, having detailed info at hand (even beyond normal log retention) is wise.

**7. Phase Rollout with Metrics:** Implement these changes in phases and track metrics closely. As one guide suggested, start with **Day 0–30 visibility**: establish your baseline of current bot/agent traffic and where your detection blind spots are[88]. Then within 90 days, aim to have some **intent-based detection** live and measure outcomes like *conversion lift from trusted automation* and *reduction in false positives*[89][90].

Key metrics to watch include: **percentage of orders from non-humans**, approval rates for those vs humans, fraud chargeback rates segmented by human vs agent, and **false decline rates** for agent orders (how often an agent's attempted purchase is blocked and later proven legitimate)[91]. Improvements in these metrics indicate your strategy is working – e.g., you might aim to see false declines drop and a modest uptick in revenue from new automated customers as you ease their path.

Use these data to refine your rules and to demonstrate to senior management the value of accommodating agentic commerce (it's not just risk – it's also an opportunity to grab **new revenue streams** as consumers adopt AI shopping).

**8. Stay Abreast of Regulations and Influence Policy:** In NA and EU, regulatory bodies are beginning to pay attention. The EU AI Act, for example, emphasizes transparency in AI interactions, which might extend to requiring that automated agents **clearly reveal they are not human** when interacting commercially.

Additionally, **consumer protection laws** may evolve – Europe could mandate, say, that agents obtain explicit consent for each purchase above a certain value, or that there's a clear path for a consumer to get recourse if an agent mis-buys something (this could tie into chargeback rights). Merchants should keep an eye on these developments and even engage through industry groups to shape reasonable standards (ensuring regulations aren't so heavy-handed that they stifle innovation).

Being proactive in *self-governance* (via frameworks like we've discussed) will put the industry in a better position to argue it can manage the risks without overly prescriptive laws.

In the meantime, ensure **compliance with existing laws**: data privacy (don't collect more data on agent interactions than you can justify under GDPR, etc.), and payments compliance (still follow PSD2 SCA where applicable – "the machine did it" is not a legal exemption!).

# Operational Implications

Adopting the above measures will have implications across the organization: -

**Fraud/Risk Teams:** They will need new training and possibly new hires skilled in bot detection analytics. Expect an uptick in manual review caseload initially for odd-looking transactions (until models catch up). Fraud analysts must learn to **"think like an agent"** to differentiate genuine agent patterns from fraud – essentially a new domain of expertise. -

**IT/Engineering:** Implementation of agent-specific APIs, monitoring systems, and data pipelines for the extra signals will require engineering effort. There may also be performance considerations: e.g., allowing high-frequency API calls from agents might spike traffic – ensuring your infrastructure (and third-party services) can handle it is key. Coordination with DevOps to not accidentally block partner agent traffic in security configs (firewalls, CDN rules) is needed. -

**Customer Experience:** There is a chance of *increased customer inquiries* initially ("Why do I have to log in? My AI used to just buy it." or "Why was my agent blocked?"). However, if done well, the long-term customer experience should be smoother – customers who prefer AI assistance will gravitate to merchants that accommodate it with minimal hassle.

**User education** can help: perhaps a brief FAQ on your site, "Shopping with AI Assistants? Here's what to know."

# Revenue and Growth

By positioning your business to safely accept AI-driven purchases, you could tap into new growth. Early evidence suggests merchants can see **conversion gains by welcoming automated buyers that competitors turned away**[40]. In highly competitive sectors (airline tickets, electronics), being agent-friendly could become a selling point (even a B2B2C marketing angle: “Our site is optimized for your AI assistants!”).

On the other hand, one must track the **fraud loss trend** – if losses creep up due to agent abuse, it may be necessary to momentarily tighten controls and iterate. The success criterion is balancing fraud and friction such that overall profit (revenue minus fraud minus operational costs) grows.

# Toward a Declared-Path Future: Conclusions and Next Steps

The rise of AI “agentic” commerce represents a profound shift – one that promises more convenience for consumers and new revenue streams for merchants, **if and only if trust can be established**. The last year’s trends show growing consumer interest and the first real implementations of agent-based buying, accompanied by a chorus of warnings from risk experts about the accompanying fraud dangers. We have outlined how anti-bot technology and fraud prevention must evolve: pivoting from a mindset of *“block all bots”* to *“differentiate good agents from bad bots”*. This involves investing in adaptive AI-driven defenses, sharing data across the ecosystem, and developing **clear trust frameworks** so that legitimate AI agents are recognized and unwanted imposters are swiftly dealt with.

consent).

