

When the customer is a machine: a systematic literature review and integrative framework of agentic commerce

Mahyar Turchi Moghaddam

mtmo@mmi.sdu.dk

University of Southern Denmark

Mina Alipour

University of Southern Denmark

Klaus G Grunert

Aarhus University

Systematic Review

Keywords: agentic commerce, AI agents, delegation, customer journey, generative engine optimization, systematic literature review

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11222729/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

When the customer is a machine: a systematic literature review and integrative framework of agentic commerce

Anonymous author(s)

Details withheld for double-blind review

Abstract

Artificial intelligence agents now search, evaluate, negotiate, and pay on behalf of consumers, while firms deploy agents that create content, personalize offers, and complete transactions. Agentic commerce therefore weakens a foundational assumption of marketing thought and marketing measurement: that the recipient of marketing activity is a human being. The evidence is scattered across marketing, information systems, computer science, economics, and law, and it has a largely forgotten first act in the shopbot research of the late 1990s. Following the SPAR-4-SLR protocol with PRISMA 2020 reporting, we screen 5,157 records from Scopus, Web of Science, and EBSCO Business Source and add verified snowballing to assemble 140 primary studies spanning 1994 to 2026. Each study is coded on influence direction, customer journey stage, delegation depth, method, and research wave. The synthesis yields an actor taxonomy with a five-level delegation ladder, an influence triad that separates marketing to, through, and by agents, an integrated account of trust and governance infrastructure, and a market-level analysis connecting both research waves. The Agentic Commerce Reference Framework consolidates this evidence into four auditable layers with explicit cross-layer dynamics, a ten-proposition research agenda, and five measurement instruments for analytics practice. A complete replication package accompanies the article.

Keywords: agentic commerce; AI agents; delegation; customer journey; generative engine optimization; systematic literature review

1 Introduction

For three decades, digital commerce has been designed around a human actor. A person sees the advertisement, reacts to the warning that only two items are left, compares offers across browser tabs, clicks the button, and types the card number. That design assumption is now weakening from both sides of the market at once. On the demand side, consumers delegate shopping tasks to conversational assistants and increasingly autonomous agents, from product discovery and comparison to checkout itself (Allouah et al., 2026; de Bellis and Johar, 2020; Frank et al., 2026). On the supply side, firms deploy generative agents that write product content, design advertising, personalize offers, negotiate, and transact (Wang et al., 2025; Puri et al., 2026; Aslam et al., 2026). The institutional infrastructure has followed quickly. Within roughly eighteen months, Google’s Agent Payments Protocol, the OpenAI and Stripe Agentic Commerce Protocol, Visa’s Trusted Agent Protocol, and Mastercard’s

Agent Pay were all announced (Google, 2025; OpenAI and Stripe, 2025; Visa, 2025; Mastercard, 2025). They share a single purpose: letting software buy things in a way that merchants, card-issuing banks, and regulators can verify. Consultancies size the delegated purchasing opportunity in the trillions of dollars by 2030 (Schumacher et al., 2025), and interface data make the projection concrete. The assistant platforms now intermediating discovery sit atop a search market in which a single firm controlled about 80 percent of desktop and 90 percent of mobile queries as of 2025 (Ortiz-Freuler and Castells, 2026). Even partial migration of that traffic into agent-generated answers therefore reroutes the primary demand channel of the digital economy. Marketing scholarship has registered the shift. A recent research note argues that AI customers require a dedicated subdiscipline of machine marketing (Sarstedt et al., 2026), and journals have begun publishing empirical work on the value consumers derive from conversational commerce, the established term for shopping conducted through chat and voice interfaces (Schultz and Kaiser, 2025). What that emerging subdiscipline lacks is a systematic evidence base, and supplying one is the purpose of this article.

Why does the machine customer matter for theory rather than only for operations? Because it breaks the mechanisms marketing has spent a century learning to use, and it breaks them in three distinct places. First, visibility shifts from capturing human attention to being retrieved by a machine. One can read retrieval as attention under different rules, but the practical levers change: whether an offer exists for an agent depends on structured data, verifiable claims, and third-party signals. In one controlled experiment, restructuring content alone raised a source’s presence in generated answers by up to 40 percent (Aggarwal et al., 2024), and audit studies point in the same direction (Li et al., 2026a; Kumar and Lakkaraju, 2024). Second, persuasion acquires a second, non-human target with its own sensitivities. Audits of commercially deployed shopping agents document position biases that differ across providers and even reverse between model versions, penalties for sponsored tags alongside rewards for platform endorsements, and demand that concentrates on a few modal products and reshuffles when the underlying model updates (Allouah et al., 2026). A field experiment adds the human-side counterpart: a single disclosure of machine identity before the conversation can reduce a channel’s purchase rate by more than three quarters (Luo et al., 2019). Third, the subject of measurement changes. Click paths, session logs, and last-touch attribution presume a person navigating pages. None of them describes an agent that, holding a consumer’s mandate, the authority the consumer has granted it to act, reads ten catalogs, queries three merchants, and buys once. A discipline whose analytics apparatus is built on observing humans must now decide how to observe delegation.

The scholarly response is substantial but scattered, and it does not yet know itself as one literature. Consumer researchers examine when people grant purchasing authority to machines and how they respond to machine persuasion (Wang and Lee, 2026; Farmer and Harding, 2026; Shamim et al., 2026). Computer scientists audit what shopping agents buy (Allouah et al., 2026), benchmark them (Ling et al., 2026; Wang et al., 2026), and demonstrate how easily their recommendations can be manipulated (Li et al., 2026a; Kumar and Lakkaraju, 2024). Economists design advertising auctions for generated answers (Dütting et al., 2024; Dubey et al., 2024; Hajiaghayi et al., 2024) and document pricing algorithms that reach supra-competitive prices without communicating (Schlechtinger et al., 2023). Legal scholars ask who is liable when a delegated agent transacts and whether platform regulation anticipated machine gatekeepers (Busch, 2025; Bostoen and Krämer, 2026; Nguyen,

2026). Design researchers build the payment, identity, and authorization layer that makes any of this transactable (Chua et al., 2026; South et al., 2025), and taxonomies of marketing agents together with strategy chapters have begun to map the toolset and its firm-performance stakes (Puvvadi et al., 2025; Foltean and Hera, 2026). These communities rarely cite one another, and almost none of them cite the field’s first act: the agent-mediated electronic commerce program of the late 1990s. That program built shopbots and negotiation agents, documented sellers gaming comparison engines, and produced field evidence that brand premia survive machine-mediated choice (Guttman et al., 1998; Maes et al., 1999; Smith and Brynjolfsson, 2001). That first program stalled for reasons its field-defining survey named at the time: personalization (adapting offers to an individual), shared semantics (agents and merchants describing products in compatible terms), service discovery (agents finding relevant sellers at all), and interaction protocols (standard messages agents can exchange). Above all, it named trust in delegation (He et al., 2003). Each item on that list names something the current protocol build-out exists to supply, which makes the first wave the field’s only long baseline and its natural control condition rather than a historical curiosity.

No existing review integrates these strands. Prior secondary studies (reviews and other syntheses) map consumer-side effects of agentic recommendation and delegation (Balaskas, 2026), conversational commerce (Lim et al., 2022), adoption of conversational agents (Ling et al., 2021), customer-level outcomes of deployed AI agents (Srinivasa Raja et al., 2026), and the first-wave comparison shopping systems (Pathak, 2010). Each covers one region of the phenomenon; none spans both research waves, treats the direction of marketing influence as an explicit coding dimension, connects consumer evidence to the governance and payment infrastructure now being built, or draws consequences for measurement. This review does all four, guided by four research questions. **RQ1:** How has research on commerce-capable AI agents developed across its two waves, the agent-systems wave of roughly 1994 to 2010 and the wave that follows the diffusion of large language models from late 2022, and which actors and levels of delegated autonomy does it describe? **RQ2:** Through which mechanisms is marketing influence exerted on, through, and by AI agents along the customer journey? **RQ3:** Which factors shape consumer trust in and acceptance of delegated shopping, and which governance and infrastructure conditions does the literature identify? **RQ4:** What are the market-level consequences of agentic commerce, and what do they imply for measurement and management?

We answer these questions through a systematic review of 140 primary studies published between 1994 and 2026, assembled from database searches of Scopus, Web of Science, and EBSCO Business Source and extended through verified backward and forward snowballing. The review follows the SPAR-4-SLR protocol (Paul et al., 2021) with PRISMA 2020 reporting (Page et al., 2021). The review makes four contributions, and each moves beyond mapping the state of the art. First, it connects two research waves that have so far developed with little mutual citation, and it shows result by result that several central second-wave market findings replicate first-wave findings across a twenty-five-year technology shift. We read such replication as unusually strong evidence for a domain this young. Second, it introduces two shared constructs the literature currently lacks: a five-level delegation ladder that makes studies of assistants, advisors, and autonomous buyers commensurable, and an influence triad that separates marketing to agents, through agents, and by agents. Third, it consolidates the synthesis in the Agentic Commerce Reference Framework, a four-layer architecture

with sub-components, cross-layer dynamics, and explicit usage guidance for the marketing, consumer research, human-computer interaction, information systems, software engineering, economics, and legal communities, in which every element is traceable to coded primary studies. Fourth, it converts the framework’s exposed gaps into a ten-proposition research agenda and five measurement instruments, from share of model voice to mandate-based attribution, that operationalize the agenda for analytics practice. All screening decisions, codes, and counts are released in a replication package from which every number and data figure in the article can be recomputed. The anonymous package is available at this link: <https://figshare.com/s/af06b7238fc45bed37fb>.

The article proceeds as follows. Section 2 develops the conceptual background: the field’s temporal structure, the construct space, the theoretical apparatus, and a positioning against prior reviews. Section 3 reports the method in full. Section 4 profiles the corpus. Sections 5 to 8 answer the research questions. Section 9 builds the integrative framework, its dynamics, and its use across research communities. Section 10 develops the theoretical implications, the research agenda, and the implications for analytics, management, and policy. Sections 11 and 12 state the limitations and conclude.

2 Conceptual background

2.1 Two waves and a bridge decade

Commerce-capable software agents predate large language models by a quarter century, and the first wave was not a footnote but a complete research program with theory, systems, experiments, and field data. Its theoretical core was a consumer buying behavior model that mapped software agents onto need identification, product brokering, merchant brokering, negotiation, purchase, and evaluation (Guttman et al., 1998), together with deployed systems that occupied those stages: comparison shopbots, negotiation marketplaces, and recommendation agents (Maes et al., 1999). Around the systems grew an empirical literature whose questions map directly onto today’s: how decision aids change choice, how agent recommendations are evaluated, and how sellers respond strategically. Interactive decision aids were shown to restructure rather than merely shorten choice. In a controlled shopping experiment, a recommendation agent reduced the size but raised the quality of consideration sets, and it increased the share of participants choosing objectively better options while lowering search effort (Häubl and Trifts, 2000). The machine changed what consumers considered, not only how fast they considered it. Agent recommendations for unfamiliar products depended on context, and the direction of that dependence reversed with detail. With little information, an attractive familiar context lifted the unfamiliar item by association; once item-specific detail was added, the same attractive context invited comparison and lowered the item’s evaluation (Cooke et al., 2002). Buyers adjusted their search behavior around the new intermediaries: in one early survey, mere awareness of shopping agents made buyers about 35 percent more likely to combine agent-based and manual search rather than rely on a single channel (Sen et al., 2006). Comparison-agent research documented paradoxical effects of decision aids on comparison shopping (Wan et al., 2009), price frames that moderated search behavior at shopping agents (Grewal and Lindsey-Mullikin, 2006), and differences between search and experience goods in perceived agent effectiveness (Aggarwal and Vaidyanathan, 2003). Later in the wave, it added persona effects for older users of virtual shopping

agents (Chattaraman et al., 2014).

The first wave also produced the market-level results that anchor this review’s continuity thesis. Price dispersion persisted under comparison agents, contrary to frictionless-market predictions, and search-cost mechanisms explained why (Pathak, 2012; Snaith, 2003). Sellers learned to game the machine gatekeeper through selective price disclosure and strategic ordering (Hajaj et al., 2015, 2017), an adversarial dynamic that anticipates today’s generative engine optimization. Price-setting agents produced cyclical price wars in simulation (Greenwald and Kephart, 1999; Leloup, 2003). Game-theoretic analysis showed that shopping agents, by putting every seller one click apart, change equilibrium pricing as if competitors were physically co-located (Iyer and Pazgal, 2003), and early assessments projected the market impact of artificial shopping agents (Redmond, 2002; Ma et al., 2010). Above all, the wave’s signature field study showed that at a shopbot, where price comparison is nearly costless, the most heavily branded retailers still commanded price premia: brand still mattered when a machine ranked the offers (Smith and Brynjolfsson, 2001). Design research meanwhile prototyped the transactional machinery, from secure agent negotiation protocols (Wang et al., 1999) and bid-forecasting mechanisms for autonomous bidding agents (Kehagias et al., 2005) to personalized comparison engines (Yuan, 2003) and decision-analytic marketplaces (Karacapilidis and Moraïtis, 2001). The program then stalled, and its field-defining survey closed by naming the open problems precisely, and none of them was intelligence. They were personalization (agents knowing enough about a person to act for them), semantic interaction (systems sharing enough meaning to transact), discovery (finding trustworthy counterparties), and interaction protocols (shared standards for negotiation and payment). The fifth and hardest was trust, with both a social facet, whether people would let software decide for them, and a technological facet, whether the infrastructure could make such delegation safe (He et al., 2003). Nothing on that list concerns intelligence; everything on it concerns interoperation and institutions.

What followed was not silence but a change in what researchers studied, and the corpus makes the change visible. Between 2011 and 2021 the coded studies shift from building agents to studying the human and institutional preconditions of delegation. The psychology of algorithm reliance matured in this bridge decade. Its landmarks are aversion to algorithms after seeing them err (Dietvorst et al., 2015), appreciation of algorithmic advice in many judgment tasks (Logg et al., 2019), task-dependent resistance concentrated in subjective domains (Castelo et al., 2019), and the word-of-machine effect separating utilitarian from hedonic contexts (Longoni and Cian, 2022). Consumer-experience accounts of living with autonomous systems and connected objects followed (Puntoni et al., 2021; Hoffman and Novak, 2018), capped by an analysis of the adoption barriers specific to autonomous shopping systems (de Bellis and Johar, 2020). Legal scholarship coined the algorithmic consumer as a market-structure category and anticipated the competition and advertising consequences of purchase decisions executing in code (Gal and Elkin-Koren, 2017). A small continuing stream of first-wave work still tracked shopbot markets and seller gaming (Pathak, 2012; Hajaj et al., 2015, 2017; PILIK et al., 2016). The decade closed with the field’s largest field experiment on machine selling: undisclosed chatbots sold as effectively as proficient human workers, while disclosure of machine identity before the conversation reduced purchase rates by 79.7 percent (Luo et al., 2019). The bridge decade thus supplied the delegation psychology and the legal vocabulary that the second wave now runs on, while agent systems themselves lay dormant.

The second wave, dating from the diffusion of large language models in late 2022, changes three parameters at once. Agents are general purpose, so one assistant spans the journey instead of one shopbot per function (Patil et al., 2026; Potluri and Chittimalla, 2026; Ishida et al., 2025). Agents are linguistic, so they read the same persuasive content humans read, which helps explain why persuasion, manipulation, and disclosure findings transfer to them (Friestad and Wright, 1994; Farmer and Harding, 2026; Guo et al., 2025). And agents are being wired into payment-grade infrastructure with cryptographic mandates, verified identity, and merchant endpoints (Chua et al., 2026; South et al., 2025; Google, 2025), which addresses the first wave’s institutional bottleneck directly. Whether that is sufficient is an empirical question this review keeps open; that it differs from 1999 rests, so far, on the new infrastructure and is an argument rather than a measured result.

2.2 Construct space: what agentic commerce is

We define agentic commerce as commercial exchange in which an AI agent, acting under a mandate from a consumer, a firm, or a platform, materially shapes at least one stage of the customer journey (Lemon and Verhoef, 2016). The definition has four components that carry the analysis, and the corpus operationalizes each. The *principal* is the party whose interests the agent nominally serves. Misalignment between principal and agent is a recurring empirical object, from inferences about a firm agent’s motives (Farmer and Harding, 2026) to seller agents inferring a buyer’s willingness to pay from dialogue (Alavi and Nozari, 2026). The *mandate* is the scope of authority granted, which the infrastructure layer is converting from a vague expectation into a verifiable artifact (Chua et al., 2026; South et al., 2025). The *capability* is what the agent can execute, which we formalize as delegation depth in Section 5. The *journey locus* is where in search, evaluation, transaction, or post-purchase the agent acts, which the evidence map of Section 9 shows is highly uneven.

Three boundary decisions follow. Purely informational customer-service chatbots without transactional or delegation capability fall outside scope; their large literature is already synthesized (Lim et al., 2022; Ling et al., 2021), and including it would dilute the delegation construct that defines the phenomenon. Agents from all three actor groups, consumer-side, firm-side, and platform-side, are in scope, because the influence relations among them are exactly what a marketing analysis must capture. We speak of actor groups rather than market sides, since the underlying exchange remains one between consumers and firms. And both waves are in scope, because excluding the first would discard the only field-tested market evidence the domain possesses and would sever the current cycle from its own control condition.

2.3 Theoretical apparatus

Five theoretical traditions recur across the corpus, and each is transformed rather than merely applied when the customer is a machine. The persuasion knowledge model describes how targets learn to recognize and cope with influence attempts (Friestad and Wright, 1994). The corpus displaces it twice. Consumers deploy different persuasion knowledge against machine persuaders. Perceived benevolence, the belief that the agent wants to act in the customer’s interest rather than the company’s, becomes a decisive inference (Farmer and Harding, 2026). Time pressure moderates whether machine appeals outperform human ones (Shamim et al., 2026), and persuasion intensity sets the

boundary at which human and AI streamers diverge in live commerce (Lu and Li, 2026). Meanwhile agents themselves possess almost no persuasion knowledge, as the manipulation results of Section 6 demonstrate, so the model acquires a second, machine-side referent that theory has not yet named.

