IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510128

Agentic Commerce: A Survey of Emerging Landscape

Janhavi Manore[1*], Sharvari Jamkar[1], Nupoor Joshi[1], Sukanya Gupta[1], Parag Jambhulkar[1]

Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India[1]

*Corresponding Author

Abstract: The rapid development of artificial intelligence (AI) has moved electronic commerce from using traditional algorithmic recommendations to relying on autonomous agentic systems. In this new approach, referred to as agentic commerce, large language model (LLM)-powered agents are able to independently discover, assess, negotiate, and carry out financial transactions on behalf of human users. The present survey offers a thorough overview of the developing agentic commerce environment. It looks at the basic architecture and the interoperability protocols that allow smooth machine-to-machine transactions, examines the change in the user experience going from manual browsing to intent-driven delegation, and systematically reviews the existing evaluation frameworks, pointing out the gap between shopping scenarios in simulation and actual real-world performance. The study also investigates the security, privacy, and economic issues specific to this system, such as intent misinterpretation, adversarial manipulation of the marketplace, risks associated with dynamic pricing, and the need for dynamic trust governance. Finally, it identifies the major open challenges and the key research areas that are needed if secure, transparent, and user-aligned agentic commerce platforms are to be developed. This work acts as a fundamental reference for researchers, developers, and industry professionals who are dealing with the next generation of automated digital trade.

Keywords: Agentic commerce, LLM agents, Autonomous transactions, E-commerce protocols, AI security and trust, Agent evaluation frameworks

I. INTRODUCTION

Large language model (LLM)-based agents are moving on from being employed for answering questions to performing actions. In the area of e-commerce, these agents can carry out the action of browsing catalogues, compare products, manage the shopping cart, and complete the checkout procedure with only a small amount of input from the user. Although this kind of delegation can reduce the amount of effort needed, it also introduces new requirements in the areas of reliability, security, authorization, and oversight.

Work is currently being carried out to set up the required infrastructure for this environment. The Agent Payments Protocol (AP2) provides authorization on the basis of a mandate for payments that are initiated by agents, in cases where humans are not present [1]. The Universal Commerce Protocol (UCP) defines a standard method by which agents, businesses, and payment providers can interact [2], while the Agentic Commerce Protocol (ACP) allows for agent-facing checkout and delegated payment [3]. Recent surveys and architectural studies show how fast the development is within this ecosystem [4–6].

The topic covers research in the areas of LLM security, agent evaluation, e-commerce, payment infrastructure, governance, and marketplace economics. It has already been demonstrated that there are risks linked to prompt injection and manipulation [7–10], and the shopping and marketplace benchmarks examine agent behaviour in ever more realistic environments [11–15]. Additional research focuses on protocol security, trust, consumer protection, and the economic effects of autonomous purchasing [16–19].

The survey covers the different developments in the following fields: (i) architectures and protocols, (ii) applications, (iii) security and reliability, (iv) evaluation and benchmarking, (v) trust and governance, and (vi) economic and market implications; it also mentions the open challenges and proposes avenues for future research.

The paper is structured as follows: Section II describes the survey methodology; Section III provides the background and includes details on the existing protocols; Section IV examines security and threats, and Section V looks at applications and evaluation. Sections VI to VIII in turn consider trust and governance, economic implications, and reliability. The final sections look at the open challenges, future directions, and conclude the paper.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510128

II. SURVEY METHODOLOGY

A. *Research Questions*

This survey is guided by the following research questions:

1. What are the main architectures and features of emerging agentic commerce systems?
2. What suggestions have been made regarding trust, security, authorization, and verification in the context of agentic commerce?
3. How are autonomous transactions, payments, and settlement supported in agentic commerce?
4. What are the protocols and mechanisms which enable interoperability between agents, users, and commerce platforms?
5. What impact do agentic commerce systems have on consumer behaviour, on competition, and on multi-agent marketplaces?
6. What are the major challenges and future research directions in agentic commerce?

B. *Sources and Search Strategy*

Relevant literature was identified through IEEE Xplore, ACM Digital Library, Scopus, Web of Science, ScienceDirect, SpringerLink, and arXiv. Google Scholar was additionally used for supplementary discovery and citation tracking. Search terms covered major aspects of the field, including “agentic commerce”, “agentic AI”, “LLM agent”, “autonomous transaction”, “agentic payment”, “agent security”, “agent verification”, “multi-agent marketplace”, “agent negotiation”, and “agent interoperability”. The search window covers publications from 2023 to 2026.