# Fundamentally, it comes down to Declaring vs. Hiding

Declared agents

Those that operate transparently and accountably can be integrated into the commerce flow with low friction, bringing benefits like faster checkouts and personalized payment optimization.

Undeclared bots

If allowed to proliferate would force merchants into a defensive crouch, adding friction everywhere and hurting user experience.

Our recommendations urge merchants, platforms, and payment providers to collaborate now on establishing the **“rules of the road” for agentic commerce**. This includes technical standards (for identity, signing, data sharing) and policies (for authentication, liability, and user

# The **path forward** can be envisioned in phases

**Short term** **(next 6–12 months)**, focus on readiness: upgrade fraud systems, start capturing agent-related data, implement basic agent recognition, and perhaps whitelist known partner agents. Put guardrails like spending limits in place and require user login/MFA for agent purchases to keep risk manageable[92]. Essentially, get the fundamentals right (like requiring accounts instead of guest checkout for agent users, so there's at least one checkpoint of user presence[92]).

Medium term **(1–2 years)**, as agent usage grows, work within industry groups to formalize trust signals (that payment flag, shared reputation databases for agents, etc.). Introduce more **seamless solutions**: for example, by this time you might replace overt CAPTCHAs with invisible checks entirely, having confidence your backend can spot bad bots without user friction[93]. Roll out refined ML models that are **80–90% accurate in distinguishing agent vs human vs bot**, and adjust your customer journey accordingly (perhaps even customizing content: e.g. show a different simplified page to a detected agent vs a human).

Long term **(3–5 years)**, aim for a mature declared-agent ecosystem. We might see certifications or ratings for AI agents ("Verified Shopping Agent" trustmark) that merchants can rely on. Regulators might also mandate certain practices, which by then you will have anticipated. Ideally, by this stage, **agentic commerce is mainstream and safely boring** – much like credit cards became ubiquitous after security measures (CVV, chip-and-PIN, 3DS) curbed the worst fraud. The merchants who adapt early will not only suffer fewer growing pains but will actively shape how this plays out, likely capturing loyal users who prefer agent-driven shopping

# A final thought on trust and user acceptance

As Bain's research highlighted, **trust is the sticking point** – only a quarter of consumers today feel comfortable letting AI fully handle purchases[5]. Building that trust will require the entire commerce ecosystem to demonstrate that convenience doesn't come at the expense of security or accountability. Every stakeholder has a role: agents must shop *"freely, but not anonymously or without control"*[94], merchants must innovate in fraud prevention without raising barriers for good customers, and platforms should be transparent with users about how, when, and why an AI agent transacts for them.

If we get this balance right – **declaring who/what is at the checkout and protecting both sides of the transaction**[94] – agentic commerce could indeed be the next major leap in e-commerce, delivering the effortless experiences consumers crave. If we get it wrong, it may well become a cautionary tale of innovation outpacing trust.

The recommendations in this report aim to ensure it's the former: a future where **human and AI shoppers coexist securely**, to the benefit of customers, merchants, and the digital economy at large.

# Sources:


- Bain & Co., *Agentic AI Commerce Hinges on Consumer Trust* – consumer survey on AI in shopping[5][6].
- Signifyd, *What is Agentic Commerce and How do Retailers Prepare?* – expert commentary on shifts in shopping and fraud challenges[9][86][47].
- Ravelin, *Agentic commerce fraud: How to protect your online shop* – recommendations on adapting fraud strategy (false positives, ML, signals)[28][56].
- Spec, *Agentic AI & The Bot Problem You Can't See* – best practices for distinguishing good vs bad automation and reducing friction[40][36][93].
- PSE Consulting, *Behind the Mask – How can merchants identify bots behind AI agents?* – trust framework ideas: agent verification, payment flags, short-term vs medium-term measures[22][95][69].
- Riskified Press Release (Aug 2025), *Riskified and HUMAN Security Partnership* – notes on a unified trust framework and early data on risk in AI-driven traffic[38][64].
- PaymentExpert, *What happens to payments when AI becomes the shopper?* – Signifyd's insights on payment method shifts, SCA, and liability in agentic flows[15][45].
- CapMonster Blog, *Captcha impact on conversion* – statistics on conversion drop due to CAPTCHAs[23].
- Riskified, *False Declines Guide* – statistics on false decline losses for merchants[26].
- Brightspot Report (Aug 2025), *Surge in Automated Bot Traffic* – bot traffic share and limits of current anti-bot tools[32][34].