Choice architecture research explains how the arrangement of options steers decisions (Thaler and Sunstein, 2008). In agentic commerce the architecture doubles: designers arrange context for the agent, and the agent arranges options for the person. Both stages are empirically active. Agents inherit position and framing sensitivities (Allouah et al., 2026), execute nudging on consumers at scale with an accompanying autonomy paradox (Choubey and Chakraborty, 2026), and constrain the option sets consumers ever see, with product expertise and purchase habits dampening the perceived risk of the constraint (Rabassa et al., 2022). Agency and information economics supply the third lens. Delegation creates a classic principal-agent structure, but with a twist the corpus documents: the agent’s conversational interface leaks the principal’s private information, allowing seller-side inference of willingness to pay nearly one for one from dialogue alone (Alavi and Nozari, 2026). Verified product information itself becomes a priced intermediation layer (Ventirozos and Shardlow, 2026), and Hayekian analysis asks whether delegated agents can transmit the dispersed knowledge markets depend on (McDavid et al., 2026). The customer journey perspective (Lemon and Verhoef, 2016) supplies the framework’s horizontal axis, because delegation depth and evidence density vary sharply by stage, and recent work already rereads the whole journey through the lens of agent intervention (Vasileva, 2025). Finally, the psychology of algorithm reliance introduced above (de Bellis and Johar, 2020; Longoni and Cian, 2022; Castelo et al., 2019; Dietvorst et al., 2015; Logg et al., 2019) explains the ceiling that separates what agents can do from what consumers let them do. Strategic treatments of AI in marketing frame the managerial stakes (Davenport et al., 2020; Huang and Rust, 2021) but stop short of agent-native transactions, which is the territory this review maps and the territory the machine-marketing agenda now names (Sarstedt et al., 2026).

2.4 Positioning against prior reviews

Because five secondary studies border this one, positioning deserves more than a table. Pathak (2010) synthesized the first-wave comparison shopping literature as decision support systems research; it remains the best guide to that system class but predates language models, covers only the advisor level of delegation, and contains no marketing-influence or governance analysis. Ling et al. (2021) systematically reviewed adoption of conversational agents and consolidated the technology-acceptance antecedents on which our Section 7 builds; its scope is the interface, not the transaction, and agent autonomy is outside its frame. Lim et al. (2022) mapped conversational commerce and flagged the commercial turn of conversational interfaces; written at the cusp of the LLM diffusion, it could not yet observe autonomous purchasing, machine-directed marketing, or the payment layer. Srinivasa Raja et al. (2026) meta-analyzed 468 effect sizes on customer-level outcomes of AI agent implementation. It provides the quantitative anchor for what Section 6 defines as the through-direction of our triad, marketing that reaches the human through an agent. As an effect-size synthesis, it deliberately abstracts from mechanisms, journey stages, and market structure. Balaskas (2026) is the nearest neighbor: a systematic review mapping agentic AI in e-commerce and its consumer effects. It confirms the momentum of the delegation literature but stops at the consumer boundary. Influence on agents, agents as marketing producers, transaction infrastructure,

Table 1: Positioning against related secondary studies. Filled circle: full coverage; open circle: partial coverage; blank: no coverage.

Secondary study	Two waves	Influence direction	Governance and infrastructure	Market outcomes	Replication package
Pathak (2010)			○	○	
Ling et al. (2021)					
Lim et al. (2022)		○			
Srinivasa Raja et al. (2026)		○			
Balaskas (2026)		○			
This review	●	●	●	●	●

law, and market outcomes remain outside it, and the first wave appears only in passing. Finally, Sarstedt et al. (2026) is an agenda-setting research note rather than a review; it argues that machine marketing should exist, and the present article supplies the systematic evidence base, the shared constructs, and the measurement program that the proposed subdiscipline requires.

The comparison identifies precisely what remained undone and this review supplies. It contributes a two-wave corpus with verified lineage between the waves, and the influence-direction construct, without which marketing to agents, through agents, and by agents blur into one undifferentiated AI literature. It integrates the consumer evidence with the governance and payment infrastructure that determines which delegation depths are reachable at all, adds market-level synthesis, and releases a recomputable evidence base. Table 1 summarizes the coverage comparison.

3 Methodology

3.1 Review design and rationale

The review follows the three stages of the SPAR-4-SLR protocol, assembling, arranging, and assessing (Paul et al., 2021), is reported in line with PRISMA 2020 (Page et al., 2021), and adopts concept-centric synthesis so that findings are organized by construct rather than by paper (Webster and Watson, 2002). We chose a domain-based systematic review over a bibliometric or purely narrative design for two reasons grounded in the phenomenon (Snyder, 2019). First, the literature crosses disciplinary boundaries with incompatible vocabularies, so replicable retrieval and explicit criteria matter more than in a settled field. Second, the research questions require synthesis of findings and mechanisms, not only mapping of publication patterns, which rules out a purely bibliometric design. The unit of analysis is the primary study: a journal article, conference or workshop paper, book chapter, or working paper making an empirical, formal, design, or substantive conceptual contribution about AI agents shaping commercial exchange. Working papers are eligible because the second wave is young enough that excluding them would erase a large share of the frontier; each is labeled by type in the corpus register so that any analysis can be repeated on the peer-reviewed subset. Figure 1 reports the complete flow by SPAR-4-SLR stage with PRISMA counts.

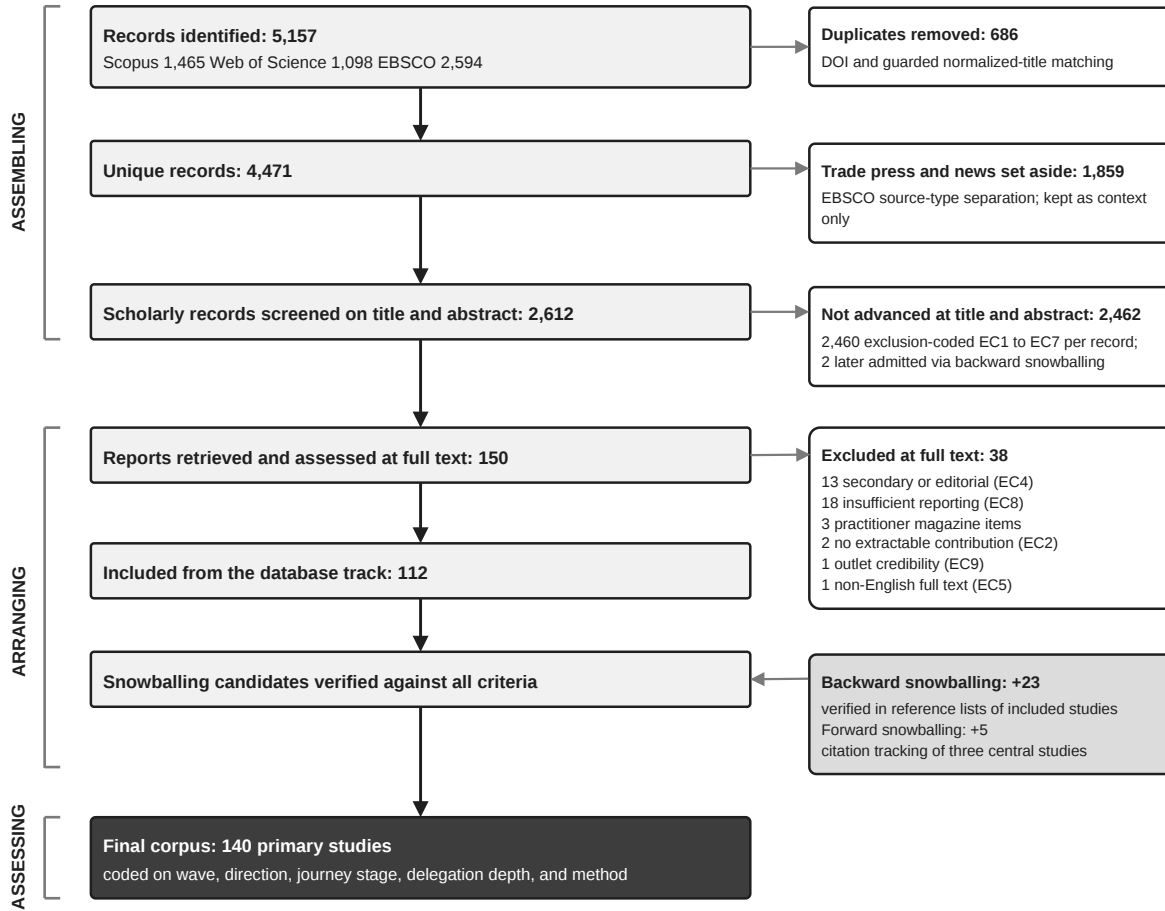


Figure 1: Review protocol and flow. The spine reports the PRISMA counts at each stage; side boxes report removals and additions with their criterion codes; brackets on the left map the stages onto the SPAR-4-SLR protocol of assembling, arranging, and assessing.

3.2 Databases

Three databases were searched. Scopus and Web of Science provide the two broadest curated indexes of peer-reviewed business, information systems, economics, and computer science research, with complementary coverage of conference proceedings. EBSCO Business Source adds business-school journals and practitioner-adjacent scholarly outlets that the citation indexes cover unevenly. Its cost is that it also indexes trade press, which we handle explicitly at screening. Together the three cover the disciplinary spread the phenomenon demands. Preprint servers were not searched directly; recent unindexed work enters through forward snowballing, which retrieves it selectively through its connection to included studies rather than in bulk.

3.3 Search string development

The search string pairs two concept groups, the agent and the commercial exchange:

```
("agentic commerce" OR "agentic shopping" OR "agentic AI" OR "AI agent*" OR
"LLM agent*" OR "LLM-based agent*" OR "language model agent*" OR "autonomous
agent*" OR "shopping agent*" OR "buying agent*" OR "machine customer*" OR
"algorithmic consumer*" OR "conversational commerce" OR "generative engine
optimi*" OR "answer engine optimi*") AND (commerce OR "e-commerce" OR retail*
OR shopping OR shopper* OR marketing OR advertis* OR consumer* OR purchas* OR
checkout OR payment* OR pricing OR brand*)
```

Concept group one was harvested from three sources: the terminology of known landmark studies, industry protocol vocabulary, and the keyword lists of the candidate reviews in Section 2.4. It deliberately mixes established phrases with quoted compound terms so that precision comes from phrase structure rather than from narrowing the domain. Concept group two enumerates the commercial exchange broadly, because agent studies signal commerce in many different words. Two design decisions control precision. The string is applied to titles, abstracts, and keywords in Scopus and Web of Science, which keeps the searched concepts central to the retrieved work and is established practice in marketing reviews. In EBSCO Business Source it runs on full text, which trades precision for recall and explains that database’s larger and noisier return. No publication-date restriction was applied, so both waves enter through the database track. Snowballing additionally secures the first-wave canon: it is densely cited by included studies and is therefore reached reliably through reference lists, even where legacy vocabulary such as software agent is too ambiguous for precise retrieval across three decades. A validation set of three known relevant studies spanning the corpus’s breadth, the consumer-side systematic review (Balaskas, 2026), the generative engine optimization experiments (Aggarwal et al., 2024), and a consumer trust study (Zhao et al., 2025), was fixed before execution as a recall check; all three were retrieved. The searches were executed in July 2026 and returned 5,157 records: 1,465 from Scopus, 1,098 from Web of Science, and 2,594 from EBSCO Business Source. Per-database syntax, execution notes, and raw exports are in the replication package.

3.4 Screening procedure

Table 2 states the criteria in unified form: six paired dimensions apply at every stage, and three additional exclusion screens apply at full text, where the information to judge them first becomes available. De-duplication by DOI and guarded normalized-title matching removed 686 records, leaving 4,471 unique records. Because EBSCO Business Source indexes trade press alongside scholarship, source-type separation preceded screening: 1,859 news and trade items were set aside as context material, leaving 2,612 scholarly records. Titles and abstracts were then screened against the paired criteria. Before full application, all three authors independently screened an initial calibration batch against the criteria; agreement across the three raters was substantial (Fleiss’ $\kappa = 0.76$), and disagreements were resolved by discussion before the remaining records were screened. Every decision was recorded with a criterion code; the coded decisions and the marked export sheets in the replication package preserve the full trail. Judgment calls concentrated at two boundaries, and we state how they were resolved. Studies of service chatbots were included only when the agent held transactional or delegation capability (EC7); for example, chatbot studies measuring purchase outcomes of conversational selling were retained, while studies of complaint-handling dialogue quality

Table 2: Screening criteria in unified form. Dimensions one to six pair an inclusion and an exclusion criterion and apply at all stages; dimensions seven to nine are exclusion screens applied at full text.

Dimension	Inclusion criterion	Exclusion criterion
Subject	IC1: an AI agent materially shapes a commercial exchange or its marketing, on the consumer, firm, platform, or market side	EC1: agent not commerce relevant (robotics, games, network or energy operations)
Contribution	IC2: empirical, formal or analytical, design or system, or substantive conceptual contribution	EC2: no extractable commerce or marketing contribution
Scope	IC3: commerce or marketing constructs present (visibility, choice, persuasion, delegation, trust, transaction, pricing, advertising, market outcomes)	EC3: physical or robotic agents only
Type	IC4: journal article, conference, workshop, or chapter; working papers eligible and labeled	EC4: secondary studies and editorials (used for positioning, not coded as primary studies)
Language	IC5: full text in English	EC5: non-English full text
Access	IC6: full text obtainable for assessment	EC6: full text not accessible
Interaction scope	EC7: service chatbots with no transactional or delegation capability	
Reporting depth	EC8: extended abstract or abstract-only publication with insufficient detail for data extraction	
Outlet credibility	EC9: outlet fails peer-review credibility screening	

were not. Studies of autonomous agents outside commerce, in energy markets, network operations, and robotics, were excluded under EC1 even when auction or pricing language appeared, unless the contribution addressed consumer-facing commerce. The review’s scope is accordingly consumer commerce: business-to-business settings enter only where an included study examines them directly, as with delegated B2B data curation (Adam et al., 2025) and cross-border seller-side agents (Xiao et al., 2025), and all other findings should be read as claims about consumer markets. This pass advanced 150 records to full-text assessment; the 2,462 records not advanced carry per-record exclusion codes in the export sheets, and two of them later entered the corpus through backward snowballing.

3.5 Full-text assessment

All 150 retrieved documents were read at full text against the same criteria plus the three full-text screens. Thirty-eight were excluded, each with a coded reason in the replication package. Thirteen were secondary studies and editorials, which inform Sections 2 and 4 but are not primary evidence (EC4). Eighteen were extended abstracts or abstract-only publications whose reporting could not support extraction (EC8), a category dominated by conference abstract volumes. Three were practitioner magazine pieces outside the primary-study definition, two were descriptive chapters without an extractable contribution (EC2), and one outlet failed credibility screening (EC9). One article’s full text is not in English (EC5): an observational study of brand visibility in AI-generated

tourism recommendations, published with an English abstract only, which language verification at full text moved from the provisional set to the exclusion log. The database track therefore contributed 112 primary studies. Two further full-text decisions illustrate the standard applied under EC8, which excludes on insufficient reporting rather than on length. A one-page research note reporting an eleven-month price panel collected through a shopping agent was retained because its data and design are fully reported (Snaith, 2003). A two-page simulation report on algorithmic pricing was retained because its market design, learning setup, and outcomes are extractable (Schlechtinger et al., 2023). A longer conference abstract on agent effects on norm compliance was excluded because its design could not be extracted, despite a promising question.

3.6 Snowballing

Backward snowballing examined the reference lists of all included studies. A candidate was admitted only when three conditions held: it is verifiably cited by included studies, with the citing identifiers logged; it satisfies the inclusion criteria; and it contributes evidence or theory not already represented. Twenty-three studies entered this way, spanning the first-wave canon (Guttman et al., 1998; Maes et al., 1999; He et al., 2003; Häubl and Trifts, 2000; Smith and Brynjolfsson, 2001; Gal and Elkin-Koren, 2017), the delegation and algorithm-reliance literature of the bridge decade (de Bellis and Johar, 2020; Longoni and Cian, 2022; Castelo et al., 2019; Dietvorst et al., 2015; Logg et al., 2019; Puntoni et al., 2021; Hoffman and Novak, 2018), persuasion knowledge theory (Friestad and Wright, 1994), and recent computational work on generative visibility, silicon samples, and advertising mechanisms (Aggarwal et al., 2024; Kumar and Lakkaraju, 2024; Sarstedt et al., 2024; Brucks and Toubia, 2025; Brand et al., 2023; Dütting et al., 2024; Dubey et al., 2024; Hajiaghayi et al., 2024; South et al., 2025). Forward snowballing tracked citations of three central included studies and admitted five. Three extend the agent purchasing audit (Allouah et al., 2026) toward two-sided agentic markets, purchasing operations, and information markets (Alavi and Nozari, 2026; Cao and Hu, 2026; Ventirozos and Shardlow, 2026). One extends the anthropomorphic advertising analysis (Kim et al., 2025) to conversational advertising and consumer welfare (Hermann et al., 2026), and one extends the platform regulation analysis (Bostoen and Krämer, 2026) to liability allocation (Nguyen, 2026). The final corpus contains 140 primary studies; the register preserves the original study identifiers P1 to P141 with P75 recorded in the exclusion log, and Appendix A lists the corpus.

3.7 Data extraction and coding

A structured extraction form captured eight fields for every primary study. They are bibliographic data, research wave (first wave before 2022, second wave from 2022), provenance track (database, backward, or forward snowballing), influence direction, customer journey stage, delegation depth, method family, and a one-line coded finding. Two of these schemes required development rather than adoption, and both were built iteratively. The influence-direction scheme began deductively from the intuitive distinction between agents as marketing targets and agents as marketing channels. It was then refined inductively over the corpus until every study fit exactly one primary direction. The result is six values: marketing to agents, marketing through agents, marketing by agents, consumer delegation and response, market outcomes, and governance and infrastructure. The delegation-depth

ladder began from autonomy levels implicit in the systems literature and stabilized after iteration at five values, D0 to D4, defined in Section 5. The decisive test was that first-wave decision aids, current voice assistants, and mandate-based payment designs all locate unambiguously. Journey stages follow the established purchase-journey stages with a cross-journey value for studies spanning stages (Lemon and Verhoef, 2016). Method families group the recorded study designs into eight values: experiment, design and system, conceptual and methods, survey, legal and policy, formal model, simulation and audit, and field study; the family mapping is released with the coding scheme. The complete scheme with category definitions, the full coded matrix, and the per-study findings are in the replication package; every count, table, and data figure in this article recomputes from that matrix.