C. *Inclusion and Exclusion Criteria*

A paper was included if it addressed at least one relevant aspect of agentic commerce, including architectures, trust and security, authorization and verification, autonomous transactions and payments, protocols and interoperability, multi-agent marketplaces, negotiation, consumer behaviour, economic implications, or evaluation of commerce-related agents. Relevant conceptual, architectural, analytical, and review studies were also considered.

Papers were excluded if they focused solely on general-purpose LLM capabilities without relevance to agentic commerce, were unrelated to agents or commerce, duplicated an included publication, lacked sufficient technical or analytical content, or were not available in English. Preprints were included when they satisfied the same relevance criteria; where a peer-reviewed version was available, it was preferred.

D. *Limitations and Threats to Validity*

Agentic commerce is a rapidly evolving field, and terminology, protocols, and system architectures continue to change. A significant portion of recent work is also available as preprints. Differences in database coverage and terminology may result in relevant studies being missed. Therefore, the findings should be interpreted as a snapshot of the literature available during the survey period.

III. BACKGROUND

A. *Agentic Commerce*

Agentic commerce refers to commerce workflows in which software agents act on behalf of buyers to discover products, evaluate alternatives, and assist with or execute transactions [4, 11]. Unlike conventional e-commerce, where the user performs search, comparison and checkout directly, an agentic system delegates some of these activities to an AI agent, which may interpret a natural-language request, retrieve product information, compare alternatives, manage a cart or checkout session, and interact with payment infrastructure. Broader reviews of agentic AI describe architectural capabilities such as perception, planning, memory and tool use [5], while reviews of generative AI in e-commerce document applications and emerging research gaps [20]. Process-automation systems further illustrate the efficiency motivations for agentic commerce on the merchant side [6, 21].

A typical system involves several roles: the user, who expresses a purchasing goal or constraints; a buyer agent, which interprets the request and acts on the user’s behalf; a merchant or merchant service, which provides product information and fulfils the order; and one or more payment or credential providers. A simplified transaction proceeds through stages in which the agent discovers and evaluates products, a product is selected, a cart or checkout session is created, authorization is established, and payment and fulfilment are completed (Fig. 1). Some systems keep the user in the loop at checkout, whereas human-not-present flows allow an agent to complete a transaction within previously authorized conditions.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510128

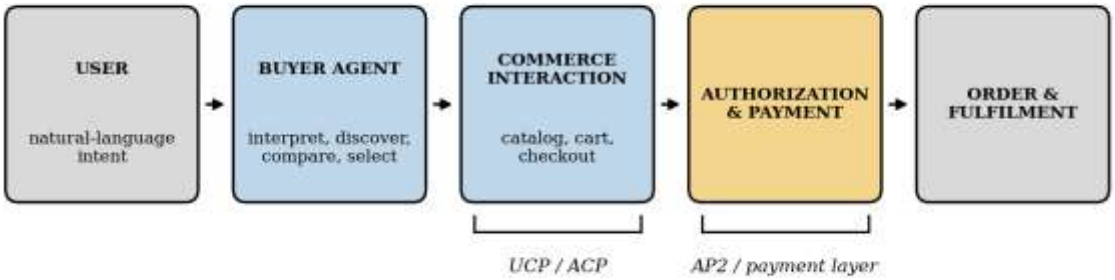


Fig. 1 Simplified agentic-commerce transaction flow

The increasing delegation of commerce activities introduces requirements beyond conventional transaction processing, including reliability, security, interoperability, transparency, and appropriate authorization and oversight [4, 16, 22].

B. *Protocols and Standards*

Several protocols and standards have emerged to support agent-mediated commerce. They address different layers of the ecosystem, including agent identity and trust, commerce interoperability, authorization, checkout, and machine-to-machine payments, and are therefore complementary rather than interchangeable (Table I). The growing number of protocol initiatives reflects the broader development of infrastructure for communication and transactions between users, agents, merchants, and automated services [4, 16].

