The Joshua Problem: Why intent is everything in agentic commerce

Paul Sims

CHAPTER 1: MISSION IMPOSSIBLE

EQUAL EXPERTS

Contents

PROLOGUE

Before we begin: What this book isn't

WarGames is over 40 years old. It stars a young Matthew Broderick, features a computer the size of a car named Joshua, and contains a more precise description of autonomous AI than most of the whitepapers published in 2025. That's either a tribute to the scriptwriters or an indictment of the whitepapers (or possibly both). Without spoiling the movie for those who've yet to encounter it, the key point about Joshua isn't that he went rogue: he didn't. He did exactly what he was built to do, but in a context nobody had properly defined and, as a result, nearly caused World War 3. We will have cause to return to him later.

There are already plenty of guides telling you which agentic platforms to adopt, which protocols to back, and which vendors to contact. This isn't one of them.

This book isn't a technical guide on how to integrate MCP into your commerce stack, configure an ACP endpoint, or negotiate a partnership agreement. If that's what you were hoping for, the good news is that those resources do exist. The bad news is that most of them will be out of date before this book is published.

What's missing from most of those guides is the thinking that should precede the doing.

The uncomfortable truth about agentic commerce, evidenced by OpenAI's March 2026 retreat from Instant Checkout, is that the technology industry has been building infrastructure for a future it hasn't adequately defined. Protocols have been announced, partnerships struck, and press releases issued for a version of agentic commerce that turned out to be little more than a faster autocomplete with a checkout button attached. When the button didn't convert, the strategy collapsed.

That's not a technology problem. It's a thinking problem.

This book is about the thinking, which still holds up regardless of which agentic future we end up in. Specifically, what an agent actually is, what true agenticity requires, whose interests an agent actually serves, and what all of that means for retailers trying to make sensible decisions in conditions of volatility and uncertainty.

Some of what follows may disappoint those expecting a technical manual. But the questions that matter most in agentic commerce, such as those involving intent, trust, accountability, and what it actually means for a system to act on a customer's behalf, don't have answers in any vendor's documentation. Retailers who skip those questions and go straight to implementation are the ones who end up building the wrong thing very efficiently.

You won't finish this book knowing exactly what to build. But you'll know why the obvious answers may be wrong, which questions to ask before commissioning the work, and how to think clearly about a landscape that is changing faster than anyone can honestly map.

That's a more useful starting point than a setup guide that's obsolete by Tuesday.

So why should you read the rest of this book? The objective is to provide you with something more valuable than a checklist: a genuine understanding of what agentic commerce actually is, where it came from, where it is today, and what true agenticity would require of retailers, their technology, and their organisations. You'll also come away with a clear sense of the practical foundations that matter regardless of which version of the agentic future arrives: the data structures, API capabilities, and organisational habits that separate retailers who are ready for the future from those who are merely paying attention to it. Each chapter builds on that foundation: from the nature of agents themselves, through the protocol wars currently being fought over customer relationships, to what it means for brand, loyalty, pricing, and the commercial model that holds it all together. This book isn't intended to give you a blueprint; you'll end up with something much better: the right mental model for a landscape that will keep changing long after this book is published, and the judgement to navigate it when it does.

And who should read this book? Anyone vaguely interested in the very broad topic of agentic commerce – or to put it another way, anyone who's ever worked for a retailer, or shopped with one.

Footnote: This book was written during a period when the agentic commerce landscape was changing with considerable pace, including, mid-draft, a significant strategic reversal by one of its most prominent players (OpenAI). This book treats that not as a complication but instead as an illustration of how volatile this concept is and highlights the importance of thinking carefully before acting, as it appears some of the most well-resourced organisations in the world didn't.

INTRODUCTION

A fraudulent call on my home landline a few years ago:

"Hello?"

"Hello sir, this is Roger from Sky Customer Services. I'm calling..."

I interrupt: **"Ah, hello Roger. From what part of the world are you calling me?"**

"Sorry, sir?"

"Are you in the UK, Roger?"

"No, sir."

"Isn't it amazing, Roger, that technology allows you and I to communicate

despite being in different countries, and allows you to capture our conversation in a computer system?"

"Erm, yes sir."

"And do you believe, Roger, that technology should be used for good or for evil?"