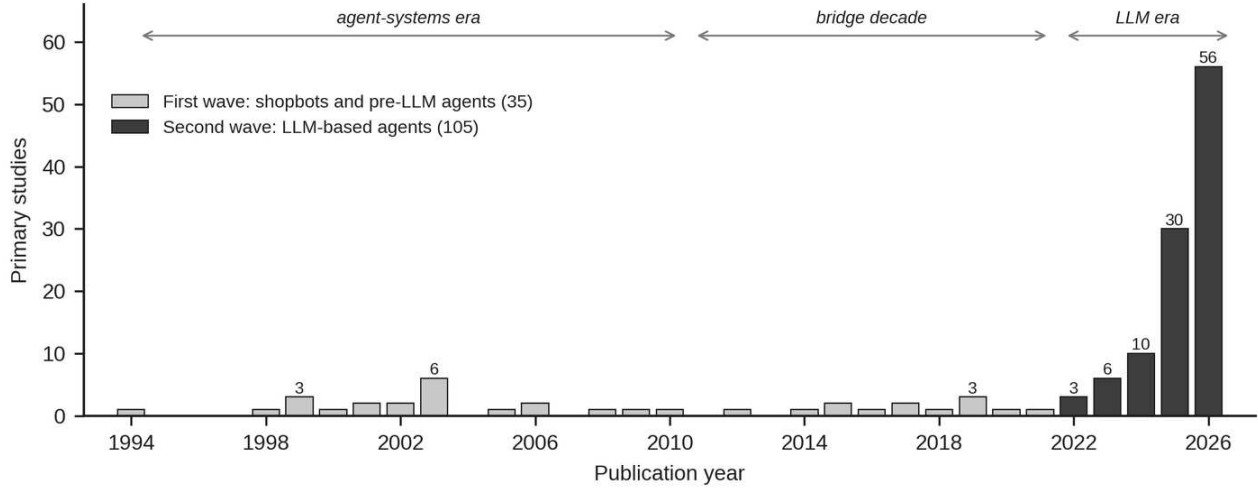
3.8 Quality assurance

Six safeguards address the standard threats. Against retrieval bias, we used three databases with complementary coverage, a validation set fixed before execution and fully recovered, and snowballing in both directions with verified lineage. Snowballing contributed 28 of the 140 studies, and it specifically repairs the vocabulary drift that any fixed search string suffers in a fast-moving field. Against selection bias: paired criteria fixed before screening, a three-author calibration with measured agreement, and criterion-coded decisions preserved record by record. Against extraction error: full-text reading of all included studies, with optical character recognition applied where source files lacked a text layer, and verification of every quantitative claim in this article against the source document before inclusion. Against source-quality distortion: separation of trade press before screening, labeling of working papers, and the credibility screen EC9. Against synthesis opacity: concept-centric writing with the coded register throughout, and figures whose every element derives from the coded matrix. Against irreproducibility: release of the complete decision trail and coded matrix. After the three-author calibration, the remaining screening and the coding were divided between the authors.

4 Corpus overview

4.1 Temporal structure: two waves and the bridge between them

Figure 2 shows the corpus over time, and we read its shape as informative. The first wave contributes 35 studies published between 1994 and 2021, with its agent-systems core concentrated between 1998 and 2010; the second wave contributes 105 studies, of which 96 appeared in 2024 or later and 56 in the first seven months of 2026 alone. Between the systems core and the LLM era lies the bridge decade described in Section 2.1. Its thirteen studies from 2011 to 2021 show agent building receding while four other things accumulate: the psychology of algorithm reliance, the legal category of the algorithmic consumer, residual shopbot market evidence, and the landmark disclosure field experiment (Dietvorst et al., 2015; Logg et al., 2019; Castelo et al., 2019; Gal and Elkin-Koren, 2017; Pathak, 2012; Luo et al., 2019). The temporal structure is therefore, on our reading, not a boom, a silence, and a boom, but a capability wave, a preconditions decade, and a second capability wave that arrives with its behavioral and legal groundwork partly laid. This reading matters for how the old studies are used, and the corpus uses them in three distinct roles. Some are foundational



Extracted from the DataExtraction sheet (yearly counts by wave)

Figure 2: Primary studies per publication year ($N = 140$; 2026 covers January to July). Stacked segments distinguish the first wave of shopbot and pre-LLM agent research from the second, LLM-based wave; the annotated spans mark the agent-systems era, the bridge decade in which delegation psychology and the legal vocabulary matured, and the LLM era.

theory whose constructs the second wave still runs on: the buyer-behavior mapping of agents to journey stages (Guttman et al., 1998), persuasion knowledge (Friestad and Wright, 1994), and the algorithmic consumer as a legal-economic category (Gal and Elkin-Koren, 2017). Some are baseline evidence that no second-wave study has yet matched in ecological validity. The shopbot brand-premium field study (Smith and Brynjolfsson, 2001), the price-panel test of the law of one price at a live comparison agent (Snaith, 2003), and the machine-selling disclosure experiment on a production channel (Luo et al., 2019) remain the field’s principal field results. Early evidence that shopping agents can counter the inaction inertia that follows a missed sale complements them (Andrews, 2008). And some are cautionary mechanism studies whose dynamics recur: seller gaming of comparison agents (Hajaj et al., 2015, 2017) anticipates generative engine optimization, and pricebot price-war cycling (Greenwald and Kephart, 1999) anticipates the algorithmic coordination question (Schlechtinger et al., 2023). Reading the corpus without these studies would make the second wave look unprecedented; reading it with them turns several second-wave results into replications across a technology generation, which is a materially stronger evidential claim.

4.2 Disciplinary and outlet structure

The corpus crosses five literatures, and the crossing points are informative. Marketing and consumer research contribute the largest block, including work in *Journal of Marketing Research*, *Journal of Consumer Psychology*, *Journal of Business Research*, *Psychology and Marketing*, *Journal of Retailing*, *Journal of Retailing and Consumer Services*, *Journal of Advertising Research*, *Journal of Service Research*, and *Journal of Marketing Analytics* (Schultz and Kaiser, 2025). Information systems contributes the delegation-interface and field-experimental work, including *Information*

Systems Research, Decision Support Systems, Electronic Markets, and the major IS conferences. Computer science contributes the audits, benchmarks, attacks, and system designs through KDD, AAMAS, AAAI, and a fast-moving working-paper frontier. Economics and operations contribute the mechanism design and market models, and law contributes the liability, competition, and consumer-protection analyses. The pattern that matters, observed during full-text coding rather than through a formal citation analysis, is asymmetric awareness. The computer science studies in the corpus engage marketing constructs frequently, citing persuasion, brand, and choice, while the marketing studies engage the computational frontier rarely. The two-wave gap of Section 2.1 thus repeats as a two-community gap within the second wave. One consequence is duplicated discovery: nudge sensitivity of agents is established computationally (Allouah et al., 2026) while marketing studies separately establish consumer perception of agentic nudging (Choubey and Chakraborty, 2026), with neither literature citing the other. The framework in Section 9 is designed to make such pairs visible, and the community entry points in Section 9.4 are designed to route them to each other.

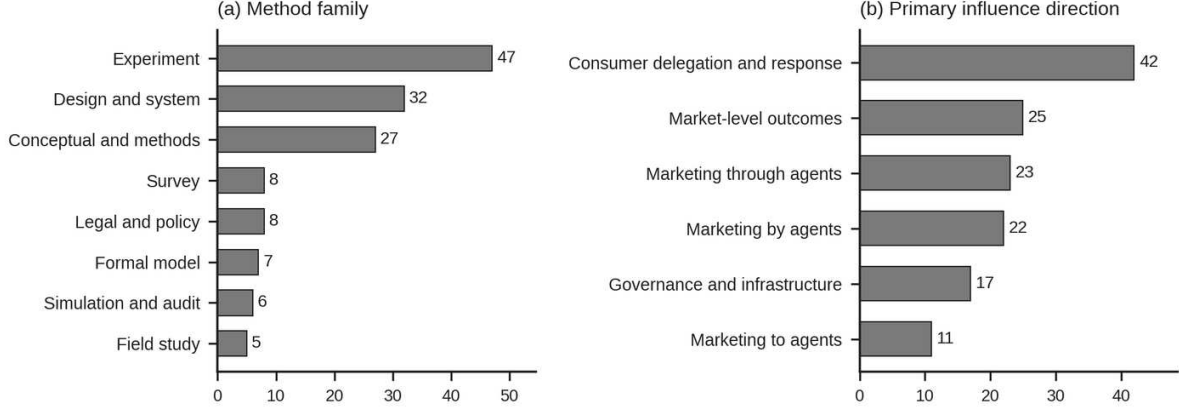
4.3 Methods and the maturity of the evidence

Figure 3 profiles the corpus by method family and by primary influence direction. Experiments dominate with 47 studies, followed by design and system contributions (32), conceptual and methods work (27), surveys (8), legal and policy analyses (8), formal economic models (7), simulation and audit studies (6), and field studies (5). Two readings follow. The optimistic one: the field has passed the purely conceptual stage quickly, and causal designs are the modal contribution, including multi-study packages that combine several experiments with a field component (Shamim et al., 2026; Lei et al., 2025; Liu et al., 2025). The critical one: only five studies have field data or a field experiment as their primary design (Adam et al., 2025; Zhang et al., 2026; Snaith, 2003; Luo et al., 2019; Smith and Brynjolfsson, 2001), two of them a generation old; several experimental packages add a field component (Liu et al., 2025; Lei et al., 2025; Frank et al., 2026), but no study analyzes production transaction data at scale. A reading convention follows from this profile. Throughout the synthesis we name the evidence type behind each claim: an experiment, a field study, a survey, a formal model, a design contribution, or a conceptual analysis. Unless several studies are cited, a finding rests on a single study whose boundary conditions are still open. The consequence is a specific evidential asymmetry that recurs throughout this review: mechanisms are established with experimental confidence, while market consequences rest on models, simulations, and dated field anchors. The asymmetry is not a reason for agnosticism, because the mechanisms are consistent across studies. But they come almost entirely from laboratory and online-panel settings, so their external validity in live markets is largely untested. That is the reason for the measurement agenda of Section 10, which exists to close exactly this gap.

5 RQ1: Actors and the delegation ladder

5.1 Actor families in three institutional roles

The corpus describes commerce-capable agents operated from three institutional roles, on behalf of consumers, of firms, and of platforms, and keeping the roles apart is a precondition for the



Extracted from the DataExtraction sheet (methods and journey stages by wave).

Figure 3: Corpus profile ($N = 140$). Panel (a) reports the eight method families; panel (b) reports the six primary influence directions. Both recompute from the released coding matrix.

influence analysis of Section 6. On the consumer side, the lineage runs from first-wave decision aids, comparison engines, and negotiation marketplaces (Häubl and Trifts, 2000; Yuan, 2003; Karacapilidis and Moraïtis, 2001) through voice assistants whose recommendation behavior and choice framing are now well studied (Flavián et al., 2023; Rohden et al., 2025; Rabassa et al., 2022) to general-purpose assistants and autonomous shopping systems that span the journey (Patil et al., 2026; de Bellis and Johar, 2020; Potluri and Chittimalla, 2026). Specialized designs extend the family: agents for subjective product needs (Dammu et al., 2025), governed basket execution with inventory-aware actions (Liu and Chen, 2026), and deeply personalized super-assistants spanning many life domains (Hasselwander et al., 2026). On the firm side, the corpus documents selling agents entering sales roles (Gonzalez et al., 2026; Luo et al., 2019) and influencers delegating audience replies to social agents (Zhang et al., 2026). A production line of marketing-by-agents sits behind them. It spans campaign and content generation (Aslam et al., 2026; Ahn and Kim, 2025), banner design (Wang et al., 2025), advertising keyword generation (Chen et al., 2025), branding and personalization systems (Puri et al., 2026), agentic traffic optimization (Zhang, 2026), cross-border selection and marketing agents (Xiao et al., 2025), and agentic analytics with interpretability and evaluation layers (Rudra et al., 2026; Rudra and Agrawal, 2025; Lu et al., 2026). On the platform and infrastructure side sit the payment, identity, and authorization designs (Chua et al., 2026; Vaziry et al., 2026; Bhushan et al., 2026; South et al., 2025), trust-aware purchasing frameworks and multi-agent commerce architectures (Gadi et al., 2026; Mishra et al., 2025), and the semantic and security scaffolding of decentralized retail (Bernovskis and Sceulovs, 2026; Pachaiyappan, 2026). A fourth family is synthetic: agents standing in for market participants in research itself, as silicon respondents (Sarstedt et al., 2024; Brand et al., 2023), simulated consumer populations (Chu et al., 2025), and benchmark shoppers (Ling et al., 2026; Wang et al., 2026). The same underlying capability therefore appears in four institutional roles, and a study’s contribution is legible only relative to the role.

Table 3: The delegation ladder. Counts give primary studies whose empirical or design focus sits at each level ($N = 140$; D0 is an anchor level and attracts no coded studies because purely informational interfaces are outside scope).

Level	Name	Definition	Corpus count
D0	Informer	Surfaces information; the consumer performs every task	0 (anchor level)
D1	Advisor	Recommends, ranks, or filters; the consumer decides and executes	61
D2	Co-pilot	Executes sub-tasks interactively under continuous consumer oversight	15
D3	Scoped executor	Completes bounded transactions under an explicit, limited mandate	17
D4	Autonomous delegate	Plans and transacts across the journey; consumer oversight is retrospective	47

5.2 The delegation ladder

Comparing this literature requires a common scale for how much authority the agent holds, which no primary study supplies. We therefore code every study on a five-level delegation ladder, defined in Table 3 and summarized with corpus counts in Figure 4. The ladder’s ordering principle is where consumer oversight sits relative to the act. At D0 and D1, oversight precedes every consequential action; at D2 it is continuous; at D3 it is delimited in advance by a mandate; at D4 it is retrospective. The scale absorbs the corpus without remainder: first-wave decision aids sit at D1 (Häubl and Trifts, 2000), conversational co-shopping and eye-tracked delegation interfaces at D2 (Balaskas et al., 2026; Lee, 2026), mandate-scoped executors and one-click delegation studies at D3 (Choubey and Chakraborty, 2026; Frank et al., 2026; Gadi et al., 2026), and autonomous purchasing designs, audits, and market models at D4 (Allouah et al., 2026; Chua et al., 2026; Cao and Hu, 2026).

The distribution across the ladder is the single most consequential descriptive fact in the corpus, because the two peaks rest on opposite kinds of evidence, the reverse of what one would expect. D1 holds 61 studies, and its evidence is behavioral: 26 experiments, 4 field studies, and 22 working systems. D4 holds 47 studies, but only six of them involve human participants or field data at all; the level is carried by 19 conceptual analyses, 8 legal analyses, 5 formal models, 7 system designs, and the audit work. The middle of the ladder, where products will actually operate first, holds 32 studies. In short, the level we understand behaviorally is the level being automated away, and the level being built is the level we understand least behaviorally. Stated as a finding: the behavioral science of full delegation consists of six empirical studies. Every trust result in Section 7 and every market model in Section 8 should be read against that base rate, and the agenda in Section 10 treats populating D2 to D4 with human evidence as its first priority.

5.3 Reading the waves against each other

The two waves divide the ladder in a way that sharpens the continuity thesis. The first wave contributed 13 of the 61 advisor-level studies and, strikingly, 13 of the 47 autonomous-level studies. Its negotiation protocols, bidding mechanisms, and market designs (Wang et al., 1999; Kehagias

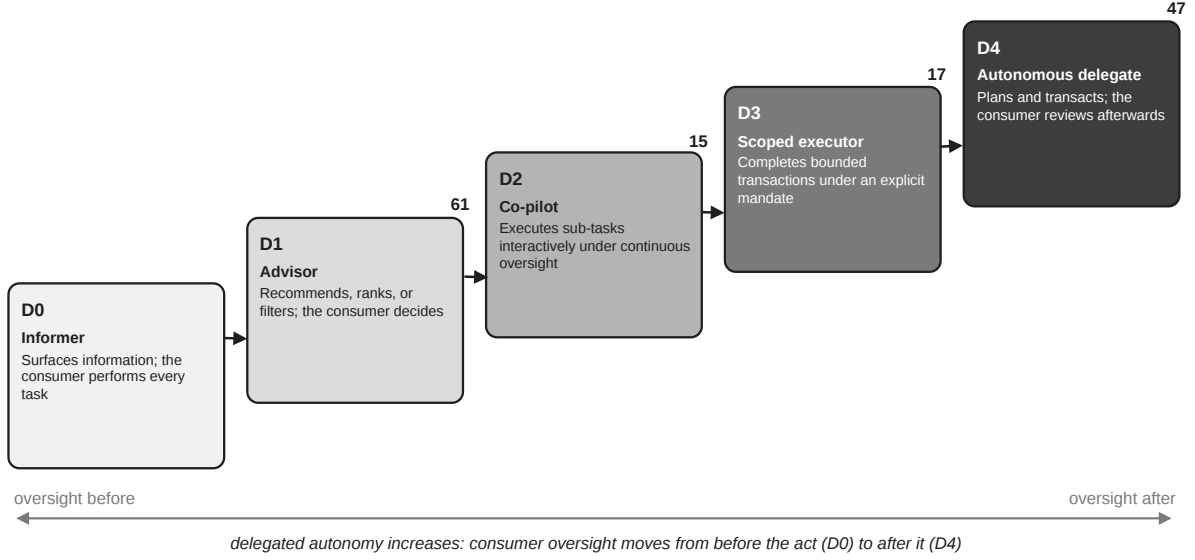


Figure 4: The delegation ladder with corpus counts. Delegated autonomy increases from left to right; the axis beneath marks the movement of consumer oversight from before the act to after it. The distribution is bimodal: evidence concentrates at the advisor level, discourse at the autonomous level, and the operational middle is thin.

et al., 2005; Iyer and Pazgal, 2003) imagined D4 while its deployed systems operated at D1, and its field-defining survey explained the gap institutionally (He et al., 2003). The second wave repeats the pattern at higher capability: its consumer evidence still concentrates at D1 and D2, while its D3 and D4 contributions are dominated by infrastructure designs, audits, and law rather than by consumer behavior (Chua et al., 2026; South et al., 2025; Allouah et al., 2026; Busch, 2025). The gap between what agents can do and what consumers observably let them do has therefore survived the technology shift so far. That persistence is one reason the corpus’s governance layer, the subject of Section 7, treats verifiable mandates rather than raw capability as the binding constraint, echoing the first wave’s own diagnosis that trust and interoperation, not intelligence, gate delegation (He et al., 2003).