TABLE I AGENTIC COMMERCE PROTOCOLS AND THEIR PRIMARY SCOPE

Protocol	**Primary purpose**	**Main mechanisms and scope**
AP2 [1]	Payment authorization	Mandates, verifiable authorization, transaction binding, receipts, and human-present and human-not-present payment flows
UCP [2]	Interoperable commerce and checkout	Cart management, checkout, identity linking, order lifecycle operations, multiple transports, and payment handlers
ACP [3]	Agent–business checkout and delegated payment	Checkout-session lifecycle and delegated payment mechanisms for agent-mediated transactions
ERC-8004 [16]	Agent identity and trust	Agent identification, trust, and reputation mechanisms for interactions involving autonomous agents
APP [16]	Agent payments	Agent-to-agent and agent-to-service payments, including metering and escrow mechanisms
x402 [16, 23]	Machine payments	Hypertext Transfer Protocol (HTTP)-native payment flows for application programming interfaces (APIs) and digital services
MPP [16]	Machine payments	Challenge–credential–receipt flows for machine-initiated payments
ERC-8183 [16]	Agentic commerce	Framework for agent-mediated commercial interactions and transaction workflows

These protocols collectively illustrate the layered nature of the emerging agentic-commerce infrastructure, with different standards contributing to interoperability, trust, authorization, and transaction execution.

IV. THREAT LANDSCAPE AND SECURITY CHALLENGES

A. *Security and Reliability Risks*

The increasing autonomy of commerce agents expands the security and reliability requirements of conventional e-commerce systems. An agent may interpret user requests, interact with external tools and services, process merchant-provided information, and initiate transactions with limited human intervention. Consequently, failures can arise not only from conventional software or infrastructure vulnerabilities but also from the interaction between reasoning, external information, tools, and transaction mechanisms [4, 5, 16].

Recent research identifies several broad areas of concern, including prompt and tool-related manipulation, insecure agent interactions, unauthorized actions, information disclosure, identity and trust failures, and weaknesses in the infrastructure

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IJARCCE.2026.1510128

supporting autonomous transactions [7, 8, 16, 17]. These risks are particularly important in commerce because errors can directly affect financial transactions, personal information, contractual obligations, and consumer choice.

B. *Manipulation and Adversarial Environments*

Commerce agents work in an environment where the information they encounter may not be entirely under the control of the agent provider. The product descriptions, search results, external tools, websites and other service responses they interact with can contain misleading or adversarial content. Work on tool-integrated LLM agents has shown that external content can influence agent behaviour and work on web agents has demonstrated that the interface itself can impact agent decisions [7, 8, 10].

Red-teaming studies of payment enabled agents do show that manipulation is possible even in the case where the underlying payment infrastructure continues to enforce its own authorisation requirements. The AP2 red teaming study is a case in point, showing that adversarial inputs can be used to manipulate the behaviour of a shopping agent without the attacker needing to in any way break the protocol's cryptographic authorisation mechanisms [9]. This serves to underline the fact that you can't just treat payment security as a sufficient safeguard for the entire workflow; you have to look at the agent, the environment it operates in and the transaction infrastructure as a whole.

C. *Protocol, Identity, and Transaction Risks*

Agentic commerce introduces new dependencies on protocols for identity, authorisation, payments, checkout and the communication between the various autonomous services. A weakness or ambiguity at any one of those interfaces can have an impact on the downstream transaction behaviour. The SoK on the security of autonomous LLM agents in agentic commerce highlights that risk can propagate across multiple layers: agent integrity, transaction authorisation, inter-agent trust, market interaction and compliance.

Protocol level research likewise shows the need to assess security across a range of agentic-commerce platforms rather than looking at individual agents in isolation. AIP-Bench provides a cross-platform study of protocol level attacks and defences, highlighting the increasing importance of a standardised security evaluation for agent mediated transactions [17]. The issues raised are of course not limited to payment authorisation but extend right through the chain of identity, delegation, communication and transaction execution.

D. *Privacy, Trust, and Consumer Protection*

Autonomous purchasing also brings with it non-technical risks. The agent will be processing personal preferences, account details, payment credentials and transaction history as it makes decisions on behalf of the user. A failure in privacy or identity management could therefore result in the exposure of information or the execution of actions which the user didn't expect.

Consumer protection research also highlights the issues of consent, transparency, liability and redress. The current consumer protection regime is predicated on the notion that a human is observing and taking part in the transaction. Agentic commerce can of course place an agent between the consumer and the merchant and this raises all sorts of questions about responsibility in the event of an agent making an inappropriate purchase, misreading the consumer's requirements or acting on conflicting incentives.

Trust is therefore affected not just by whether a transaction is technically secure but also by whether the user can understand, control and contest what is being done on their behalf. Governance research in this area also stresses the importance of appropriate human oversight, accountability, monitoring and mechanisms for managing autonomous systems throughout their life cycle [22].

E. *Broader Market and Governance Risks*

The effects of agentic commerce can extend beyond individual transactions. Autonomous agents may influence product visibility, consumer search, merchant competition, pricing, and market participation. Studies of agentic shopping have begun examining model-dependent purchasing behaviour, position effects, personalization, and the implications of delegating product discovery and selection to AI systems [11, 12, 19].