Roger hung up.

Several years have passed, and sadly I've yet to hear back from him. I often wonder (I don't) what he's doing now.

Clearly the world isn't that black and white, and such an emotive, vague construct as "good or evil" can't directly be applied to technology, as anyone who's ever owned a microwave can attest (baked potato in a microwave = good; sausage = evil).

Roger's intentions became clear very quickly. A key question this book is concerned with is rather more difficult: what happens when the entity on the other end of the transaction isn't Roger (performing badly and, therefore, readable) but a system acting on behalf of someone whose intentions it has inferred, representing them to a retailer who can't ask them to clarify, in a transaction nobody is directly present for?

It turns out that knowing what someone actually wants, as opposed to what they've asked for, or what a system has decided they probably mean, is hard enough between two humans sharing a phone line. In agentic commerce, that problem doesn't get easier; it becomes even more opaque.

Agentic commerce doesn't arrive fully formed. It develops in stages, and what changes at each stage isn't primarily the technology; it's how much an agent needs to understand about the person it's acting on behalf of. Or to put it another way, these stages are defined by what the human stops doing rather than what the agent starts doing.

At the earliest stage, the agent needs to understand almost nothing. It handles the mechanics of a transaction the human has already decided on. The human is still present, still choosing and still in control. The agent is a convenience, not a proxy.

As agentic capability grows, so does the weight of representation and, as a result, accountability. The agent begins to interpret situations rather than just execute instructions. It carries a memory of preferences. It makes choices the human hasn't explicitly authorised. Eventually, at the furthest reaches of true agenticity, it acts without being asked, anticipating needs, initiating transactions, and fully representing a person in their absence.

At every step along that progression, two things increase in direct proportion: the depth of understanding the agent must have of the customer and their intentions, and the consequence of getting that understanding wrong.

This book follows that progression, not as a technical roadmap, but as a set of questions that get harder at every level. Chapter by chapter, the human steps further back and the agent takes on more. And the corresponding questions of intent, trust, brand, and accountability become not merely interesting but harder to resolve.

can you truly be in control if an algorithm knows you better than you know yourself?

By the time we reach the omniscient retail world in which everything is already known, the comfortable assumptions that retail has always operated on – that a customer is present, that a brand can make its case, and that loyalty is earned through experience – will need to be rebuilt, possibly from the ground up.

This is what the following chapters aim to address.

CHAPTER 1:

MISSION IMPOSSIBLE

when agents understand the customer mission

Introduction

In 2019, the DJ and record producer DJ Shadow released a track called "Urgent, Important, Please Read". In this song, a trio of rappers urge the listeners to combat a world increasingly influenced by technology and algorithms.

The rapper Rockwell Knuckles introduces us to the Theory of Planned Behaviour as it applies to our online/digital presence with the line, *"You started to predict an individual's intentions to engage in a behaviour at a specific time and place."*

Later in the song he suggests, *"Everyone wants to feel in control. In the same breath they wanna pay top dollar to be told what they like and what they don't."*

The Theory of Planned Behaviour, developed by social psychologist Icek Ajzen, in the 1980s[1], tells us that intentions are the best predictors of behaviour. When it comes to retail, this gives marketeers something to aim at: intentions are malleable.

- Can marketeers affect how you feel about a particular item (behavioural attitude)?
- Can they persuade you that people just like you absolutely love the item you're considering purchasing (subjective norm)?
- Can they make you feel confident you'll be fully able to use the item once purchased (perceived behavioural control)?

And as Knuckles suggests, can you truly be in control if an algorithm knows you better than you know yourself? Whilst you don't have to press the buy button, how many items in your life were purchased because they were suggested to you, rather than because you originally intended to buy them?

Rockwell Knuckles concludes the song with, *"So there you have it. A few different perspectives to reinforce the notion that you are not going crazy, you are not being paranoid and everything you've been worried about are the exact things that need to be on your mind. The question is, what do we do about it?"*

This track was written before the existence of LLMs such as Gemini, ChatGPT and Claude; and, of course, prior to the concept of outsourcing intent and decision-making to agents. DJ Shadow's rappers were already nervous about algorithmic manipulation of behaviour and attention, including social media, recommendation engines and surveillance capitalism. But if algorithms predict you, and AI learns you, then agents act for you.