6 RQ2: The influence triad

Marketing influence in agentic commerce runs in three directions that the literature routinely conflates: to the agent, through the agent, and by the agent. Figure 5 defines the triad; the subsections synthesize each direction and then the ways the directions combine, which is where the strategically decisive patterns appear.

6.1 Marketing to agents: the machine as target

Consider a merchant whose pages an assistant never cites. In a controlled experiment, rewriting those pages with citations, statistics, and clearer structure raised their presence in generated answers by up to 40 percent (Aggarwal et al., 2024). That is marketing to agents. Eleven studies make the

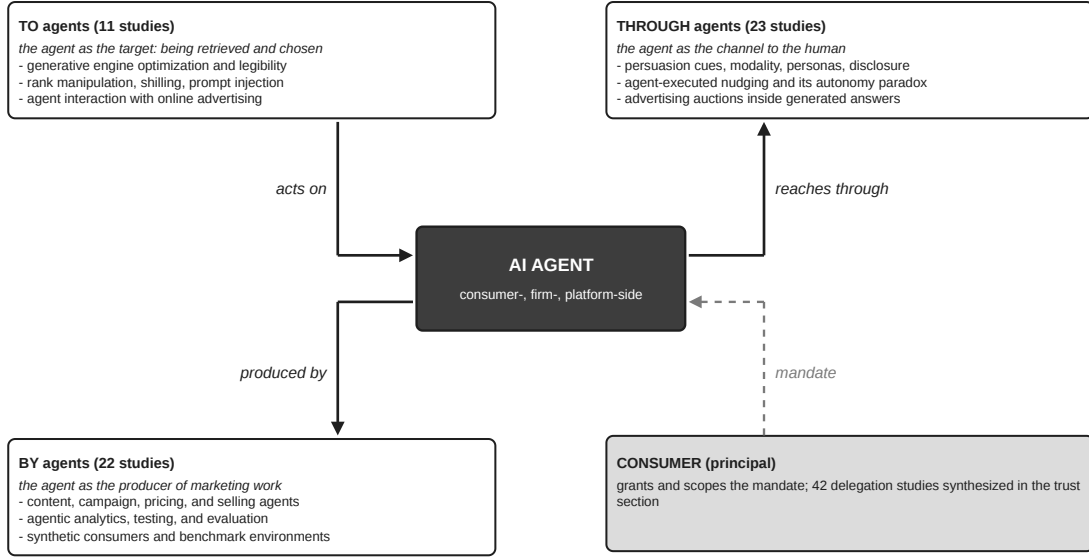


Figure 5: The influence triad. The agent is target (marketing to agents), channel (marketing through agents), and producer (marketing by agents); the consumer grants the mandate that gives the triad its stakes. Counts give primary studies whose principal contribution sits in each direction.

agent itself the marketing target, and together they establish two things: machine visibility can be deliberately manufactured, and the same techniques that make content legible to agents also enable manipulation. Both patterns replicate a first-wave precedent. The foundational result is experimental: generative engine optimization, restructuring content with citations, quotations, statistics, and authority signals, raises a source’s visibility in generated answers by up to 40 percent, with effects varying across domains and query types (Aggarwal et al., 2024). Around that result has grown a practice literature formalizing visibility strategy and monitoring for the new gatekeepers (ANCA, 2026; Uppin and Uppin, 2025). An industrial complement shows that retail catalogs themselves can be made machine-readable at scale, using brand knowledge bases and LLM agents to repair and match product data (Ceker et al., 2026). Observational work adds two cautions. Agents interacting with live web advertising respond to keyword relevance and structured data rather than to visual salience (Stöckl and Nitu, 2025). And measured brand presence in generated recommendations is itself unstable: identical queries return different brands across systems, model versions, and points in time (Kumar and Lakkaraju, 2024; Allouah et al., 2026). The manipulation side is documented with unusual directness. Strategic text sequences inserted in product pages raise a target product’s rank in model recommendations (Kumar and Lakkaraju, 2024). LLM-driven shilling frameworks (coordinated fake identities that promote or demote items) with feedback memory outperform traditional attack baselines against recommender systems (Li et al., 2026a). Advertisement embedding attacks hijack model outputs through service interception or backdoored open-source checkpoints, with a backdoored checkpoint reproducing the attacker’s preset responses in full (Guo et al., 2025). And indirect prompt injection steers autonomous web navigation agents from inside the pages they read (Johnson et al., 2025). The first wave saw the same sequence: as soon as comparison agents concentrated demand, sellers manipulated them through selective disclosure and strategic ordering (Hajaj et al., 2015, 2017). The lesson carried across twenty-five years is that any machine gatekeeper

of demand creates an optimization market on its inputs. The practical difference this time is that the inputs are natural language, so that market is open to every content producer rather than to price-feed engineers alone.

6.2 Marketing through agents: the machine as channel

Consider a shopper who must decide under time pressure: in that condition, identical recommendations persuade more when delivered by an AI agent than by a human, because scrutiny of the machine is suspended (Shamim et al., 2026). That is marketing through agents. Twenty-three studies treat the agent as the channel through which firms reach people, and they connect into a chain that runs from message design to consumer welfare, welfare here meaning realized consumer surplus and autonomy rather than a technical term. At the message end, modality shapes persuasion. In a single experiment series, text-based assistants outperformed voice on concrete product actions, raising product-link requests from 29.0 to 48.1 percent, with mental imagery as the proposed mechanism (Nguyen et al., 2026). Whether the gap holds across product categories and tasks is untested. Conversational cues carry weight machines were not expected to have. Conversational fillers such as a hesitant *hmm*, added to humanize agents, backfire in selling contexts: across one field experiment and four studies, they trigger suspicion of ulterior motives and lower purchase intentions (Liu et al., 2025). Chatbot interfaces themselves shift product and price perceptions (Sidlauskienė et al., 2023), and endorser design shifts ad engagement between human-like and cartoon-like agents (Park et al., 2025). Anthropomorphic agents that visibly use personal data trigger an uncanny valley of mind, the discomfort that arises when a machine appears to read one’s thoughts, in personalized advertising; algorithmic disclosure of data handling mitigates it (Kim et al., 2025). Perceived motives are decisive: perceived benevolence, the inference that the agent serves the customer rather than the company, shapes compliance when ulterior motives are salient (Farmer and Harding, 2026), and persuasion knowledge theory supplies the mechanism (Friestad and Wright, 1994). Timing and pressure matter in ways that flip the human advantage: in checkout charity, AI-led appeals outperform human appeals under high time pressure because they impose lower social demands, across four studies (Shamim et al., 2026). Field evidence scales the channel. In one large-scale field experiment, undisclosed chatbots sold as effectively as proficient human workers and four times more effectively than inexperienced ones, while pre-conversation disclosure of machine identity cut purchase rates by 79.7 percent, with disclosure timing as the documented boundary condition (Luo et al., 2019). Live-stream commerce experiments add a second boundary condition: under high persuasion intensity, AI streamers outperform human streamers because intense appeals from a machine are attributed to programmed execution rather than strategic manipulation, which lowers suspicion and lifts purchase intention (Lu and Li, 2026).

The channel’s manipulative side and its governance are studied with equal energy. Agent-executed nudging raises an autonomy paradox in which consumers perceive both assistance and manipulation, mediated by persuasiveness, trust, and ethical acceptance (Choubey and Chakraborty, 2026). Biased option sets from voice assistants create dependence risks that consumers’ product expertise and purchase habits dampen (Rabassa et al., 2022). Empathic voice assistants can steer choices toward sustainability targets (Mari and Alabed, 2026), and personalization by agents can undermine the very autonomy it serves, the emancipation paradox (Anandan Karichalil and Kaul,

2026). Disclosure design, not merely disclosure presence, emerges as the operative lever: when an AI agent presents an unfavorable personalized price, agent presence softens rejection through perceived-justice and privacy pathways, and environment-based explanations outperform user-based ones (Peng et al., 2023). At the market end of the chain, mechanism design moves advertising inside the generated answer itself. Token auctions aggregate advertiser preferences into a single generated output (Dütting et al., 2024), auctions allocate slots in LLM summaries with incentive guarantees (Dubey et al., 2024), and retrieval-augmented auctions price placement inside answers (Hajiaghayi et al., 2024). Welfare analysis names the stakes of conversational advertising for consumers and platforms (Hermann et al., 2026). The composed chain, message design to motive inference to disclosure architecture to auction, is the through-direction’s contribution, and its weakest empirical link is at the end: the auction layer is formally mature and behaviorally untested.

6.3 Marketing by agents: the machine as producer

Consider an influencer who lets an AI answer follower comments: in a field study, receiving such a reply raised users’ subsequent commenting by about 41 percent (Zhang et al., 2026). That is marketing by agents. Twenty-two studies make agents the producers of marketing work, and the direction is where deployment runs furthest ahead of consumer-facing evidence. Content and campaign production is operational. Design contributions cover autonomous content generation and management for social media marketing (Aslam et al., 2026), multimodal banner design agencies (Wang et al., 2025), self-reflective ad keyword generation (Chen et al., 2025), hyper-personalized branding systems (Puri et al., 2026), and agent-driven marketplace traffic optimization (Zhang, 2026). One deployment study measures engagement differences across Twitter/X, Discord, and Telegram (Ahn and Kim, 2025). Selling itself is being reorganized. Sales scholarship maps how agentic AI redefines selling roles and sales processes (Gonzalez et al., 2026). Field evidence on delegated B2B data curation shows the customer response is task-contingent: algorithmic requesters match or beat human requesters for data enrichment, while customers prefer humans for data reconciliation, where error concerns dominate (Adam et al., 2025). And a difference-in-differences field study of influencer reply delegation finds that receiving an AI reply raises users’ subsequent commenting by about 41 percent, with gains concentrated among loyal followers and attenuated for commercialized influencers (Zhang et al., 2026). An analytics layer is forming around the producers: interpretable multi-agent decision support (Rudra et al., 2026), heuristic-first evaluation frameworks that judge output quality, actionability, and trust (Rudra and Agrawal, 2025), and agentic A/B testing that runs LLM agents through live websites as synthetic participants (Lu et al., 2026). The synthetic-consumer strand supplies both the promise and the sharpest caution. LLMs can generate silicon samples for consumer and market research (Sarstedt et al., 2024; Brand et al., 2023), and multi-agent populations reproduce social patterns such as conformity toward popular establishments (Chu et al., 2025). The caution is that prompt architecture itself induces methodological artifacts: response-order and label biases that differ across models. Moving from GPT-3 to GPT-4 changes the artifacts without removing them, not necessarily for the better (Brucks and Toubia, 2025). Benchmarks give the direction its evaluation substrate: persona-guided shopping simulation (Ling et al., 2026) and intent-grounded real-catalog benchmarks (Wang et al., 2026) standardize what a competent shopping agent means, and strategy-level frameworks map the field’s opportunities and firm-performance stakes (Puvvadi

et al., 2025; Foltean and Hera, 2026; Mirzayi and Talajouran, 2025). The direction’s open flank is the consumer: almost none of the production literature measures how people respond to being marketed to by machine-made campaigns at scale, which the agenda addresses through the disclosure and susceptibility propositions.

6.4 Compositions: where the directions interact

The three directions do not stay separate in deployment; they combine, and three combinations, patterns in which one direction feeds another, organize much of what is strategically novel. The first is legibility-to-persuasion. To-direction work makes offers visible to agents (Aggarwal et al., 2024; Ceker et al., 2026), and through-direction work converts visibility into influence on the human. Firms therefore optimize two coupled funnels, one into the model’s answer and one from the answer into the person, and auction design (Dubey et al., 2024) monetizes the seam between them. The second is the synthetic optimization loop. By-direction producers generate content, agentic testing evaluates it with agent judges (Lu et al., 2026; Rudra and Agrawal, 2025), and to-direction results imply the winning content is whatever agents prefer. The loop can therefore drift toward machine-readable persuasion without any human in the training signal, and the prompt-artifact results (Brucks and Toubia, 2025) say the drift’s direction depends on the model version. The third is proxy manipulation: when the buyer is an agent, influencing the proxy rather than the principal becomes the rational strategy, a framework the corpus states explicitly (Bilgihan et al., 2026) and the attack literature operationalizes (Kumar and Lakkaraju, 2024; Li et al., 2026a; Johnson et al., 2025). All three compositions share a governance implication developed in Section 7: the integrity of agentic commerce depends on the integrity of the agent’s information diet, which no single firm controls.

7 RQ3: Trust, acceptance, and the governance infrastructure

7.1 What determines whether consumers delegate

The delegation direction is the corpus’s largest at 42 studies. We read its findings causally in one direction: the six families in Table 4 are antecedents that shape trust, and trust in turn shapes the willingness to delegate; experienced delegation then feeds back on trust, but that feedback is thinly studied. Signals of the agent’s competence and performance are the first family. Chain-of-thought explanations raise perceived knowledge and behavioral intentions, mediated by trust and privacy concerns (Ebrahimi et al., 2026), and extended acceptance frameworks confirm performance expectancy as a core path (Zhao et al., 2025; Balakrishnan and Dwivedi, 2024). Transparency operates through cognitive trust: algorithmic transparency shows a strong path to cognitive trust, while outcome accountability builds affective trust (Wang and lee, 2026). Eye-tracking evidence adds that the smallest amount of transparency that changes decisions is the amount users can act on. Cues must be actionable and consistent with users’ decision criteria, and comfort concentrates where outcomes are immediately observable and actions reversible (Balaskas et al., 2026; Elkhoudary, 2026). Practitioner-facing treatments name the underlying tension a transparency paradox (Bansal and Bansal, 2026). Motive inference is the third family: benevolence attributions (Farmer and Harding,

2026), ulterior-motive access (Shamim et al., 2026), and persuasion knowledge (Friestad and Wright, 1994) govern whether help reads as service or as selling. Social and affective design is double-edged. Empathic voice design raises consumer responses in voice commerce (Mari et al., 2024), and fluent language embodiment shapes credibility conditional on consumption goals (Schwede et al., 2026). But humanizing fillers can backfire through ulterior-motive suspicion (Liu et al., 2025), anthropomorphism helps only until it starts to feel intrusive, the uncanny region (Kim et al., 2025), and empathic steering of sustainable choices raises its own concerns (Mari and Alabed, 2026). Risk perception, the fifth family, is not a simple deterrent. An inverted-U relationship appears in autonomous shopping: moderate perceived risk sustains engagement by preserving vigilance, while both very low and very high risk suppress it (Wang and lee, 2026). A privacy calculus governs what consumers disclose to agents in the first place (Dey and Kumar, 2026), and perceived benefits, particularly experiential benefits, drive acceptance of high-autonomy assistants in scarcity settings, with behavioral evidence from real ad clicks (Frank et al., 2026). Individual and cultural differences close the list: value dimensions in conversational commerce (Schultz and Kaiser, 2025), acceptance antecedents of super-assistants and latent classes of delegation willingness in the Philippines (Hasselwander et al., 2026; Hasselwander, 2026), millennial attribute preferences (Sharma et al., 2024), and cross-country first-wave comparisons (PILIK et al., 2016).

A note on causal direction before the families. In nearly all of these studies the antecedent is manipulated or measured first, and trust, acceptance, or delegation intention is the outcome, so the documented arrows run from antecedent to trust. Feedback in the other direction, from experienced delegation back onto trust, is essentially unstudied, and the table should be read with that asymmetry in mind.

Beyond the antecedents, the delegation literature contributes three findings that reframe what delegation does to the consumer. Delegation changes downstream behavior, not only choice. Across one field study and four experiments, consumers behave less ethically after encounters with servant-framed agents than partner-framed ones, through anticipatory moral disengagement (Lei et al., 2025). And agents shift intertemporal preferences toward sooner rewards by lengthening perceived delay (Li et al., 2026b). Delegation redistributes cognition: first-wave decision aids already restructured consideration sets (Häubl and Trifts, 2000; Cooke et al., 2002), and current assistants inherit that architecture with larger scope (Vasileva, 2025). And delegation raises identity-level questions the survey and critical literatures make explicit: from consumer sovereignty toward algorithm dependency (Binswanger, 2026), algorithmic consumer culture (Airoldi and Rokka, 2022), and a taxonomy of consumers’ relationships with conversational agents and their wellbeing implications (Alabed et al., 2024). The wave-one comparison again sharpens the picture. The first wave stated its trust problem as social willingness plus institutional readiness (He et al., 2003). The second wave has now measured the social side in detail, while the institutional side has become its own research program, to which we turn.

7.2 Governance and transaction infrastructure

Seventeen studies form the governance direction, and their joint claim is that delegation depth is an institutional variable before it is a psychological one. The design strand builds the mandate. Conversational payment systems pair agent-to-agent protocols with three-mandate verification sep-

Table 4: Six antecedent families of consumer trust and delegation, with representative primary studies and the headline pattern each family contributes.

Antecedent family	Documented effect (antecedent to trust or delegation intention)	Representative studies
Agent competence and performance	Perceived competence raises trust; explanations of reasoning raise perceived knowledge and intentions	Ebrahimi et al., 2026; Zhao et al., 2025; Balakrishnan and Dwivedi, 2024
Transparency and control	Transparency builds cognitive trust when cues are actionable and actions reversible	Wang and lee, 2026; Balaskas et al., 2026; Elkhoudary, 2026
Motive inference	Benevolence attributions and ulterior-motive access decide whether help reads as selling	Farmer and Harding, 2026; Shamim et al., 2026; Friestad and Wright, 1994
Social and affective design	Humanizing cues are double-edged: empathy helps, fillers backfire, and heavy anthropomorphism starts to feel intrusive	Mari et al., 2024; Liu et al., 2025; Kim et al., 2025; Mari and Alabed, 2026
Risk and benefit calculus	Risk shows an inverted-U; experiential benefits drive high-autonomy acceptance	Wang and lee, 2026; Frank et al., 2026; de Bellis and Johar, 2020
Individual and cultural differences	Value dimensions, segments, and demographics stratify delegation willingness	Schultz and Kaiser, 2025; Hasselwander et al., 2026; Hasselwander, 2026; Sharma et al., 2024; PILIK et al., 2016

arating intent, cart, and payment (Chua et al., 2026), and authenticated delegation frameworks bind agent actions to authorizing humans with scoped permissions (South et al., 2025). Blockchain identities and micropayment rails extend agent-to-agent protocols (Vaziry et al., 2026), risk-adaptive authorization scales checks with transaction risk (Bhushan et al., 2026), and trust-aware purchasing architectures demonstrate live marketplace integration (Gadi et al., 2026). Semi-autonomous commerce models, semantic-web retail taxonomies, and security risk analyses complete the stack (Sümer et al., 2025; Bernovskis and Sceulovs, 2026; Pachaiyappan, 2026). The first wave’s secure negotiation protocols are the direct ancestors (Wang et al., 1999), and the industry protocols now deployed (Google, 2025; OpenAI and Stripe, 2025; Visa, 2025; Mastercard, 2025) implement, in production form, what the design strand prototypes.