These developments introduce broader questions concerning fairness, competition, accountability, and the distribution of decision-making power between consumers, merchants, platforms, and agent providers. As agentic commerce becomes more integrated with existing digital markets, technical safeguards will therefore need to be complemented by governance mechanisms and consumer-protection frameworks.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510128

Overall, the literature indicates that the security of agentic commerce cannot be reduced to payment authorization or protection against individual attack techniques. It is a system-level concern spanning agent behaviour, external information, tools, protocols, identity, privacy, transaction infrastructure, and the surrounding market and regulatory environment [4, 16, 18]. These interconnected risks motivate the need for evaluation and governance approaches that consider both technical performance and the broader consequences of autonomous commercial decision-making.

V. BENCHMARKS AND EVALUATION ENVIRONMENTS

A. *Prompt-Injection Benchmarks*

AgentDojo provides an extensible environment for evaluating prompt-injection attacks and defences against tool-using agents, populated with 97 realistic tasks and 629 security test cases [7]. InjecAgent focuses on indirect prompt injection in tool-integrated agents and reports 1,054 test cases across 17 user tools and 62 attacker tools [8].

B. *Tool–Agent–User Interaction Benchmarks*

τ-bench simulates dynamic conversations between an LLM-simulated user and an agent equipped with domain application programming interface (API) tools and policy documents, across retail and airline domains [24]. Rather than judging an individual response, it compares the database state at the end of the conversation with an annotated goal state, giving a deterministic pass or fail instead of relying on a second LLM. It also proposes the pass^k metric, which repeats each task k times to measure reliability across trials. Its extension, τ^2-bench, studies a dual-control telecom setting in which both agent and user can invoke tools, and finds that performance drops significantly once users act rather than remain passive [25]. End-state comparison is close in spirit to the deterministic intent constraints discussed in Section VIII, but neither benchmark is adversarial.

C. *Deceptive-Interface Studies*

DECEPTICON studies how dark patterns affect web agents rather than human users, with 700 web navigation tasks (600 generated, 100 from real websites) designed to isolate individual dark patterns [10]. Effectiveness correlates positively with model size and test-time reasoning, and leading countermeasures, including in-context prompting and guardrail models, did not consistently reduce attack success.

D. *Shopping-Agent Benchmarks*

A distinct family of benchmarks asks whether shopping agents satisfy user goals, which is the capability that intent preservation must protect. ShoppingBench presents a real-world, intent-grounded benchmark in which tasks derive from user shopping intents [13]. EComAgentBench targets long-horizon shopping tasks with distributed hidden intent, i.e. user requirements revealed gradually across a conversation [14]. RealWorldShop evaluates and improves conversational shopping agents on real e-commerce tasks [26]. AgenticShop benchmarks agentic product curation for personalized web shopping [12]. Customer simulation has also been explored: LLM agents acting as digital twins of shoppers can evaluate conversational shopping assistants without relying exclusively on live traffic [27]. These benchmarks measure task success, personalization or conversational competence under benign conditions. They treat the user's goal as the reference for correctness, which is directly relevant to intent preservation, but none of the descriptions we reviewed injects adversarial mutations or models protocol-level objects.

E. *Marketplace and Economic Testbeds*

ACES pairs a platform-agnostic buyer agent with a programmable mock marketplace to study what AI agents choose to buy and why [11]. Randomized experiments over position and listing attributes show that agent choices can concentrate on few products, that all evaluated models favour the top row of a listing but disagree on preferred columns, and that models penalize sponsored tags, reward endorsements and respond to price, ratings and reviews in directionally expected but model-dependent ways. Magentic Marketplace offers an open-source environment for studying agentic markets with agents on both sides [28]. A dynamic multi-attribute auction benchmark asks whether LLM agents can price competitively [29], and Agentic Commerce World proposes an auditable, verifiable environment in which commerce runs can be inspected after the fact [15]. Role-coherence work studies whether agents keep to the role assigned by the principal in AI-mediated markets [30].

An earlier analysis of market concentration and position bias in LLM-driven marketplaces reports related findings [31]. ACES is, to our knowledge, the closest testbed to the controlled commerce environment needed for adaptive adversarial testing, since both pair a buyer agent with a controllable marketplace. The aims differ: ACES treats purchasing behaviour as an object of empirical study without adversarial manipulation, protocol-level objects or an explicit user-intent constraint. The biases it documents, particularly sensitivity to position and sponsored tags, nonetheless suggest realistic mutation classes, since content that a human would recognize as a paid promotion already steers agent choices without

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510128

any injected instruction. Auditable environments such as Agentic Commerce World are promising substrates for deterministic oracles because the full event record is available for checking.