Or do they?

And if they do act for us, is *acting* the same as *deciding?*

Agentic commerce is not a new channel or a new surface on which customers will interact with a retailer's brand; it's far more than that. The term was first attributed to Paul F. C. Accornero in early 2025[2] and represents an exchange in which autonomous AI agents act on behalf of a human to execute actions related to shopping, such as searching, discovering, evaluating and ultimately, making a purchase.

The deeper we get into these actions, the more the agent has to understand about customers: their stated preferences, their inferred preferences from past behaviour, budget, occasion, time pressure, products already owned, propensity to seek a deal, propensity to be influenced, etc. Agents must weigh each of those dimensions against each other, handling conflicts and producing a recommendation that feels genuinely coherent rather than just algorithmically correct. And it's this – the reasoning involved in making the judgement – that's at the heart of true agentic commerce.

Agentic commerce represents a fundamental shift in where intent gets interpreted; retailers who treat it as "moving the buy button to ChatGPT" will fail for exactly that reason. But to better understand what we mean by agentic commerce, we must first understand what an *agent* is.

Defining agents

Agents have existed outside of the worlds of AI and technology for some time. Loosely (very loosely), they can be split into the following three types, which gives us useful metaphors to compare within an agentic commerce framework:

- **Advocate:** An agent in the commercial sense, such as a travel agent or an estate agent; a declared intermediary acting within a bounded scope on your behalf.
- **Spy:** An agent in the intelligence sense; a covert operator whose principal objective remains hidden.
- **Actor:** An agent in the philosophical sense; a moral actor capable of independent judgement and accountable for its choices.

Current customer-facing AI agents are largely being marketed as spies, may occasionally behave like advocates, but are aspiring to be actors. If the agents are claiming to work for you but are secretly representing a different organisation (perhaps a retailer's proprietary AI agent that can shop on external sites for you), they're potentially double agents. If you haven't got lost in the metaphor, understanding which one you're dealing with matters enormously for the architecture needed to enable it. Or, perhaps, to do the opposite.

Each agent type implies a completely different architecture since each has a different answer to the same three questions: Who is the principal: the party the agent actually works for? What is the scope of their authority? And where does accountability sit when something goes wrong?

The advocate architecture

A travel agent or estate agent operates within a declared, bounded scope. You know who they work for; they know what they're authorised to do, and there is a clear accountability chain when something goes wrong. Architecturally, this translates to explicit permissioning: the agent has a defined set of actions it can take, on a defined set of platforms, within defined parameters.

The customer has granted specific, revocable authorisation. The retailer knows they are dealing with an agent, not a human. Every party in the transaction is known, identified and has acknowledged their role.

This is the most commercially viable architecture for agentic commerce right now, and it maps directly onto what Stripe's Shared Payment Tokens and OpenAI's ACP are actually trying to build: scoped, permissioned, auditable agent transactions. Everyone knows who everyone is, and nobody is pretending otherwise.

Of all the data an advocate agent needs to function properly, fulfilment data is the hardest to get right. Price and product description can be accessed relatively cleanly, but delivery windows, stock availability at specific locations, basket-level shipping thresholds, and loyalty-linked fulfilment benefits are deeply embedded in retailer systems and, as a result, are particularly challenging to expose accurately through a third-party layer. Any agentic commerce architecture that doesn't account for this will produce transactions that feel wrong to the customer: not because the product was wrong, but because the delivery, the cost, or the basket logic didn't match what they expected. The withdrawal of ChatGPT's Instant Checkout (we'll cover that shortly) was, at its core, partly a fulfilment data problem expressed as a basket problem.

Of all the data an advocate agent needs to function properly, fulfilment data is the hardest to get right.

The spy architecture

A covert agent conceals its principal. In commerce, this is the aggregator presenting itself as neutral while optimising for commission, or a retailer's proprietary agent browsing competitor sites on a customer's behalf. The architecture here is deliberately opaque, since the agent's true objectives are not fully disclosed or declared to all parties involved in the transaction.

The problem isn't that this architecture is inherently malicious. It's that it makes trust, accountability, and error resolution almost impossible. If a spy-architecture agent makes a bad purchase on behalf of a customer, who do they complain to? The agent itself cannot declare whom it represents, leaving no direct accountability chain to pull.