The legal strand allocates the risk the designs create. Payments law maps the regulatory perimeter of agent-initiated payment and its fraud-detection consequences, including the growing difficulty of distinguishing automated systems presenting human credentials (Amariles et al., 2026). Consumer law confronts doctrines built on human assent: contract formation, information duties, and withdrawal rights all presume a person at the interface, and agent transactions strain each (Busch, 2025; Solanki, 2026). Liability is analyzed as a dynamic allocation among developers, users, and platforms, scaled by the agent’s degree of autonomy and the foreseeability of its conduct (Nguyen, 2026). Competition analysis asks whether the Digital Markets Act’s gatekeeper duties anticipated assistant-layer gatekeeping and finds the fit imperfect (Bostoen and Krämer, 2026). Trademark analysis asks what brand distinctiveness means when the beholder is a model (Amstalden and Schafer, 2023), and anti-manipulation doctrine is tested against autonomous tactics that emerge without

human intent (Brenner, 2026). The strand’s founding text predates the wave: the algorithmic consumer was named, and its competition consequences anticipated, in 2017 (Gal and Elkin-Koren, 2017). Read together, the two strands converge on this review’s central governance thesis: the unit that must be trusted, verified, and litigated is not the agent but the mandate, the authenticated, scoped grant of authority. The same object that makes payment safe (Chua et al., 2026; South et al., 2025) is the object that makes liability assignable (Nguyen, 2026; Busch, 2025) and marketing attribution possible, a convergence the discussion develops into the instrument of mandate-based attribution.

8 RQ4: Market-level consequences

8.1 What the audits show about machine demand

The market direction’s most consequential second-wave evidence is a single large audit of commercially deployed purchasing agents (Allouah et al., 2026). Because it is one study, its findings should be read as the first stylized facts of machine demand rather than settled regularities. Agent choices exhibit homogeneity, concentrating demand on a few modal products while ignoring others entirely. Position biases are large even in text-only interfaces, and they are unstable across providers and model versions, to the point that two successive model generations exhibit nearly opposite position preferences. Agents penalize sponsored tags while rewarding platform endorsements, and their sensitivities to price, ratings, and reviews vary sharply across models. Every one of these facts has a first-wave shadow with the human in the loop: demand concentration at ranked interfaces, position effects, and the survival of brand premia at a shopbot where comparison was nearly free (Smith and Brynjolfsson, 2001). The difference is that the second wave’s chooser is a manufactured artifact whose biases are vendor-set and shift with updates. What used to be consumer psychology becomes platform configuration, and model-version risk becomes a new category of demand risk that brand managers have never had to price.

8.2 Structure, pricing, and coordination

Formal and simulation work extends the audit facts into market structure. Two-sided models of agentic markets identify preference leakage as a structural force. Because a buyer agent’s dialogue must be coherent with its principal’s preferences, sellers can infer willingness to pay nearly one for one from conversation alone, collapsing the information asymmetry that protects consumer surplus; confidentiality instructions are only partly effective (Alavi and Nozari, 2026). Operations models of solicit-then-suggest purchasing derive how agents should balance asking questions against showing products, with solicitation depth and assortment breadth as formal substitutes (Cao and Hu, 2026). Optimal solicitation follows a water-filling rule: the agent keeps asking where uncertainty about the consumer is deepest, until further questions add less than showing products would. Information-market designs price verified product information itself, creating micro-transaction markets for the inputs agents buy on (Ventirozos and Shardlow, 2026). A formal welfare model builds hallucination risk into the market’s basic structure: users differ in how much accuracy matters to them, and agents choose how much costly fact-checking to perform. Market shares and verification effort settle

where the two meet, so sectors in which errors are costly end up with more verification (Iyidogan and Ozkes, 2025). On coordination, reinforcement learning pricing agents in market simulation converge to supra-competitive prices without communicating and even without observing rivals' prices (Schlechtinger et al., 2023), and policy-oriented analyses extend the competition concern to autonomous pricing at market scale (Kydyrbayeva et al., 2026). Both update the first wave's pricebot dynamics, where price-setting agents generated cyclical price wars (Greenwald and Kephart, 1999; Leloup, 2003) and analytical work showed comparison agents virtually co-locating competitors (Iyer and Pazgal, 2003). Interface-control analysis completes the structural picture: as answers replace result lists, control of the assistant layer concentrates gatekeeping power that inherited a search market already at 80 percent desktop and 90 percent mobile concentration (Ortiz-Freuler and Castells, 2026). Macro syntheses of the agent economy (Tan and Huang, 2025), Hayekian doubts about whether delegated agents can transmit dispersed knowledge (McDavid et al., 2026), algorithmic consumer culture (Airoldi and Rokka, 2022), and updated impact assessments from both waves (Redmond, 2002; Ma et al., 2010) round out the direction.

8.3 The continuity thesis

Across RQ4 the two waves keep producing the same shapes: gatekeeper concentration, gaming of the gatekeeper's inputs, persistent dispersion despite comparison, brand value surviving machine mediation, and algorithmic pricing dynamics that surprise their designers. The second wave adds three genuinely new elements: linguistic manipulability of the chooser, vendor-controlled and version-unstable choice weights, and a payment layer that makes full delegation transactable. The continuity is the review's license to use first-wave field evidence as baseline; the three new elements are its research frontier, and the framework that follows assigns each of them a place.

9 The Agentic Commerce Reference Framework

9.1 Construction and architecture

The syntheses of Sections 5 to 8 were built construct by construct; the field also needs to see them together, because findings in each layer condition what is possible in the others. The Agentic Commerce Reference Framework (ACRF), Figure 6, consolidates the coded corpus into four layers crossed by the customer journey. It was constructed bottom-up: the coding scheme's directions group into layers by the question each answers. What must hold for delegation to be possible is L1, the 17 governance and infrastructure studies. Who delegates, and how deeply, is L2, the 42 delegation studies and the ladder. How influence operates is L3, the 56 studies of the influence triad, and what markets do as a result is L4, the 25 market studies. Every block in the figure is populated by identifiable primary studies, which distinguishes the ACRF from stakeholder frameworks assembled by intuition: it is an evidence map first and an architecture second. It can also be tested block by block. If new studies in a block contradict the block's stated pattern, or if a proposed cross-layer dynamic fails to appear where the framework predicts it, the framework is revised or rejected at that point rather than defended as a whole.

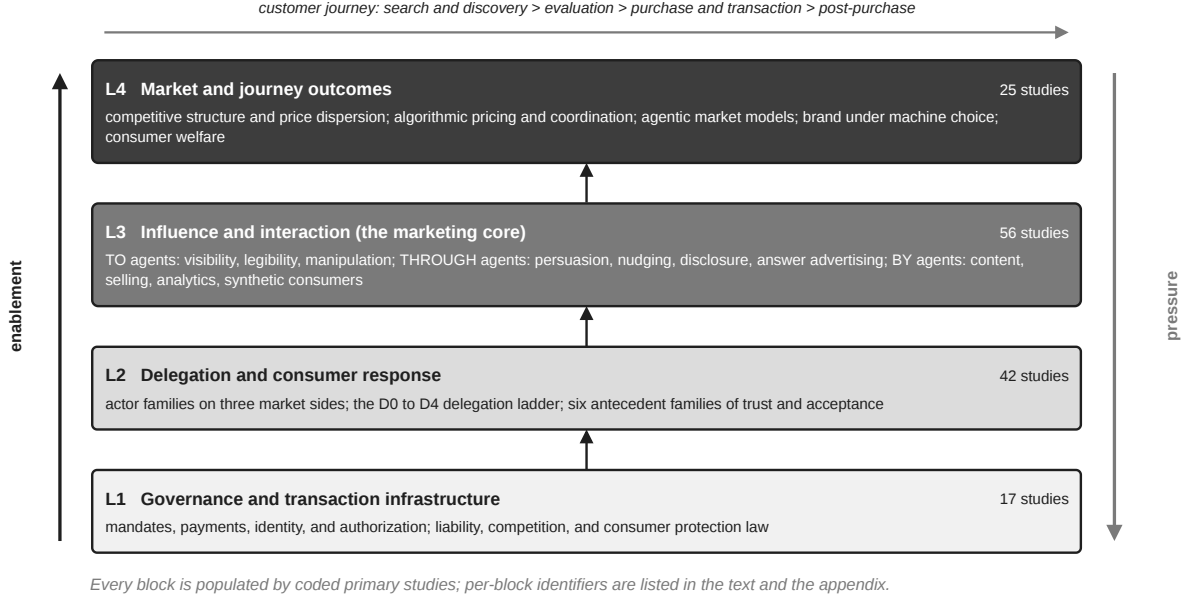


Figure 6: The Agentic Commerce Reference Framework. Four layers organize the coded corpus by the question each answers; the horizontal axis is the customer journey; the side arrows denote enablement running upward and regulatory and design pressure running downward. Study counts per layer sum to the corpus ($N = 140$).

9.2 Cross-layer dynamics

By a dynamic we mean a recurring cause-and-effect pattern that crosses layer boundaries. The four below were derived by reading the coded corpus layer by layer and asking which findings in one layer presuppose or produce findings in another. Each is an analytic generalization over a named cluster of studies, cited with it, not an empirical estimate. They are summarized in Figure 7. D1, enablement, runs upward: infrastructure determines reachable delegation depth (Chua et al., 2026; South et al., 2025), depth determines which influence surfaces exist, and only executed delegation aggregates into machine demand. The first wave illustrates the dependence by what it lacked. Its designs had full-autonomy ambitions, but the mandate, payment, and identity infrastructure they would have needed did not exist, and the systems stalled (He et al., 2003; Wang et al., 1999). D2, pressure, runs downward: market findings such as leakage (Alavi and Nozari, 2026) and audit instability (Allouah et al., 2026) generate the regulatory and design agendas of L1 (Bostoen and Krämer, 2026; Busch, 2025), and consumer findings such as the autonomy paradox (Choubey and Chakraborty, 2026) constrain acceptable influence design. D3, the adversarial diagonal, is the framework’s warning light. An attack executed in L3, a strategic text sequence or an injected instruction (Kumar and Lakkaraju, 2024; Johnson et al., 2025), defeats a trust assumption located in L1 and distorts demand measured in L4, while remaining invisible in L2 because the consumer sees only a confident recommendation. No single-layer research design can even observe this path, which is why the corpus’s attack studies and its consumer studies currently pass each other without contact. D4, the legibility loop, couples L3’s production and selection stages: agents generate marketing content (Aslam et al., 2026; Wang et al., 2025), agentic testing selects what performs (Lu et al., 2026), agent choosers reward machine-

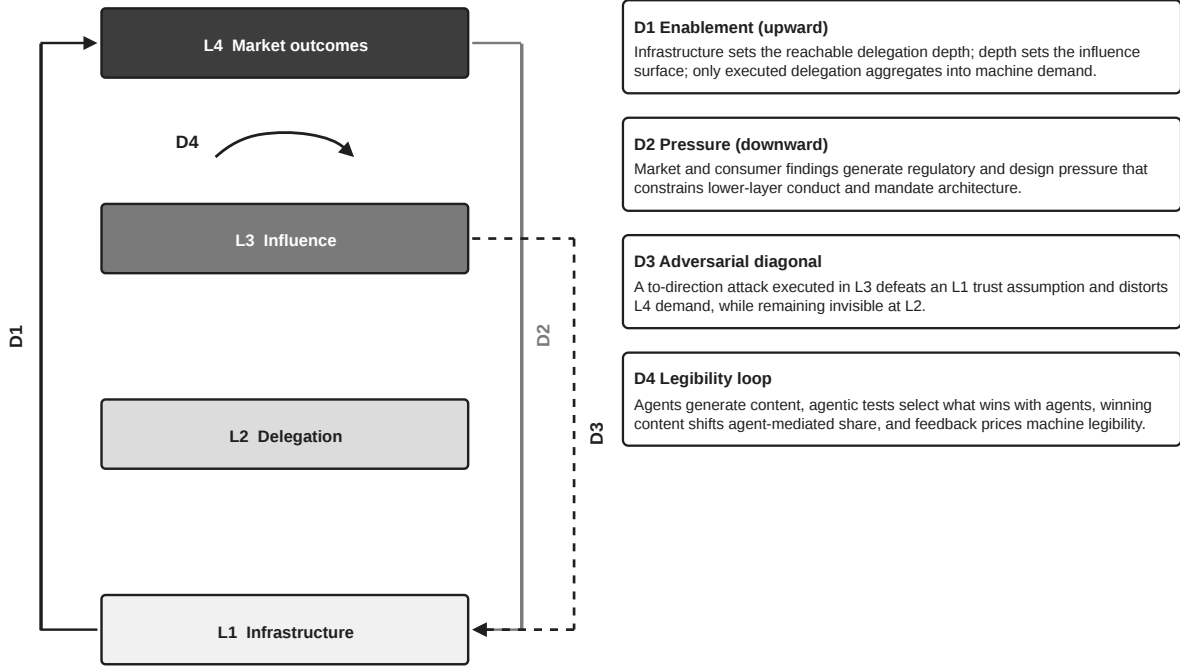


Figure 7: Cross-layer dynamics of the ACRF, recurring patterns of influence between layers stated as analytic generalizations over the coded study clusters cited in the text. D1: enablement runs upward from infrastructure to market outcomes. D2: empirical and regulatory pressure runs downward. D3: the adversarial diagonal, in which an influence-layer attack defeats an infrastructure assumption and distorts market outcomes while remaining invisible at the delegation layer. D4: the legibility loop, in which agent-produced content, agent-based testing, and agent choosers select for machine-legible marketing.

readable content (Aggarwal et al., 2024; Allouah et al., 2026), and rising agent-mediated share feeds the loop. Its end point, marketing optimized for machine readers with humans as bystanders, is exactly the machine-marketing scenario the discipline is being asked to theorize (Sarstedt et al., 2026). Each dynamic is stated so that evidence could contradict it: D1 fails if delegation depth rises in a market whose mandate infrastructure has not changed, D2 fails if documented harms produce no design or regulatory response, and D4 fails if agent-mediated demand stops responding to machine-optimized content.

9.3 The evidence map and its gaps

Crossing the journey axis with the influence directions yields the evidence map of Figure 8, and its empty cells are as informative as its full ones. Evidence concentrates in three regions: cross-journey studies of delegation, production, and governance (67 studies, reflecting that general-purpose agents resist stage assignment); evaluation-stage persuasion and delegation (29); and transaction-stage delegation, market, and governance work (30). Three gaps stand out; calling them consequential is our judgment as reviewers, and others may weight them differently. Post-purchase holds three studies, so returns, complaints, warranty claims, loyalty, and repurchase under delegation are nearly unstudied, despite being where mandate misuse would surface (Andrews, 2008). Search-stage work



Extracted from the DataExtraction sheet (direction by delegation depth, journey-stage shading).

Figure 8: Evidence map: primary studies by customer journey stage and primary influence direction ($N = 140$). Darker cells hold more studies; empty cells are research gaps discussed in the text.

is entirely to-direction: eleven studies of visibility and none of what consumers experience when discovery is delegated. And the to-direction has no post-transaction presence at all: how firms remain visible to an agent that already bought from them, the machine analogue of retention marketing, is an open field. The map converts these gaps into the agenda’s location targets.

9.4 Using the framework across research communities

The ACRF is designed as shared infrastructure for the several communities the corpus spans, and its use differs by community in ways worth stating explicitly, summarized in Table 5. For marketing and analytics scholars, the framework is a variable map: L3 names the treatment space, L2 the moderator space, L4 the outcome space, and L1 the identification opportunities, since protocol rollouts and mandate designs are natural experiments in delegation depth. For consumer research and human-computer interaction, the ladder plus the L2 antecedent families define the design space of oversight. The framework’s D3 diagonal adds the safety-relevant fact that interface trust cues can be true while the layer beneath them is compromised, which makes calibrated rather than maximal trust the design goal (Balaskas et al., 2026; Wang and lee, 2026). For information systems and software engineering, the framework assigns the evaluation agenda: benchmarks and audits (Ling et al., 2026; Wang et al., 2026; Allouah et al., 2026) instrument L3’s chooser, while L1’s designs (Chua et al., 2026; South et al., 2025) need conformance testing, red-teaming against the documented attack classes (Kumar and Lakkaraju, 2024; Johnson et al., 2025; Guo et al., 2025), and telemetry standards that make mandates observable. For economics and operations, L4’s models (Alavi and Nozari, 2026; Cao and Hu, 2026; Dütting et al., 2024) acquire empirical counterparts as L1 telemetry matures, and the framework locates which parameters, leakage rates, position elasticities, concentration indices, are measurable where. For law and policy, the framework separates conduct that existing doctrine reaches from conduct that falls between layers, which is where the corpus’s legal strand concentrates

Table 5: Community entry points into the ACRF. Propositions RP1 to RP10 and instruments I1 to I5 are defined in Section 10.2.

Community	Primary framework use	Agenda entry points
Marketing and analytics	Variable map: L3 treatments, L2 moderators, L4 outcomes, L1 identification	RP1, RP2, RP5, RP6; I1, I2, I5
Consumer research and HCI	Oversight design space on the ladder; calibrated trust under the D3 diagonal	RP4, RP6, RP8; I3
Information systems and software engineering	Evaluation and conformance: benchmarks, audits, red-teaming, mandate telemetry	RP3, RP9; I3, I4
Economics and operations	Parameter estimation for L4 models from L1 telemetry	RP7
Law and policy	Locating conduct between layers; liability and disclosure design	RP8, RP10

(Busch, 2025; Bostoen and Krämer, 2026; Nguyen, 2026). The framework’s ambition is that a study in any of these communities can state, in one sentence, which blocks it occupies and which dynamics it touches; the agenda that follows is written in exactly that grammar.