F. *Comparison*

Table II compares the benchmarks along common dimensions.

TABLE II COMPARISON OF BENCHMARKS AND ENVIRONMENTS

Benchmark[a]	**Domain**	**Adversarial content**	**Stateful**	**Protocol-aware**	**Correctness judged by**
AgentDojo [7]	Email, banking, travel	Prompt injection (629 cases)	Yes	No	Task-specific security and utility checks
InjecAgent [8]	Tool use	Indirect injection (1,054 cases)	Limited	No	Rule-based check of injected call
τ-bench [24]	Retail, airline	None	Yes	No	Final database state vs. goal state
τ²-bench [25]	Telecom	None	Yes	No	Final state vs. goal state
DECEPTICON [10]	Web navigation	Dark patterns (700 tasks)	Yes	No	Success rate vs. dark-pattern effect
ACES [11]	Marketplace	None (bias study)	Yes	No	Not applicable (revealed preference)
ShoppingBench [13]	Shopping	–	–	No	Intent-grounded task success
EComAgentBench [14]	Shopping	–	Multiturn	No	Hidden-intent satisfaction
Magentic Mkt. [28]	Two-sided market	–	Yes	No	–
AIP-Bench [17]	Commerce platforms	Protocol-level attacks	–	Yes (per title)	–

[a] "–" = not determined from the sources reviewed

G. *Observations*

Security benchmarks excel at prompt-injection vectors (AgentDojo, InjecAgent) and user-interface-level manipulation (DECEPTICON). Shopping benchmarks measure goal satisfaction but under benign conditions, and marketplace testbeds expose behavioural biases. A broader evaluation framework that jointly considers security, user-intent fidelity, transaction correctness, and marketplace behaviour remains an important direction for future research.

VI. INTENT SPECIFICATION AND VERIFICATION

A central challenge in agentic commerce is determining whether an agent's actions and final outcomes actually satisfy the user's intended requirements. Unlike conventional software, where expected behaviour can often be expressed through explicit rules, commerce agents may operate on natural-language preferences, contextual information, and changing external states. An agent may therefore complete a transaction successfully while still violating an important user preference.

A. *Approaches to Outcome Verification*

Methods currently used for assessing the results of agents can generally be categorized as model-based or specification-based. When LLMs are used as judges, they assess the actions or final outcomes of an agent with the aid of another language model and give a qualitative rating or decision. Although such approaches are flexible and appropriate for subjective requirements, the judgments they produce can differ from one model, prompt, or repeated evaluation to another.

Instead, specification-based approaches express user requirements in the form of explicit, machine-verifiable constraints. They enable deterministic and reproducible verification of requirements such as price limits, product attributes, authorization conditions, or transaction states. By checking properties against sequences of actions and system states, formal methods and runtime verification go beyond this approach and become useful for policies that involve temporal or state-dependent requirements.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510128

B. *Hybrid Approaches and Open Challenges*

There is no one verification method that can deal with all kinds of user intent. While explicit constraints offer reliable verification for clearly defined requirements they are less effective when it comes to subjective or context-dependent preferences. Although model-based evaluation is more flexible it does lead to variability among evaluators and demands thorough validation. It is for these reasons that hybrid approaches can be used to combine deterministic checks for important constraints with model-based evaluation for qualitative requirements.

A key open challenge in the area of agentic commerce is the creation of reliable representations of user intent that are capable of supporting both flexible natural-language preferences and machine-checkable transaction constraints. The future evaluation frameworks should also take into account user-intent fidelity, transaction correctness, security, and changing commerce states all at the same time rather than assessing these aspects separately.

VII. EVALUATION METRICS AND EXPERIMENTAL PRACTICE

A. *Metrics in the Literature*

Commonly reported measures include attack success rate, number of unique failures found, tests to first failure, coverage of states and attack categories, failure diversity, false-positive rate and cost. Reliability-oriented work adds pass^k over repeated trials [24]. A simple efficiency measure for a testing strategy s is

$$Efficiency(s) = \frac{unique\ violations\ discovered}{total\ tests\ executed} \quad (1)$$

B. *Fair Comparison*

Strategies should be compared using the same tasks, agent configuration, mutation operators, test budget and stopping rules.