[1] [2] [4] [28] [29] [30] [43] [56] [57] [60] [83] [84] [85] Agentic commerce fraud: Protecting your business | Ravelinhttps://www.ravelin.com/blog/agentic-commerce-fraud-ai-agent-fraud

[3] [14] [15] [16] [44] [45] [46] Agentic Commerce: How AI Agents Reshape Payments & SCAhttps://paymentexpert.com/2025/08/19/agentic-commerce-ai-payments-sca-liability/

[5] [6] [17] Agentic AI Commerce Hinges on Consumer Trust | Bain & Companyhttps://www.bain.com/insights/agentic-ai-commerce-hinges-on-consumer-trust/

[7] [8] [67] Why Agentic Commerce Needs Crypto to Scale | Coinbasehttps://www.coinbase.com/en-au/developer-platform/discover/launches/agentic-commerce

[9] [10] [11] [12] [13] [47] [49] [50] [51] [52] [53] [55] [58] [59] [86] [87] What is Agentic Commerce and How do Retailers Prepare? https://www.signifyd.com/blog/agentic-commerce/

[18] AI Agents in Ecommerce | Salesforce USHttps://www.salesforce.com/commerce/ai/agentic-commerce/

[19] Core Concepts - Agent Commerce Kithttps://www.agentcommercekit.com/overview/concepts

[20] What is Agentic Commerce? | Visa Navigatehttps://navigate.visa.com/europe/future-of-money/what-is-agentic-commerce/

[21] Consumers welcome AI agents across the shopping journeyhttps://www.emarketer.com/content/consumers-welcome-ai-agents-across-shopping-journey-

[22] [41] [42] [54] [62] [65] [68] [69] [77] [78] [92] [94] [95] Behind the Mask - How can merchants identify bots and bad actors behind AI agents? - PSE Consultinghttps://pseconsulting.com/insights/articles/behind-the-mask-how-can-merchants-identify-bots-and-bad-actors-behind-ai-agents/

[23] [24] [35] Captcha and its impact on conversion: how to protect your site without losing users | CapMonster Bloghttps://capmonster.cloud/en/blog/bypass-captcha/captcha-impact-on-conversion

[25] CAPTCHAs Are Destroying Your Conversion Rate - There's A Better ...https://authenticityleads.com/captchas-hurt-your-conversion-rate/

[26] [27] False declines: A guide for ecommerce merchants - Riskifiedhttps://www.riskified.com/learning/ecommerce-checkout-optimization/false-declines/

[31] 13 Top Bot Management Software for 2025 | Indusface Bloghttps://www.indusface.com/blog/top-bot-management-software/

[32] [33] [34] How to respond to the growing wave of AI-driven bot traffic - Brightspothttps://www.brightspot.com/cms-resources/technology-insights/brightspot-automated-bot-traffic-report-and-recommendations

[36] [40] [61] [71] [72] [73] [74] [75] [76] [88] [89] [90] [91] [93] Agentic AI & The Bot Problem You Can’t See | Spechttps://www.specprotected.com/blog/agentic-ai-the-bot-problem-you-cant-see

[37] [38] [39] [48] [63] [64] [79] [80] [81] Riskified Joins Forces with HUMAN to Help Merchants Embrace Trusted AI Shopping Agent Commerce - Riskifiedhttps://www.riskified.com/press/riskified-joins-forces-with-human-to-help-merchants-embrace-trusted-ai-shopping-agent-commerce/

[66] [70] Mastercard's Agent Pay: The Cryptographic Shield Redefining E-Commerce Security and Market Dominancehttps://www.ainvest.com/news/mastercard-agent-pay-cryptographic-shield-redefining-commerce-security-market-dominance-2505/

[82] Agentic commerce: The next revolution in online buying - Miraklhttps://www.mirakl.com/blog/agentic-commerce-the-next-revolution-in-online-buying