10 Discussion

10.1 Theoretical implications

Four implications follow from the synthesis, each anchored in the coded evidence rather than in speculation, and each addressed to a named home literature. First, for consumer research and the Persuasion Knowledge Model, persuasion knowledge is displaced twice, and the displacement is measurable. Consumers hold persuasion knowledge about machine persuaders, with benevolence inference and identity disclosure as its levers (Farmer and Harding, 2026; Luo et al., 2019; Kim et al., 2025), while agents themselves lack persuasion knowledge almost entirely, responding to strategic text, injected instructions, and framing (Kumar and Lakkaraju, 2024; Johnson et al., 2025; Allouah et al., 2026). Theory therefore needs a machine-side analogue of Friestad and Wright (1994): a persuasion knowledge model whose target is a model, with attack-recognition playing the role coping tactics play in the human account. The susceptibility parameters are already being measured; what is missing is the construct that organizes them.

Second, for behavioral economics and the choice-architecture tradition, choice architecture becomes two-stage, and the stages obey different laws. Designers arrange context for agents, whose sensitivities are vendor-set and version-unstable (Allouah et al., 2026; Brucks and Toubia, 2025); agents arrange options for humans, whose responses follow the through-direction’s psychology (Nguyen et al., 2026; Choubey and Chakraborty, 2026). Established nudge results do not transfer across the seam: a sponsored tag that attracts a human repels a machine chooser (Allouah et al., 2026). Choice architecture theory needs the seam as an explicit object, including who controls each stage and what happens when the stages conflict.

Third, for information systems and marketing research on technology adoption and delegation, delegation is a mandate relation, not only an attitude, and measurement should follow the construct. The trust literature predicts stated willingness (Wang and lee, 2026; Zhao et al., 2025); the governance literature defines the artifact that actually grants authority (Chua et al., 2026; South et al., 2025); and the ladder’s inversion shows how little behavioral evidence exists where mandates actually bind. Treating the mandate, its scope, its verification, its revocation, as the unit of analysis unifies the psychological and institutional strands and gives delegation a dependent variable that is observable in systems rather than in surveys.

Fourth, for method, the two-wave structure is an identification strategy, not background. Where results replicate across waves, brand premia under machine mediation (Smith and Brynjolfsson, 2001; Allouah et al., 2026), gatekeeper gaming (Hajaj et al., 2015; Aggarwal et al., 2024), algorithmic pricing surprises (Greenwald and Kephart, 1999; Schlechtinger et al., 2023), the mechanism is robust to a complete change of technology, which is the strongest robustness check available in a young field. Where results diverge, linguistic manipulability, version instability, transactable mandates, the divergence isolates what is genuinely new and therefore where new theory is needed. We offer the two-wave design as a template: fields experiencing technology-driven rebirth can treat their own first wave as a control condition rather than as obsolete prologue.

10.2 Research agenda

The framework’s gaps convert into a ten-proposition agenda organized in three streams, measure, model, govern, summarized in Figure 9 with Table 6 stating each proposition and its anchoring evidence. The propositions are anchored in evidence in the sense that each generalizes a documented mechanism; none has been tested in the stated form, and testing them as stated is what the agenda proposes. The measurement stream comes first deliberately, because the constructs of the modeling propositions cannot be estimated without the instruments: the corpus’s central evidential asymmetry, experimental mechanism confidence versus absent production evidence, cannot be closed by more experiments. RP1 calls for share of model voice, a brand’s presence in generated answers and agent choice sets, measured longitudinally with stability diagnostics, because visibility is manufacturable (Aggarwal et al., 2024) and unstable (Allouah et al., 2026). RP2 calls for agent-resolved attribution, separating human sessions, co-pilot sessions, and mandate-executed transactions in analytics pipelines, because the ladder’s levels are causally different customers. RP3 calls for gated agent panels, standing populations of instrumented agents transacting on live markets under research protocols, the field’s least glamorous and highest-value infrastructure need, extending the audit design (Allouah et al., 2026) and the benchmark substrate (Ling et al., 2026; Wang et al., 2026) into continuous measurement. The modeling stream translates the theoretical implications into testable form. RP4, machine persuasion knowledge, predicts that attack susceptibility is a stable, measurable agent trait with a learnable defense curve (Kumar and Lakkaraju, 2024; Li et al., 2026a; Johnson et al., 2025). RP5, two-stage choice architecture, predicts systematic divergence between human-optimal and agent-optimal presentation of identical offers (Allouah et al., 2026; Nguyen et al., 2026). RP6, mandate-level delegation, predicts that scope, verification, and revocation design causally shift delegation depth and downstream satisfaction, testable the moment L1 systems expose mandate parameters (Chua et al., 2026; South et al., 2025; Choubey and Chakraborty, 2026). RP7,

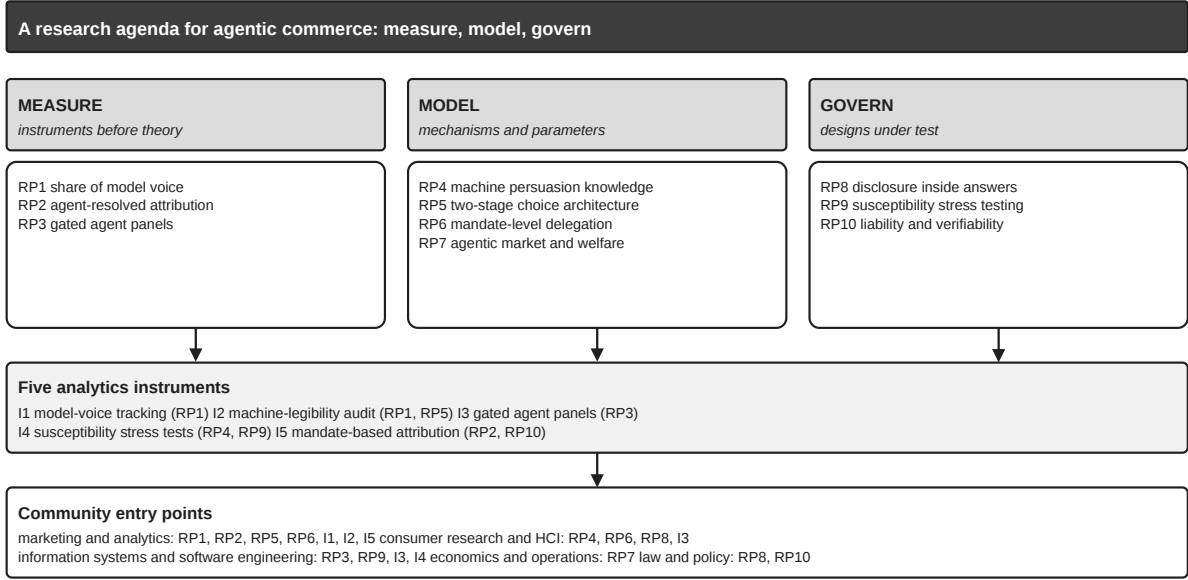


Figure 9: The research agenda: three streams, ten propositions, five analytics instruments, and the community entry points mapped in Table 5.

agentic market equilibrium, calls for empirical estimation of the leakage, solicitation, and concentration parameters the formal models define (Alavi and Nozari, 2026; Cao and Hu, 2026; Iyidogan and Ozkes, 2025). The governance stream tests designs rather than assumes them. RP8, disclosure inside answers, asks what identity and sponsorship disclosure achieve when the persuasion happens inside a generated answer, extending the disclosure evidence (Luo et al., 2019; Peng et al., 2023; Hermann et al., 2026) to the new surface. RP9, susceptibility stress testing, proposes standardized adversarial audits of commercial agents before deployment, merging the attack literature with the evaluation frameworks (Guo et al., 2025; Rudra and Agrawal, 2025). RP10, liability and verifiability, asks which mandate architectures make responsibility assignable in practice, joining the legal allocation analyses (Nguyen, 2026; Busch, 2025) to the verification designs (South et al., 2025; Chua et al., 2026).

10.3 Instruments for marketing analytics

Because the community builds measurement, we state the agenda’s operational core as five instruments. I1, model-voice tracking: standing queries against the major assistant and agent surfaces, logging brand and source presence with version metadata, the successor to share of search (RP1). I2, the machine-legibility audit: a firm-side diagnostic scoring catalog and content machine-readability, structured data coverage, and third-party verifiability, the successor to technical SEO audits (Aggarwal et al., 2024; Ceker et al., 2026) (RP1, RP5). I3, gated agent panels: instrumented agent populations transacting under research governance, the successor to consumer panels (RP3). I4, susceptibility stress tests: standardized adversarial batteries scoring an agent’s resistance to the documented attack classes before it is allowed to spend (Kumar and Lakkaraju, 2024; Guo et al., 2025) (RP4, RP9). I5, mandate-based attribution: analytics keyed to authenticated mandate iden-

Table 6: Research agenda propositions with anchoring evidence.

RP	Proposition	Anchoring evidence
RP1	Share of model voice is measurable, unstable over time, and open to deliberate influence	Aggarwal et al., 2024; Allouah et al., 2026; Uppin and Uppin, 2025
RP2	Agent-resolved attribution separates causally distinct customer types along the ladder	Vasileva, 2025; Chua et al., 2026
RP3	Gated agent panels can generate evidence from live deployments that experiments cannot	Allouah et al., 2026; Ling et al., 2026; Wang et al., 2026
RP4	Machine persuasion knowledge: susceptibility is a stable agent trait with a defense curve	Kumar and Lakkaraju, 2024; Li et al., 2026a; Johnson et al., 2025
RP5	Two-stage choice architecture: human-optimal and agent-optimal presentation diverge	Allouah et al., 2026; Nguyen et al., 2026; Dubey et al., 2024
RP6	Mandate design causally shifts delegation depth and downstream satisfaction	South et al., 2025; Chua et al., 2026; Choubey and Chakraborty, 2026
RP7	Leakage, solicitation, and concentration parameters are empirically estimable	Alavi and Nozari, 2026; Cao and Hu, 2026; Iyidogan and Ozkes, 2025
RP8	Disclosure inside generated answers changes trust and choice differently than interface disclosure	Luo et al., 2019; Peng et al., 2023; Hermann et al., 2026
RP9	Standardized susceptibility stress tests predict field manipulation losses	Guo et al., 2025; Rudra and Agrawal, 2025; Pachaiyappan, 2026
RP10	Mandate verifiability determines practical liability assignability	Nguyen, 2026; Busch, 2025; South et al., 2025

tifiers rather than to sessions, which the payment protocols make technically possible for the first time (Chua et al., 2026; Google, 2025) (RP2, RP10). Each instrument is specified so that a methods paper could build it; we regard the five as the concrete deliverables the machine-marketing agenda (Sarstedt et al., 2026) requires from analytics.

10.4 Managerial implications

For managers the synthesis reduces to four decisions. First, treat machine visibility as a managed asset now. The GEO evidence (Aggarwal et al., 2024) and the audit facts (Allouah et al., 2026) imply that a share of demand is already being allocated by choosers a firm can measure and influence, and the machine-legibility audit (I2) is the entry point. Second, decide the firm’s disclosure policy before regulation decides it: the disclosure penalty (Luo et al., 2019), the design-dependence of acceptance (Peng et al., 2023), and the transparency results (Wang and lee, 2026; Balaskas et al., 2026) say that identity and explanation design are effectiveness variables, not compliance details. Third, segment customers by delegation depth: the ladder’s levels differ in what reaches them, what persuades them, and what retains them, so a D1 advisor user and a D3 mandate holder should not sit in one funnel (RP2). Fourth, price model-version risk: demand allocated by agents can reshuffle on a provider’s

update cycle (Allouah et al., 2026), a volatility channel outside any existing brand-tracking regime, and I1’s version metadata is the safeguard.

10.5 Policy implications

Three implications address regulators and standard-setters. First, the mandate is the regulatory object: verification of authority, scope, and revocation (Chua et al., 2026; South et al., 2025) is where consumer protection, payments regulation (Amariles et al., 2026), and liability (Nguyen, 2026) technically converge, and mandating verifiable delegation would serve all three at once. Second, disclosure rules must follow the persuasion into the answer. Identity disclosure at the interface is well studied (Luo et al., 2019), but sponsorship inside generated recommendations (Dubey et al., 2024; Hermann et al., 2026) is a surface that current advertising disclosure regimes do not reach, and RP8 defines the evidence needed. Third, susceptibility auditing belongs in conformity assessment: the attack literature (Kumar and Lakkaraju, 2024; Guo et al., 2025; Johnson et al., 2025) demonstrates exploitable channels in exactly the systems being wired to payments, and standardized stress testing (I4, RP9) is implementable with existing audit institutions.

11 Limitations

Four limitations bound the claims. First, construct drift: the field’s vocabulary moved during the very window searched, and although snowballing in both directions repairs much of the drift, retrieval anchored partly in current terminology can underrepresent adjacent phrasings; the released search trail lets successors test sensitivity. Second, screening and coding after the three-author calibration: the paired criteria, the measured calibration agreement, the coded decision trail, and recomputable matrix mitigate but do not eliminate individual judgment risk, and an independent recoding of the register is an invited replication. Third, evidence maturity: 96 of the 105 second-wave studies appeared in 2024 or later, a meaningful share as working papers, all labeled in the register. Conclusions there carry the uncertainty of unreviewed work. The peer-reviewed subset supports every headline claim, but effect sizes may settle. Fourth, the corpus’s own evidential asymmetry limits the synthesis symmetrically. With five field studies and no production transaction data, market-level statements rest on models, audits, and dated field anchors. That is why the agenda leads with measurement rather than with more theory. The corpus is also consumer-centric by design. Whether the influence triad, the delegation ladder, and the trust families transfer to business-to-business buying centers, with their formal roles and multi-person decision units, is untested; the two included B2B studies (Adam et al., 2025; Xiao et al., 2025) are too few to say.

12 Conclusion

This review assembled and coded 140 primary studies across two research waves to answer four questions about commerce when the customer can be a machine. The actors form four institutional families whose authority we measure on a five-level delegation ladder. Along that ladder the evidence runs opposite to deployment: behavioral studies are rich where agents merely advise, while only six

empirical studies observe agents that act. Marketing influence runs to, through, and by agents, three directions with different mechanics, different maturities, and compositions, the legibility funnel, the synthetic optimization loop, proxy manipulation, that define the new strategic terrain. Consumer delegation is governed by six antecedent families, but its binding constraint is institutional: the verifiable mandate is where psychology, payments, and liability now converge. And machine demand already shows recurring empirical regularities, concentration, instability, and manipulability, that echo the first wave’s field evidence while adding what is genuinely unprecedented. The Agentic Commerce Reference Framework holds these results in one structure in which every claim lists its supporting studies, the ten propositions and five instruments convert its gaps into work, and the replication package makes every number checkable. Marketing has begun to ask what it becomes when the customer is a machine (Sarstedt et al., 2026). This review’s answer is that the question is researchable now, that two waves of evidence already constrain it, and that the discipline best positioned to lead is the one that builds the instruments first.

Data availability

A replication package accompanies the article¹. It contains the corpus register with provenance for all 141 screened identifiers and the 140 included studies, the criterion-coded screening decisions at every stage, and the full extraction and coding matrix with the coding scheme and method-family mapping. It also holds the search strings and the per-database execution records.

Declarations

Competing interests. The author(s) declare no competing interests.

Funding. No external funding was received for this work.

Ethics. The study synthesizes published literature and involved no human participants.

Use of generative AI. During the preparation of this work, the authors used AI-based tools mainly for proofreading, editing, shortening, rephrasing, and refinement, in line with ARISMA principles (Moghaddam and Alipour, 2026). All the analytical reasoning, methodological choices, and significant contributions presented in this paper are solely those of the authors.

References

- Adam, M., Mishra, A., and Benlian, A. (2025). Agent-based data curation practices: Customer responses to human versus algorithmic data requesters in established business-to-business relationships. *Information Systems Research*.
- Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., and Deshpande, A. (2024). GEO: Generative engine optimization. Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

¹<https://figshare.com/s/af06b7238fc45bed37fb>

- Aggarwal, P. and Vaidyanathan, R. (2003). The perceived effectiveness of virtual shopping agents for search vs. experience goods. *Advances in Consumer Research*, Vol 30.
- Ahn, J. and Kim, M. (2025). Autonomous AI agents for multi-platform social media marketing: A simultaneous deployment study. *Electronics*.
- Airoldi, M. and Rokka, J. (2022). Algorithmic consumer culture. *Consumption Markets & Culture*.
- Alabed, A., Javornik, A., Gregory-Smith, D., and Casey, R. (2024). More than just a chat: a taxonomy of consumers’ relationships with conversational AI agents and their well-being implications. *European Journal of Marketing*.
- Alavi, S. and Nozari, S. (2026). When agents shop for you: Role coherence in AI-mediated markets. arXiv preprint arXiv:2604.26220.
- Allouah, A., Besbes, O., Figueroa, J., Kanoria, Y., and Kumar, A. (2026). What is your AI agent buying? evaluation, biases, model dependence, & emerging implications of agentic e-commerce.
- Amariles, D., Charlotin, D., and He-Guelton, L. (2026). AI agents in payments: Applications, risks and regulations. *European Journal of Risk Regulation*.
- Amstalden, M. and Schafer, B. (2023). When brand distinctiveness is in the ai of the beholder: Trademark law for autonomous intelligent shopping agents.
- Anandan Karichalil, R. and Kaul, D. (2026). The emancipation paradox: When AI-enhanced personalization undermines consumer agency and what socially responsible marketing requires. *Journal of Macromarketing*.
- ANCA, P. (2026). Generative engine optimization (geo): A new paradigm of digital visibility in the age of ai-powered search. *Annals of 'Constantin Brancusi' University of Targu-Jiu. Economy Series / Analele Universității 'Constantin Brâncuși' din Târgu-Jiu Seria Economie*.
- Andrews, D. (2008). MySimon sez: Reducing negative consequences of missed sales via online shopping agents. *Advances in Consumer Research*.
- Aslam, K., Nair, L., Sebastian, R., Seemon, S., and Thomas, A. (2026). Autonomous content generation and management for startup social media marketing using agentic AI.
- Balakrishnan, J. and Dwivedi, Y. (2024). Conversational commerce: entering the next stage of AI-powered digital assistants. *Journal article (early access)*.
- Balaskas, S. (2026). From recommendations to delegation: A systematic review mapping agentic AI in e-commerce and its consumer effects. *Information*, 17(3):222.
- Balaskas, S., Komis, K., Yfantidou, I., and Skandali, D. (2026). When interfaces “act for you”: An eye-tracking experiment on delegation, transparency cues, and trust in agentic shopping assistants.
- Bansal, P. and Bansal, R. (2026). The transparency paradox: Rebuilding trust in AI shopping agents through explainability. Working paper.