C. *Reproducibility and Statistical Rigour*

Model randomness, closed models and changing APIs make results hard to reproduce, and agent behaviour depends strongly on the model used [11]. Repeated runs, confidence intervals and significance tests are needed before claiming that one strategy is better.

VIII. COMPARATIVE ANALYSIS

A. *Literature Matrix*

Table III places the surveyed papers along common dimensions. Years follow the preprint or publication date given in the reference.

TABLE III MASTER LITERATURE MATRIX (CONDENSED)

Paper	Year	Primary theme	Contribution type
AgentDojo [7]	2024	Security	Benchmark
InjecAgent [8]	2024	Security	Benchmark
τ-bench [24]	2024	Reliability	Benchmark, deterministic end-state
τ²-bench [25]	2025	Reliability	Benchmark, dual control
Whispers of Wealth [9]	2026	Security (AP2)	Red-teaming attacks
DECEPTICON [10]	2026	Security (user interface)	Benchmark, dark patterns
SoK [16]	2026	Security	Systematization, 12 vectors
AIP-Bench [17]	2026	Security (protocol)	Taxonomy, benchmark, defense
ACES [11]	2026	Marketplace behaviour	Sandbox, empirical study
Magentic Marketplace [28]	2025	Marketplace	Environment
ShoppingBench [13]	2025	Shopping	Benchmark
EComAgentBench [14]	2026	Shopping, hidden intent	Benchmark
RealWorldShop [26]	2026	Shopping	Benchmark, improvement
AgenticShop [12]	2026	Curation	Benchmark
Auction benchmark [29]	2026	Pricing	Benchmark
Agentic Commerce World [15]	2026	Auditable environment	Environment
Role coherence [30]	2026	Agent behaviour	Empirical study
TessPay [32]	2026	Settlement	Infrastructure
Verifiable events [33]	2026	Auditability	Architecture

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510128

Paper	Year	Primary theme	Contribution type
Decision-centered arch. [34]	2026	Architecture	Reference architecture
Agentic Purchase [35]	2026	Trust	Framework
Economics of agentic commerce [19]	2026	Economics	Model
x402 measurement [23]	2026	Adoption	Measurement
Governance practices [22]	2023	Governance	Practices
Oracle problem [36]	2015	Testing theory	Survey
LLM-as-judge [37–39]	2023–24	Evaluation	Study, surveys

B. *Trends*

Three trends stand out. First, work has moved from general tool-use agents (2024) to commerce-specific environments (2025–2026), with a sharp increase in 2026 as protocols such as AP2, UCP and ACP appeared. Second, attention is shifting from single-prompt injection to cross-layer and protocol-level threats [16, 17]. Third, evaluation environments are becoming auditable and programmable [11, 15, 28], which is a precondition for controlled mutation testing. In parallel, architecture and trust proposals [32–35] and governance and policy analyses [18, 22, 40] are growing faster than the testing methods needed to validate them.

C. *Strengths and Weaknesses of the Field*

Strengths: there is a rich set of attack demonstrations, several mature injection benchmarks, a well-developed theory of oracles, and growing shopping and marketplace environments. *Weaknesses:* (i) benchmarks are siloed, so security benchmarks ignore commerce objects and shopping benchmarks ignore adversaries; (ii) correctness is often judged by an LLM or by task success rather than by an explicit intent constraint; (iii) test selection is almost always fixed or random, with no equal-budget comparison against adaptive strategies; (iv) many sources are recent preprints with limited replication; and (v) architecture and governance proposals state requirements without methods for testing them.

IX. OPEN CHALLENGES AND RESEARCH GAPS

Based on the literature reviewed, the following gaps appear open.

- There is no standard benchmark focused on intent preservation in commerce.
- Protocol-aware testing against AP2- and UCP-style objects is limited.
- Most studies demonstrate individual attacks, and few compare testing strategies head to head under equal budgets.
- Many studies rely on LLM judges rather than deterministic oracles.
- Multi-agent and agent-to-agent trust is under-studied.
- Human-not-present autonomous payments raise the stakes of undetected intent violations.
- Chronic, cumulative drift (e.g. O2P in [16]) is invisible to per-transaction tests.
- Fast-changing models make results hard to reproduce.
- Current agent-payment protocols leave authorization gaps that existing defences do not close; a systematization proposes a layered defence and research roadmap, converging with several gaps above from a security perspective [16].