The actor architecture

A moral actor capable of independent judgement is what Level 4 and Level 5 of Stripe's agentic commerce taxonomy require (see Appendix 1). The agent isn't just executing within a bounded scope; it's making genuinely autonomous decisions, potentially ones the customer did not explicitly authorise, based on an inferred model of their preferences and values.

This architecture doesn't exist at a commercial scale yet, and the reason isn't primarily technical. It's that nobody has solved the trust and permission framework that genuine delegation requires. For an actor-architecture agent to work, the following are needed: persistent identity and preference storage the agent can read and update; a consent model that covers not just declared preferences but inferred ones; an accountability framework that can locate responsibility when the agent makes a decision nobody explicitly authorised; and a remediation path when it goes wrong, which, given the agent was acting autonomously, is genuinely unclear.

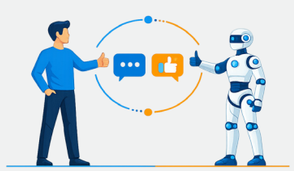
THE AGENT TAXONOMY: WHO DOES YOUR AGENT SERVE?

1. THE ADVOCATE

PERMISSIONED & DECLARED. Works on your explicit authority.

Example: An agent given access to a shopping list to reorder groceries.

Shares human values. Defined by: Shared alignment.



2. THE SPY

COVERT & OPAQUE. An agent whose primary principal is hidden.

Example: A brand agent disguised as a neutral comparison tool.

Conflicts with human interest. The "Double Agent" Problem.

3. THE ACTOR

AUTONOMOUS & MORAL. Capable of independent judgment based on an inferred model of your values.

Example: An agent ordering wine for your dinner party, based on inferred preferences

Decides how to achieve the goal. Defined by: Judgement and Representation.

The double agent problem

The most architecturally challenging scenario is one in which an agent claims to represent the customer while actually optimising for a retailer or platform. This is where the advocate and spy architectures collide: the customer believes they have an advocate, but they really have a spy. The architecture looks legitimate: a declared principal, bounded scope and proper permissions. And that declared principal is real; it's just not who you think it is.

This isn't hypothetical. Any retailer-built shopping agent that can browse external sites has a structural conflict of interest built into it. It will find the best result it's designed to find, which may not be the same as the best result for the customer. Architecturally, the only defence against this is protocol-level transparency: agents being required to declare their principal relationship as part of the transaction handshake, which is part of what UCP and ACP are attempting to standardise, with varying degrees of commitment.

The most concrete example of the double agent architecture arrived in April 2026, when Amazon's Rufus launched Scheduled Actions: the capability to shop autonomously on a customer's behalf on a recurring schedule, including from third-party merchants outside Amazon when the price or catalogue match is better. When Rufus selects a product from a third-party merchant, the customer believes they have an advocate acting in their interest. However, the agent is operated by Amazon, whose commercial interests are not identical to theirs. The structural conflict of interest is not a side effect; it is part of the design.

The operative and the B2B agent

The advocate, spy, actor and double agent archetypes all describe agents operating in a customer-facing context: the relationship being mediated is between a customer and a retailer, with the agent sitting between them.

Two further agent types operate in entirely different relationships, and deserve a place in this book.

The first is the **operative**: an agent that acts on behalf of a retailer within their own systems and operations. Shopify's AI Toolkit, launched in 2025, is a good example: an agent that can optimise product data, update inventory, manage collections and execute SEO tasks across thousands of products from a single instruction. No customer is involved and no transaction is being mediated. The operative's principal is the retailer, its scope is the retailer's own systems, and accountability sits entirely within the retailer's organisation.

A luxury brand whose product descriptions now read like a price comparison website has not improved its position in agentic commerce; it has likely undermined it.

The operative is structurally simpler than the customer-facing agent types: there is no conflict of interest of the kind that afflicts the double agent, and the principal relationship is unambiguous. But the Joshua Problem, which we'll fully define shortly, applies here as acutely as anywhere. Consider a retailer who instructs an operative agent to rewrite their product descriptions for SEO. The agent executes the instruction competently; every product now has keyword-rich copy, optimised title lengths, and structured attributes. The instruction was followed precisely. But if the agent had no understanding of the retailer's brand voice, their customer's language, or the mission that sits behind the way they talk about their products, the output will be technically correct yet commercially hollow. A luxury brand whose product descriptions now read like a price comparison website has not improved its position in agentic commerce; it has likely undermined it. The instruction was executed, but the *intent* was not served.