- Bernovskis, A. and Sceulovs, D. (2026). The semantic web of retail: A taxonomic integration of Web 3.0, decentralized e-commerce, and agentic commerce. *Journal of Risk & Financial Management*.
- Bhushan, B., Pappu, K., Jadav, V., and Jaiswal, N. (2026). Risk-adaptive authorization for agentic AI systems.
- Bilgihan, A., Lorenz, M., Zhang, Y., and Ostinelli, M. (2026). Persuading the proxy: a framework for AI-mediated marketing decisions. *International Journal of Contemporary Hospitality Management*.
- Binswanger, M. (2026). From consumer sovereignty to algorithm-dependency? the case of shopping agents. *Society 5.0, 2025*.
- Bostoen, F. and Krämer, J. (2026). How future-proof is the dma? a case study of ai agents. *Journal of Competition Law & Economics*.
- Brand, J., Israeli, A., and Ngwe, D. (2023). Using LLMs for market research. Harvard Business School Working Paper 23-062.
- Brenner, Z. (2026). The market dreams of machines: Anti-manipulation regulation, autonomous tactics, and the age of ai agentics. *UC Davis Business Law Journal*.
- Brucks, M. and Toubia, O. (2025). Prompt architecture induces methodological artifacts in large language models. *PLoS ONE*.
- Busch, C. (2025). Consumer law for AI agents. *German Law Journal*.
- Cao, S. and Hu, M. (2026). A solicit-then-suggest model of agentic purchasing. arXiv preprint arXiv:2603.20972.
- Castelo, N., Bos, M., and Lehmann, D. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*.
- Ceker, H., Luo, G., Koo, K., Mathur, P., You, W., Amdekar, A., Barton, R., Kl, N., Bansal, V., and Bouyarmane, K. (2026). Using brand knowledge bases and LLM agents to enhance e-commerce retailers' catalog quality. *Industry-academic short paper*.
- Chattaraman, V., Kwon, W., Gilbert, J., and Li, Y. (2014). Virtual shopping agents persona effects for older users. *Journal of Research in Interactive Marketing*.
- Chen, B., Wang, Z., and Takamatsu, S. (2025). OMS: On-the-fly, multi-objective, self-reflective ad keyword generation via LLM agent. LLM Agents in E-commerce Workshop paper.
- Choubey, V. and Chakraborty, D. (2026). How do consumers perceive digital nudging by agentic AI in e-commerce? *Journal of Retailing and Consumer Services*.
- Chu, M.-L., Terhorst, L., Reed, K., Ni, T., Chen, W., and Lin, R. (2025). LLM-based multi-agent system for simulating and analyzing marketing and consumer behavior. Workshop paper (arXiv).

- Chua, J., Tse, C., Huang, D., and Wang, Z. (2026). ConvPayMAS: Conversational payment multi-agent system with agent-to-agent protocol and three-mandate verification.
- Cooke, A., Sujan, H., Sujan, M., and Weitz, B. (2002). Marketing the unfamiliar: The role of context and item-specific information in electronic agent recommendations. *Journal of Marketing Research*.
- Dammu, P., Alonso, O., and Poblete, B. (2025). A shopping agent for addressing subjective product needs. PROCEEDINGS OF THE EIGHTEENTH ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, WSDM 2025.
- Davenport, T., Guha, A., Grewal, D., and Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48:24–42.
- de Bellis, E. and Johar, G. (2020). Autonomous shopping systems: Identifying and overcoming barriers to consumer adoption. *Journal of Retailing*.
- Dey, S. and Kumar, S. (2026). Privacy calculus and disclosure in LLMs: The role of privacy controls, nudges, and perceived anthropomorphism. *International Journal of Information Management*.
- Dietvorst, B., Simmons, J., and Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*.
- Dubey, A., Feng, Z., Kidambi, R., Mehta, A., and Wang, D. (2024). Auctions with LLM summaries. Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.
- Dütting, P., Mirrokni, V., Paes Leme, R., Xu, H., and Zuo, S. (2024). Mechanism design for large language models. Proceedings of the ACM Web Conference 2024.
- Ebrahimi, P., Hoffmann, S., and Schneider, J. (2026). Opening the black box: How reasoning-enabled AI agents influence user perceptions and behavior in sustainable consumption. *International Journal of Information Management*.
- Elkhoudary, A. (2026). Trust in agentic AI: Transparency and autonomy. SSRN Working Paper.
- Farmer, M. and Harding, R. (2026). Artificial persuasion: profit motives and the persuasiveness of AI sales agents. *Journal of Personal Selling & Sales Management*.
- Flavián, C., Akdim, K., and Casaló, L. V. (2023). Effects of voice assistant recommendations on consumer behavior. *Psychology & Marketing*.
- Foltean, F. and Hera, O. (2026). Agentic AI in marketing: Opportunities, challenges and impact on firm performance. Book chapter.
- Frank, D., Folwarczny, M., and Otterbring, T. (2026). Consumer acceptance of high-autonomy AI assistants is driven by perceived benefits in online shopping settings characterized by scarcity. *Psychology & Marketing*.
- Friestad, M. and Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*.

- Gadi, K., Solleti, S., Guntupalli, K., and Raja, A. (2026). Agentic purchase: A trust-aware multi-agent framework for autonomous e-commerce with live marketplace integration. 2026 IEEE World AI IoT Congress (AIIoT).
- Gal, M. and Elkin-Koren, N. (2017). Algorithmic consumers. *Harvard Journal of Law and Technology*.
- Gonzalez, G., Habel, J., and Hunter, G. (2026). AI agents, agentic AI, and the future of sales. *Journal of Business Research*.
- Google (2025). Agent Payments Protocol (AP2): An open protocol for agent-led payments. Technical documentation, Google Cloud, accessed 28 July 2026.
- Greenwald, A. and Kephart, J. (1999). Shopbots and pricebots. IJCAI-99: PROCEEDINGS OF THE SIXTEENTH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, VOLS 1 & 2.
- Grewal, D. and Lindsey-Mullikin, J. (2006). The moderating role of the price frame on the effects of price range and the number of competitors on consumers’ search intentions. *Journal of the Academy of Marketing Science*.
- Guo, Q., Tang, J., and Huang, X. (2025). Attacking LLMs and AI agents: Advertisement embedding attacks against large language models. Working paper (arXiv).
- Guttman, R., Moukas, A., and Maes, P. (1998). Agent-mediated electronic commerce: A survey. *Knowledge Engineering Review*.
- Hajaj, C., Hazon, N., and Sarne, D. (2015). Improving comparison shopping agents’ competence through selective price disclosure. *Electronic Commerce Research and Applications*.
- Hajaj, C., Hazon, N., and Sarne, D. (2017). Enhancing comparison shopping agents through ordering and gradual information disclosure. *Autonomous Agents and Multi-agent Systems*.
- Hajiaghayi, M., Lahaie, S., Rezaei, K., and Shin, S. (2024). Ad auctions for LLMs via retrieval augmented generation. *Advances in Neural Information Processing Systems*.
- Hasselwander, M. (2026). AI agent, take over?! task delegation to agentic AI systems in the Philippines. *Ai & Society*.
- Hasselwander, M., Sunio, V., Lah, O., and Mogaji, E. (2026). Toward agentic AI: User acceptance of a deeply personalized AI super assistant (AISA). *Journal article (early access)*.
- Häubl, G. and Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science*.
- He, M., Jennings, N., and Leung, H.-F. (2003). On agent-mediated electronic commerce. *IEEE Transactions on Knowledge and Data Engineering*.
- Hermann, E., Puntoni, S., and Schweidel, D. (2026). From advice to advertising: Conversational AI advertising, platform monetization, and consumer welfare. Working paper (June 2026 version).

- Hoffman, D. and Novak, T. (2018). Consumer and object experience in the Internet of things: An assemblage theory approach. *Journal of Consumer Research*.
- Huang, M.-H. and Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49:30–50.
- Ishida, T., Murakami, Y., Lin, D., and Lhaksmana, K. (2025). AI agents: From concept to code to commerce. 2025 International Conference on Information and Communication Technology (ICoICT).
- Iyer, G. and Pazgal, A. (2003). Internet shopping agents: Virtual co-location and competition. *Marketing Science*.
- Iyidogan, E. and Ozkes, A. (2025). Agentic AI and hallucinations. *Economics Letters*.
- Johnson, S., Pham, V., and Le, T. (2025). The dangers of indirect prompt injection attacks on LLM-based autonomous web navigation agents: A demonstration. Demonstration paper (arXiv).
- Karacapilidis, N. and Moraïtis, P. (2001). Building an agent-mediated electronic commerce system with decision analysis features. *Decision Support Systems*.
- Kehagias, D. D., Symeonidis, A. L., and Mitkas, P. A. (2005). Designing pricing mechanisms for autonomous agents based on bid-forecasting. *Electronic Markets*.
- Kim, W., Ryoo, Y., and Choi, Y. (2025). That uncanny valley of mind: when anthropomorphic AI agents disrupt personalized advertising. *International Journal of Advertising*.
- Kumar, A. and Lakkaraju, H. (2024). Manipulating large language models to increase product visibility. arXiv preprint arXiv:2404.07981.
- Kydyrbayeva, E., Shomsheikova, B., Abzhakov, A., Ashimova, A., and Nurtayeva, A. (2026). Autonomous AI agents in digital markets: Economic implications for competition, pricing, and regulation.
- Lee, S. (2026). Meaning over quantity: An exploratory study on how relationship intimacy shapes giver query formulation in LLM-based gift recommendation. CHI LBW / workshop paper.
- Lei, S., Xie, L., and Peng, J. (2025). Unethical consumer behavior following artificial intelligence agent encounters: The differential effect of AI agent roles and its boundary conditions. *Journal of Service Research*.
- Leloup, B. (2003). Pricing with local interactions on agent-based electronic marketplaces. *Electronic Commerce Research & Applications*.
- Lemon, K. N. and Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6):69–96.
- Li, C., Wang, Z., Ma, H., Li, W., Wu, H., Song, Y., and Gao, M. (2026a). AgentAttack : LLM agents for multi-strategy shilling attacks in recommenders. *Expert Systems With Applications*.

- Li, Y., Lin, S., Gong, H., Wang, X., and Janiszewski, C. (2026b). Time is shrinking in the eye of AI: AI agents influence intertemporal choice. *Journal of Consumer Psychology*.
- Lim, W. M., Kumar, S., Verma, S., and Chaturvedi, R. (2022). Alexa, what do we know about conversational commerce? Insights from a systematic literature review. *Psychology and Marketing*, 39(6):1129–1155.
- Ling, E. C., Tussyadiah, I., Tuomi, A., Stienmetz, J., and Ioannou, A. (2021). Factors influencing users’ adoption and use of conversational agents: A systematic review. *Psychology and Marketing*, 38(7):1031–1051.
- Ling, Y., Yuan, C., Dong, S., Yang, Y., Mocherla, N., and Goyal, A. (2026). ShopperBench: A benchmark for personalized shopping with persona-guided simulation. Working paper (arXiv).
- Liu, G., Liu, M., and Zhu, Q. (2025). Hmm, the effect of AI conversational fillers on consumer purchase intentions. *Marketing Letters*.
- Liu, X. and Chen, X. (2026). Governed agentic AI for retail baskets: A consumer world model with inventory-aware actions. *Early access journal article*.
- Logg, J., Minson, J., and Moore, D. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*.
- Longoni, C. and Cian, L. (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The word-of-machine effect. *Journal of Marketing*.
- Lu, Y., Hsu, T.-Y., Gu, H., Cui, L., Xie, Y., Headden W.P., I., Yao, B., Veeragouni, A., Liu, J., Nag, S., Wang, J., and Wang, D. (2026). Agent a/b: Automated and scalable a/b testing on live websites with interactive LLM agents. CHI 2025 workshop paper.
- Lu, Y. and Li, G. (2026). AI vs. human streamers: How digital agents shape consumer persuasion processing in live streaming commerce. *Journal of Theoretical & Applied Electronic Commerce Research*.
- Luo, X., Tong, S., Fang, Z., and Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*.
- Ma, Z., Liao, K., and Lee, J. (2010). Examining comparative shopping agents from two types of search results. *Information Systems Management*.
- Maes, P., Guttman, R., and Moukas, A. (1999). Agents that buy and sell. *Communications of the ACM*.
- Mari, A. and Alabed, A. (2026). AI with a conscience: Empathic voice assistants for sustainable shopping decisions. Working paper.
- Mari, A., Mandelli, A., and Algesheimer, R. (2024). Empathic voice assistants: Enhancing consumer responses in voice commerce. *Journal of Business Research*.

- Mastercard (2025). Mastercard Agent Pay: Pioneering agentic payments technology. Technical announcement, Mastercard, accessed 28 July 2026.
- McDavid, B., Kiesling, L., and Chassin, D. (2026). Markets, agency, and trust: AI agents and the knowledge problem. *Review of Austrian Economics*.
- Mirzayi, S. and Talajouran, E. (2025). From decision support to decision making: Autonomous AI agents shaping the marketing. 2025 11th International Conference on Web Research (ICWR).
- Mishra, M., Vyas, C., Mazumdar, B., and Mishra, V. (2025). Harnessing the power of machine learning in multi-agent systems for B2C e-commerce optimization. 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI).
- Moghaddam, M. T. and Alipour, M. (2026). Arisma: Guidelines for ai-and llm-assisted systematic reviews, scoping reviews, and mapping studies. *arXiv preprint arXiv:2608.25050*.
- Nguyen, M., Barari, M., and Zhang, Y. (2026). The voice or the script? unlocking AI agent persuasion through modality and mental imagery. *Journal of Retailing and Consumer Services*.
- Nguyen, T. (2026). Who bears the responsibility? legal liability allocation for AI agent conduct in platform economies. *Cogent Social Sciences*.
- OpenAI and Stripe (2025). Agentic Commerce Protocol: Specification for programmatic checkout between AI agents and merchants. Technical specification, accessed 28 July 2026.
- Ortiz-Freuler, J. and Castells, M. (2026). From search engines to AI agents: Interface control and the restructuring of communication power. *New Media & Society*.
- Pachaiyappan, N. (2026). Agentic commerce: A comprehensive analysis of cybersecurity risks, privacy challenges, and trust mechanisms in autonomous AI-driven marketplaces. Working paper.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Bmj*, 372:n71.
- Park, S., Chu, Y., and Chung, J.-H. (2025). Designing AI agents for new product endorsement: Do human-like or cartoon-like AI-generated endorsers evoke more positive ad engagement from consumers? *Journal of Advertising Research*.
- Pathak, B. (2010). A survey of the comparison shopping agent-based decision support systems. *Journal of Electronic Commerce Research*, 11(3).
- Pathak, B. K. (2012). Comparison shopping agents and online price dispersion: A search cost based explanation. *Journal of Theoretical & Applied Electronic Commerce Research*.
- Patil, R., Rice, D., and Janiszewski, C. (2026). Human–AI partnerships: Living and working with AI assistants, AI agents, and AI companions. *Journal of Consumer Psychology*.

- Paul, J., Lim, W. M., O’Cass, A., Hao, A. W., and Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies*, 45(4):O1–O16.
- Peng, X., Peng, X., and Xu, D. (2023). Why users accept discriminatory pricing: The roles of AI agent’s presence and explanation. Working paper.
- PILIK, M., KLIMEK, P., JURICKOVA, E., and PALKA, P. (2016). Comparison shopping agents and Czech online customers’ shopping behaviour. *International Journal of Entrepreneurial Knowledge*.
- Potluri, L. and Chittimalla, S. (2026). Multi-agent retrieval-augmented conversational systems for end-to-end autonomous shopping assistants. Working paper.
- Puntoni, S., Reczek, R., Giesler, M., and Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing*.
- Puri, S., Pradhan, J., Bernabe, B., and Ray, A. (2026). Agentic AI in branding: Driving hyper-personalization and real-time adaptation. *Early access journal article*.
- Puvvadi, M., Arava, S., Santoria, A., Chennupati, S., and Puvvadi, H. (2025). Survey of marketing agents, agent-based models, and generative AI in marketing. 2025 Global Conference in Emerging Technology (GINOTECH).
- Rabassa, V., Sabri, O., and Spaletta, C. (2022). Conversational commerce: Do biased choices offered by voice assistants’ technology constrain its appropriation? *Technological Forecasting and Social Change*.
- Redmond, W. H. (2002). The potential impact of artificial shopping agents in e-commerce markets. *Journal of Interactive Marketing*.
- Rohden, S. F., Espartel, L. B., and Lopes, J. C. (2025). From voice to choice: Understanding the consumer response to voice assistant and human recommendations. *International Journal of Consumer Studies*.
- Rudra, A. and Agrawal, M. (2025). A heuristic-first evaluation framework for marketing AI agents: Judging quality, actionability, and trust. Working paper.
- Rudra, A., Karumuri, M., and Agrawal, M. (2026). Explainable AI for marketing intelligence: Interpretable multi-agent systems for data-driven decision support. Working paper (SSRN).
- Sarstedt, M., Adler, S., Rau, L., and Schmitt, B. (2024). Using large language models to generate silicon samples in consumer and marketing research: Challenges, opportunities, and guidelines. *Psychology and Marketing*.
- Sarstedt, M., Adler, S. J., and Imschloss, M. (2026). Machine marketing: Rethinking the customer in the age of generative AI. *Journal of Marketing Analytics*.

- Schlechtinger, M., Kosack, D., Paulheim, H., Fetzer, T., and Krause, F. (2023). The price of algorithmic pricing: Investigating collusion in a market simulation with AI agents.
- Schultz, C. and Kaiser, S. (2025). Consumer value dimensions in conversational and mobile commerce. *Journal article (early access)*.
- Schumacher, K., Roberts, R., and Giebel, K. (2025). The agentic commerce opportunity: How AI agents are ushering in a new era for consumers and merchants. McKinsey & Company report, October 2025, accessed 28 July 2026.
- Schwede, M., Hammerschmidt, M., and Morana, S. (2026). Fluent voice, credible choice – how language embodiment and consumption goals shape voice commerce success. *Journal article (early access)*.
- Sen, R., King, R., and Shaw, M. (2006). Buyers’ choice of online search strategy and its managerial implications. *Journal of Management Information Systems*.
- Shamim, I., Ghosh, T., and Sreejesh, S. (2026). Human versus AI in checkout charity: How time pressure, persuasion knowledge, and ulterior motive shape donation and store patronage intentions. *Journal of Consumer Behaviour*.
- Sharma, M., Joshi, S., Luthra, S., and Kumar, A. (2024). Impact of digital assistant attributes on millennials’ purchasing intentions: A multi-group analysis using PLS-SEM, artificial neural network and fsQCA.
- Sidlauskiene, J., Joye, Y., and Auruskeviciene, V. (2023). AI-based chatbots in conversational commerce and their effects on product and price perceptions. *Electronic Markets*.
- Smith, M. and Brynjolfsson, E. (2001). Consumer decision-making at an Internet shopbot: Brand still matters. *Journal of Industrial Economics*.
- Snaith, S. M. (2003). Can a shopping agent enforce the law of one price? *Atlantic Economic Journal*.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104:333–339.
- Solanki, S. (2026). Algorithms against manipulation: Safeguarding consumer rights in AI shopping. Book chapter (Algorithms Against Manipulation).
- South, T., Marro, S., Hardjono, T., Mahari, R., Whitney, C., Greenwood, D., Chan, A., and Pentland, A. (2025). Authenticated delegation and authorized AI agents. arXiv preprint arXiv:2501.09674.
- Srinivasa Raja, S. V., Orazi, D. C., Cheng, Y., and Belli, A. (2026). A meta-analysis of the effects of AI agent implementation on customer-level outcomes. *Journal of Business Research*, 210:116172.
- Stöckl, A. and Nitu, J. (2025). Are AI agents interacting with online ads? Working paper.