X. FUTURE RESEARCH DIRECTIONS

Promising directions include standard intent-preservation benchmarks that combine the adversarial coverage of AgentDojo-style suites with the goal-grounding of shopping benchmarks; richer intent languages that handle fuzzy preferences and hidden intent; adaptive testing beyond bandits; testing across several protocols; fault injection and regression testing after model updates; auditable commerce environments with oracle hooks; and ethical, legal and regulatory analysis. A layered defence architecture and benchmark agenda has been proposed for authorization gaps in agent-payment protocols [16]; testing frameworks such as the one this survey motivates could serve as a concrete evaluation mechanism for such defences once implemented.

XI. CONCLUSION

Agents that shop and pay for users create a gap between transactions that are validly authorized and transactions that match user intent. Existing work supplies attack demonstrations [9, 10, 16], security and shopping benchmarks [7, 8, 13, 14], marketplace testbeds [11, 28], trust architectures [32, 34] and a mature theory of test oracles [36], but comparatively little evaluates testing strategies head to head using deterministic intent verification in protocol-aware commerce settings.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510128

ACKNOWLEDGMENT

Author Contributions. Janhavi Manore, Sharvari Jamkar, Nupoor Joshi, and Sukanya Gupta contributed equally to this work, including the literature review, analysis, and writing of the manuscript. Parag Jambhulkar supervised the work and reviewed the manuscript. All authors read and approved the final manuscript.
Competing Interests. Not applicable. The authors have no competing interests to declare. *Funding.* Not applicable.
Data Availability. Not applicable. This is a survey paper, and no new datasets were generated or analysed. *Research Involving Human and/or Animals; Informed Consent.* Not applicable.

REFERENCES

[1] Google and AP2 contributors, "Agent Payments Protocol (AP2) Specification, Version 0.2.0," 2026. [Online]. Available: https://ap2-protocol.org. Accessed: Sep. 28, 2026.
[2] UCP contributors, "Universal Commerce Protocol (UCP) Specification," 2026. [Online]. Available: https://ucp.dev. Accessed: Sep. 28, 2026.
[3] OpenAI and Stripe, "Agentic Commerce Protocol," 2026. [Online]. Available: https://www.agenticcommerce.dev/docs. Accessed: Sep. 30, 2026.
[4] "Agentic commerce: A survey of how AI agents are reshaping commerce," *TechRxiv* preprint, 2026, doi: 10.36227/techrxiv.176972193.39211542/v1.
[5] "The rise of agentic AI: A review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges," *Future Internet*, vol. 17, no. 9, p. 404, 2025.
[6] P. Kodali, "Agentic commerce: Architectural frameworks and governance models for AI-driven retail transactions," *ResearchGate* preprint, 2025.
[7] E. Debenedetti, J. Zhang, M. Balunović, L. Beurer-Kellner, M. Fischer, and F. Tramèr, "AgentDojo: A dynamic environment to evaluate prompt injection attacks and defenses for LLM agents," *arXiv:2406.13352*, 2024, doi: 10.48550/arXiv.2406.13352.
[8] Q. Zhan, Z. Liang, Z. Ying, and D. Kang, "InjecAgent: Benchmarking indirect prompt injections in tool-integrated large language model agents," in *Findings of the Association for Computational Linguistics: ACL 2024*, 2024, doi: 10.48550/arXiv.2403.02691.
[9] T. Debi, W. Zhu, and P. S. Gupta, "Whispers of wealth: A systematic red-teaming study of the agent payments protocol (AP2)," in *Proc. 2026 Int. Conf. Intelligent Multimedia, Networking, and Security (IMNS)*, Atlanta, GA, USA, 2026, pp. 1–6, doi: 10.1109/IMNS67862.2026.11655291.
[10] P. Cuvin, H. Zhu, and D. Yang, "DECEPTICON: How dark patterns manipulate web agents," *arXiv:2512.22894* (v2), 2026, doi: 10.48550/arXiv.2512.22894.
[11] A. Allouah, O. Besbes, J. D. Figueroa, Y. Kanoria, and A. Kumar, "What is your AI agent buying? Evaluation, biases, model dependence, & emerging implications of agentic e-commerce," in *Proc. ACM Web Conf. 2026 (WWW '26)*, 2026, pp. 8697–8700, doi: 10.1145/3774904.3792943.
[12] "AgenticShop: Benchmarking agentic product curation for personalized web shopping," in *Companion Proc. ACM Web Conf. 2026*, 2026, doi: 10.1145/3774904.3792724.
[13] "ShoppingBench: A real-world intent-grounded shopping benchmark for LLM-based agents," *arXiv:2508.04266*, 2025.
[14] "EComAgentBench: Benchmarking shopping agents on long-horizon tasks with distributed hidden intent," *arXiv:2606.17698*, 2026.
[15] "Agentic Commerce World: An auditable and verifiable environment for vibe commerce," *arXiv:2608.02441*, 2026.
[16] Q. Mao, J. Wang, Y. Liu, L. Zhu, C. Ma, and J. Yan, "SoK: Security of autonomous LLM agents in agentic commerce," *arXiv:2604.15367* (v2), 2026, doi: 10.48550/arXiv.2604.15367.
[17] "Protocol-level attacks on agentic commerce platforms: A cross-platform taxonomy, AIP-Bench, and unified defense," *arXiv:2607.21824*, 2026.
[18] CERRE, "Agentic AI and consumer protection," 2026. [Online]. Available: https://cerre.eu/publications/agentic-ai-and-consumer-protection/
[19] "The economics of agentic commerce," *arXiv:2608.08395*, 2026.
[20] "Generative AI in e-commerce: A literature review of current applications, gaps and future directions," *Asia-Pacific J. Business Administration*, 2025, doi: 10.1108/APJBA-08-2025-0575.
[21] "EcoptiAI: E-commerce process optimization and operational cost minimization through task automation using agentic AI," in *IEEE Conf. Proc.*, 2025.
[22] Y. Shavit et al., "Practices for governing agentic AI systems," OpenAI, Tech. Rep., 2023.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510128

[23] "How agentic is agentic commerce? A population-scale measurement of x402 adoption and authenticity," *arXiv:2607.12575*, 2026.
[24] S. Yao, N. Shinn, P. Razavi, and K. Narasimhan, "τ-bench: A benchmark for tool-agent-user interaction in real-world domains," *arXiv:2406.12045*, 2024, doi: 10.48550/arXiv.2406.12045.
[25] V. Barres, H. Dong, S. Ray, X. Si, and K. Narasimhan, "τ²-Bench: Evaluating conversational agents in a dual-control environment," *arXiv:2506.07982*, 2025, doi: 10.48550/arXiv.2506.07982.
[26] "RealWorldShop: Benchmarking and improving conversational shopping agents in real-world e-commerce," *arXiv:2609.38974*, 2026.
[27] Sun et al., "LLM agent meets agentic AI: Can LLM agents simulate customers to evaluate agentic-AI-based shopping assistants?," 2025.
[28] "Magentic Marketplace: An open-source environment for studying agentic markets," *arXiv:2510.25779*, 2025.
[29] "Can LLM agents price competitively? A dynamic multi-attribute auction benchmark for agentic commerce," *arXiv:2608.00102*, 2026.
[30] "When agents shop for you: Role coherence in AI-mediated markets," *arXiv:2604.26220*, 2026.
[31] A. Allouah et al., "Market concentration and position bias in LLM-driven autonomous marketplaces," 2025.
[32] "TessPay: Verify-then-pay infrastructure for trusted agentic commerce," *arXiv:2602.00213*, 2026.
[33] R. Srivastava, "Building trust in autonomous commerce: A verifiable global event architecture," *arXiv:2607.19436*, 2026.
[34] "A decision-centered reference architecture for trustworthy agentic commerce," *arXiv:2607.18347*, 2026.
[35] "Agentic purchase: A trust-aware multi-agent framework for autonomous e-commerce with live marketplace integration," in *IEEE Conf. Proc.*, 2026.
[36] E. T. Barr, M. Harman, P. McMinn, M. Shahbaz, and S. Yoo, "The oracle problem in software testing: A survey," *IEEE Trans. Softw. Eng.*, vol. 41, no. 5, pp. 507–525, 2015, doi: 10.1109/TSE.2014.2372785.
[37] L. Zheng et al., "Judging LLM-as-a-judge with MT-Bench and Chatbot Arena," in *Advances in Neural Information Processing Systems*, 2023, doi: 10.48550/arXiv.2306.05685.
[38] J. Gu et al., "A survey on LLM-as-a-judge," *arXiv:2411.15594*, 2024, doi: 10.48550/arXiv.2411.15594.
[39] K. Schroeder and Z. Wood-Doughty, "Can you trust LLM judgments? Reliability of LLM-as-a-judge," *arXiv:2412.12509*, 2024, doi: 10.48550/arXiv.2412.12509.
[40] "Agentic commerce: Opportunity or existential risk?," in *Oxford Law Pro Edited Volume*. Oxford, U.K.: Oxford Univ. Press, 2026, doi: 10.1093/9780198972877.003.0119.

 This work is licensed under a Creative Commons Attribution 4.0 International License