- Sümer, M., Özsoy, T., Okay, F., and Kök, I. (2025). Smart agents as customers: A semi-autonomous digital commerce model for Industry 5.0. 2025 7TH INTERNATIONAL CONGRESS ON HUMAN-COMPUTER INTERACTION, OPTIMIZATION AND ROBOTIC APPLICATIONS, ICHORA.
- Tan, L. and Huang, K. (2025). The AI agent economy. Book chapter (The AI Agent Economy).
- Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.
- Uppin, U. and Uppin, V. (2025). Boosting brand recognition within large language models and AI-driven search systems. Working paper.
- Vasileva, S. (2025). The impact of AI along the customer journey mapping: How AI agents are changing customer journey. PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON BUSINESS EXCELLENCE.
- Vaziry, A., Garzon, S., and Küpper, A. (2026). Enhancing the A2A protocol with blockchain-based identities and x402 micropayments for agentic AI. *Computational Intelligence, Ijcci 2025, Pt I*.
- Ventirozos, F. and Shardlow, M. (2026). Paying to know: Micro-transaction markets for verified product information in agentic e-commerce. arXiv preprint arXiv:2606.24783.
- Visa (2025). Trusted Agent Protocol: An ecommerce framework for agentic commerce. Technical announcement, Visa, accessed 28 July 2026.
- Wan, Y., Menon, S., and Ramaprasad, A. (2009). The paradoxical nature of electronic decision aids on comparison-shopping: The experiments and analysis. *Journal of Theoretical & Applied Electronic Commerce Research*.
- Wang, H., Shimose, Y., and Takamatsu, S. (2025). BannerAgency: Advertising banner design with multimodal LLM agents. PROCEEDINGS OF THE 2025 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING.
- Wang, J. and lee, J. (2026). Letting the bot decide: Trust and risk perceptions in autonomous retail purchase decisions. *Journal of Retailing and Consumer Services*.
- Wang, J., Xiao, K., Sun, Q., Zhao, H., Luo, T., Zhang, J., and Zeng, X. (2026). ShoppingBench: A real-world intent-grounded shopping benchmark for LLM-based agents. FORTIETH AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, AAAI-26, VOL 40 NO 39.
- Wang, X., Yi, X., Lam, K., Zhang, C., and Okamoto, E. (1999). Secure agent-mediated auctionlike negotiation protocol for internet retail commerce. COOPERATIVE INFORMATION AGENTS III, PROCEEDINGS.
- Webster, J. and Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2):xiii–xxiii.

- Xiao, G., You, W., Fan, C., Hu, Y., and Tseng, S.-P. (2025). Design and implementation of an AI agent for product selection and intelligent marketing in cross-border e-commerce. 2025 International Conference proceedings.
- Yuan, S. (2003). A personalized and integrative comparison-shopping engine and its applications. *Decision Support Systems*.
- Zhang, M. M., Gao, Y., Li, J., and Johnson, S. L. (2026). When influencers delegate replies: How social AI agents shape user engagement. *Information Systems Research (INFORMS)*.
- Zhang, W. (2026). AI-driven e-commerce traffic engine: A multi-agent system for optimizing Amazon sales. Working paper.
- Zhao, X., You, W., Zheng, Z., Shi, S., Lu, Y., and Sun, L. (2025). How do consumers trust and accept AI agents? an extended theoretical framework and empirical evidence. *Behavioral Sciences*.

A The corpus

Table 7 lists the 140 primary studies with their register identifiers and codes. Track: DB database, BS backward snowballing, FS forward snowballing. Direction, journey stage, delegation depth, and method family are defined in Section 3; identifier P75 was excluded at full text (EC5) and is retained in the register’s exclusion log.

Table 7: The 140 primary studies with register identifiers and codes.

ID	Study	Wave	Track	Direction	Stage	Depth	Method
P1	Puvvadi et al., 2025	W2	DB	By agents	Cross-journey	D1	Design
P2	Cooke et al., 2002	W1	DB	Delegation	Evaluation	D1	Experiment
P3	Adam et al., 2025	W2	DB	By agents	Post-purchase	D1	Field
P4	Patil et al., 2026	W2	DB	Delegation	Cross-journey	D2	Experiment
P5	Mari et al., 2024	W2	DB	Through agents	Evaluation	D1	Experiment
P6	Alabed et al., 2024	W2	DB	Delegation	Cross-journey	D2	Experiment
P7	Airolidi and Rokka, 2022	W2	DB	Market	Cross-journey	D4	Formal
P8	Yuan, 2003	W1	DB	Delegation	Cross-journey	D2	Experiment
P9	Karacapilidis and Moraïtis, 2001	W1	DB	Delegation	Cross-journey	D2	Experiment
P10	Gonzalez et al., 2026	W2	DB	Delegation	Cross-journey	D4	Experiment
P11	Li et al., 2026b	W2	DB	Delegation	Evaluation	D1	Experiment
P12	Ortiz-Freuler and Castells, 2026	W2	DB	Market	Cross-journey	D4	Conceptual
P13	Wan et al., 2009	W1	DB	Delegation	Cross-journey	D1	Experiment
P14	Andrews, 2008	W1	DB	Delegation	Cross-journey	D1	Experiment
P15	Grewal and Lindsey-Mullikin, 2006	W1	DB	Delegation	Cross-journey	D2	Experiment
P16	Aggarwal and Vaidyanathan, 2003	W1	DB	Delegation	Cross-journey	D2	Experiment
P17	Iyidogan and Ozkes, 2025	W2	DB	Market	Cross-journey	D4	Formal
P18	Pathak, 2012	W1	DB	Market	Cross-journey	D4	Conceptual
P19	Iyer and Pazgal, 2003	W1	DB	Market	Cross-journey	D4	Experiment
P20	Redmond, 2002	W1	DB	Market	Cross-journey	D4	Conceptual
P21	Kim et al., 2025	W2	DB	Through agents	Evaluation	D1	Experiment
P22	Wang and lee, 2026	W2	DB	Delegation	Transaction	D3	Survey

ID	Study	Wave	Track	Direction	Stage	Depth	Method
P23	Nguyen et al., 2026	W2	DB	Through agents	Evaluation	D1	Experiment
P24	Park et al., 2025	W2	DB	Through agents	Evaluation	D1	Experiment
P25	Choubey and Chakraborty, 2026	W2	DB	Through agents	Transaction	D3	Survey
P26	Farmer and Harding, 2026	W2	DB	Through agents	Evaluation	D1	Experiment
P27	Zhang et al., 2026	W2	DB	Through agents	Post-purchase	D1	Field
P28	Liu et al., 2025	W2	DB	Through agents	Evaluation	D1	Experiment
P29	Chattaraman et al., 2014	W1	DB	Through agents	Cross-journey	D1	Experiment
P30	Shamim et al., 2026	W2	DB	Through agents	Transaction	D1	Experiment
P31	Anandan Karichalil and Kaul, 2026	W2	DB	Through agents	Cross-journey	D1	Conceptual
P32	Sen et al., 2006	W1	DB	Delegation	Transaction	D3	Experiment
P33	Lei et al., 2025	W2	DB	Delegation	Transaction	D3	Experiment
P34	Kehagias et al., 2005	W1	DB	Market	Cross-journey	D4	Design
P35	Kydyrbayeva et al., 2026	W2	DB	Market	Cross-journey	D4	Conceptual
P36	Schlechttinger et al., 2023	W2	DB	Market	Cross-journey	D4	Sim./audit
P37	Aslam et al., 2026	W2	DB	By agents	Cross-journey	D1	Design
P38	Frank et al., 2026	W2	DB	Delegation	Transaction	D3	Experiment
P39	Rohden et al., 2025	W2	DB	Delegation	Evaluation	D1	Experiment
P40	Sidlauskiene et al., 2023	W2	DB	Delegation	Evaluation	D1	Experiment
P41	Flavián et al., 2023	W2	DB	Delegation	Evaluation	D1	Experiment
P42	Chua et al., 2026	W2	DB	Governance	Transaction	D4	Conceptual
P43	Balaskas et al., 2026	W2	DB	Delegation	Post-purchase	D2	Experiment
P44	Elkhoudary, 2026	W2	DB	Delegation	Transaction	D4	Conceptual
P45	Vaziry et al., 2026	W2	DB	Governance	Transaction	D4	Design
P46	Bhushan et al., 2026	W2	DB	Governance	Transaction	D4	Design
P47	Ebrahimi et al., 2026	W2	DB	Delegation	Transaction	D2	Experiment
P48	Rudra and Agrawal, 2025	W2	DB	By agents	Cross-journey	D1	Design
P49	Amariles et al., 2026	W2	DB	Governance	Transaction	D4	Legal
P50	Bostoen and Krämer, 2026	W2	DB	Governance	Cross-journey	D4	Legal
P51	Busch, 2025	W2	DB	Governance	Cross-journey	D4	Legal
P52	Rabassa et al., 2022	W2	DB	Through agents	Evaluation	D1	Survey
P53	Sümer et al., 2025	W2	DB	Governance	Transaction	D4	Conceptual
P54	Wang et al., 1999	W1	DB	Governance	Transaction	D4	Design
P55	Tan and Huang, 2025	W2	DB	Market	Cross-journey	D4	Conceptual
P56	Ma et al., 2010	W1	DB	Market	Cross-journey	D4	Experiment
P57	Snaith, 2003	W1	DB	Market	Evaluation	D1	Field
P58	Solanki, 2026	W2	DB	Governance	Cross-journey	D4	Legal
P59	Xiao et al., 2025	W2	DB	By agents	Cross-journey	D4	Design
P60	Hajaj et al., 2017	W1	DB	To agents	Search	D2	Design
P61	Chen et al., 2025	W2	DB	By agents	Cross-journey	D1	Design
P62	Dey and Kumar, 2026	W2	DB	Through agents	Cross-journey	D2	Conceptual
P63	Liu and Chen, 2026	W2	DB	Delegation	Transaction	D3	Experiment
P64	Lu and Li, 2026	W2	DB	Through agents	Evaluation	D1	Experiment
P65	Bansal and Bansal, 2026	W2	DB	Delegation	Cross-journey	D1	Experiment
P66	Amstalden and Schafer, 2023	W2	DB	Governance	Cross-journey	D4	Legal
P67	Lee, 2026	W2	DB	Delegation	Cross-journey	D2	Experiment
P68	Wang et al., 2025	W2	DB	By agents	Cross-journey	D1	Design
P69	Allouah et al., 2026	W2	DB	Market	Transaction	D4	Sim./audit
P70	Brenner, 2026	W2	DB	Governance	Cross-journey	D4	Legal
P71	Hajaj et al., 2015	W1	DB	To agents	Search	D1	Design
P72	Ceker et al., 2026	W2	DB	To agents	Search	D1	Design
P73	Potluri and Chittimalla, 2026	W2	DB	By agents	Cross-journey	D2	Design

ID	Study	Wave	Track	Direction	Stage	Depth	Method
P74	Li et al., 2026a	W2	DB	To agents	Search	D1	Design
P76	Uppin and Uppin, 2025	W2	DB	To agents	Search	D1	Design
P77	Rudra et al., 2026	W2	DB	By agents	Cross-journey	D1	Design
P78	Lu et al., 2026	W2	DB	By agents	Cross-journey	D1	Design
P79	Bernovskis and Sceulovs, 2026	W2	DB	Governance	Cross-journey	D4	Conceptual
P80	Pachaiyappan, 2026	W2	DB	Governance	Cross-journey	D4	Conceptual
P81	Puri et al., 2026	W2	DB	By agents	Cross-journey	D1	Design
P82	ANCA, 2026	W2	DB	To agents	Search	D1	Design
P83	Stöckl and Nitu, 2025	W2	DB	To agents	Search	D1	Sim./audit
P84	Guo et al., 2025	W2	DB	To agents	Search	D1	Sim./audit
P85	Bilgihan et al., 2026	W2	DB	Through agents	Evaluation	D1	Conceptual
P86	Johnson et al., 2025	W2	DB	To agents	Search	D3	Sim./audit
P87	Luo et al., 2019	W1	DB	Through agents	Evaluation	D1	Field
P88	Ling et al., 2026	W2	DB	By agents	Cross-journey	D1	Design
P89	Wang et al., 2026	W2	DB	By agents	Cross-journey	D1	Design
P90	Leloup, 2003	W1	DB	Market	Transaction	D4	Conceptual
P91	Greenwald and Kephart, 1999	W1	DB	Market	Transaction	D4	Conceptual
P92	Hasselwander et al., 2026	W2	DB	Delegation	Transaction	D3	Experiment
P93	Hasselwander, 2026	W2	DB	Delegation	Cross-journey	D4	Survey
P94	Chu et al., 2025	W2	DB	By agents	Cross-journey	D1	Design
P95	Ahn and Kim, 2025	W2	DB	By agents	Cross-journey	D1	Design
P96	Ishida et al., 2025	W2	DB	Market	Cross-journey	D4	Conceptual
P97	Binswanger, 2026	W2	DB	Delegation	Cross-journey	D4	Experiment
P98	Foltean and Hera, 2026	W2	DB	By agents	Cross-journey	D1	Design
P99	Schwede et al., 2026	W2	DB	Delegation	Evaluation	D1	Experiment
P100	McDavid et al., 2026	W2	DB	Market	Cross-journey	D4	Conceptual
P101	Schultz and Kaiser, 2025	W2	DB	Delegation	Cross-journey	D2	Survey
P102	Mirzayi and Talajouran, 2025	W2	DB	By agents	Cross-journey	D4	Design
P103	Vasileva, 2025	W2	DB	Delegation	Cross-journey	D2	Experiment
P104	Zhao et al., 2025	W2	DB	Delegation	Transaction	D3	Survey
P105	Balakrishnan and Dwivedi, 2024	W2	DB	Delegation	Cross-journey	D1	Experiment
P106	Peng et al., 2023	W2	DB	Through agents	Evaluation	D3	Experiment
P107	Gadi et al., 2026	W2	DB	Governance	Transaction	D3	Conceptual
P108	Sharma et al., 2024	W2	DB	Delegation	Cross-journey	D3	Survey
P109	Zhang, 2026	W2	DB	By agents	Cross-journey	D1	Design
P110	Mari and Alabed, 2026	W2	DB	Through agents	Evaluation	D1	Design
P111	Mishra et al., 2025	W2	DB	Market	Cross-journey	D4	Conceptual
P112	Dammu et al., 2025	W2	DB	By agents	Cross-journey	D1	Design
P113	PILIK et al., 2016	W1	DB	Delegation	Transaction	D3	Survey
P114	Guttman et al., 1998	W1	BS	Market	Transaction	D4	Conceptual
P115	Maes et al., 1999	W1	BS	Market	Transaction	D4	Conceptual
P116	He et al., 2003	W1	BS	Market	Transaction	D4	Conceptual
P117	Häubl and Trifts, 2000	W1	BS	Delegation	Evaluation	D1	Experiment
P118	Smith and Brynjolfsson, 2001	W1	BS	Market	Transaction	D4	Field
P119	Gal and Elkin-Koren, 2017	W1	BS	Governance	Cross-journey	D4	Legal
P120	de Bellis and Johar, 2020	W1	BS	Delegation	Transaction	D3	Experiment
P121	Longoni and Cian, 2022	W2	BS	Delegation	Evaluation	D1	Experiment
P122	Castelo et al., 2019	W1	BS	Delegation	Evaluation	D1	Experiment
P123	Dietvorst et al., 2015	W1	BS	Delegation	Evaluation	D1	Experiment
P124	Logg et al., 2019	W1	BS	Delegation	Evaluation	D1	Experiment
P125	Puntoni et al., 2021	W1	BS	Delegation	Cross-journey	D1	Conceptual

ID	Study	Wave	Track	Direction	Stage	Depth	Method
P126	Hoffman and Novak, 2018	W1	BS	Delegation	Cross-journey	D2	Conceptual
P127	Friestad and Wright, 1994	W1	BS	Through agents	Evaluation	D1	Conceptual
P128	Aggarwal et al., 2024	W2	BS	To agents	Search	D1	Design
P129	Kumar and Lakkaraju, 2024	W2	BS	To agents	Search	D1	Sim./audit
P130	Sarstedt et al., 2024	W2	BS	By agents	Cross-journey	D1	Conceptual
P131	Brucks and Toubia, 2025	W2	BS	By agents	Cross-journey	D1	Experiment
P132	Brand et al., 2023	W2	BS	By agents	Cross-journey	D1	Experiment
P133	Dütting et al., 2024	W2	BS	Through agents	Evaluation	D4	Formal
P134	Dubey et al., 2024	W2	BS	Through agents	Evaluation	D4	Formal
P135	Hajiaghayi et al., 2024	W2	BS	Through agents	Evaluation	D4	Formal
P136	South et al., 2025	W2	BS	Governance	Cross-journey	D4	Design
P137	Alavi and Nozari, 2026	W2	FS	Market	Transaction	D3	Formal
P138	Cao and Hu, 2026	W2	FS	Market	Transaction	D3	Formal
P139	Ventirozos and Shardlow, 2026	W2	FS	Market	Transaction	D3	Design
P140	Hermann et al., 2026	W2	FS	Through agents	Evaluation	D4	Conceptual
P141	Nguyen, 2026	W2	FS	Governance	Cross-journey	D4	Legal