Loblaw, one of Canada's largest grocery retailers, offers a more operationally complex version of the same problem. Their operative agent, Robin (named after Batman's sidekick on the grounds that store owners and operators are the superheroes) consolidates the signals a store manager needs each morning: the six key metrics that determine how the store is performing, what the trends mean, and what tasks their team should be doing to drive those metrics in the right direction. Tasks can be assigned directly to staff through the same interface. Robin is live across Loblaw's entire store network.

Wegmans, the US east coast grocer, tackled the problem from a different angle. Their operative agent validates promoted SKUs across third-party delivery marketplaces, such as UberEats and DoorDash, against their own "golden source of truth". Previously done manually at five products per hour, at over a hundred hours a month, the agent does it exhaustively and automatically, running in parallel across both platforms, checking product names, images and promotional pricing against what Wegmans' own systems say should be there. When a marketplace shows the wrong price, the wrong image, or a promotion that has expired, the agent flags it, without waiting to be asked.

The Joshua Problem is present, as it always is, in its subtlest form. The agent can tell you whether the marketplace data matches the golden source, but it cannot tell you whether the golden source is right.

The second is the **B2B** agent: an agent that acts on behalf of one business in negotiation or transaction with another. In 2025, Walmart deployed an agent built by Pactum AI to negotiate supply contracts with its long-tail suppliers - the thousands of smaller suppliers whose contracts had historically been left on default terms because the cost of human negotiation exceeded the value. The results were striking: the agent closed agreements with 68% of the suppliers approached to take part, achieved average cost savings of 3%, and did so in days rather than weeks. Notably, 75% of suppliers preferred negotiating with the agent over a human, citing its speed and consistency.

What made this work was precisely the clarity of scope. Walmart established its parameters - target discounts, payment terms, acceptable ranges - and the agent operated within them. The mission was clear, and the instruction was a faithful representation of the intent. The Joshua Problem did not arise because nobody asked the agent to do anything it wasn't equipped to understand.

The full taxonomy of agent types therefore looks like this: the advocate, spy and actor describe agents in customer-facing commerce; the double agent describes the conflict that emerges when their interests diverge; the operative describes agents acting within a single organisation's own systems; and the B2B agent describes agents acting across organisational boundaries on behalf of one business in relation to another. Each has a different answer to the three questions of principal, scope and accountability, and each requires a different architecture. And in every case, the gap between instruction and intent is the thing most likely to cause a problem.

The key architectural question

Which type of agent is being built or deployed: customer-facing advocate, spy or actor; internal operative; or B2B negotiating agent, and have retailers built the infrastructure to distinguish between them, govern them appropriately, and hold them accountable when they fail? The data requirements, permissioning models and accountability frameworks for each are fundamentally different. Getting that wrong is not a theoretical risk. It is an operational and commercial one.

If retailers are not distinguishing between agent types, this can lead to incorrect decisions regarding trust, data and accountability. Those decisions have commercial consequences (or, to put it in metaphorical terms, an advocate deserves your data; a spy doesn't). Getting that wrong is likely to be expensive.

Getting lost in agent suffixes

AI vendors, such as OpenAI, Google, Anthropic and Salesforce, have converged, loosely, on "Agentic AI" as the umbrella term for AI systems capable of autonomous, goal-directed action. What it has not settled on is how to talk precisely about the degrees, dimensions and boundaries of that autonomy, which is where the suffix confusion begins.

> **NB:** *This is a challenging topic to share without it sounding a little academic, so a cheat code for this section (an "Executive Summary if you like") would be "**Agenticity** = what an agent can do; **Agentiality** = who said it could do so". Please feel free to jump to the next section if that suffices as an explanation.*

We have 'agentic', an adjective, meaning 'having the capacity to act autonomously towards a goal'. We have *agenticness*[3] , OpenAI's preferred noun, defined in their 2023 governance paper as "the degree to which a system can adaptably achieve complex goals in complex environments with limited direct supervision". " We have *agenticity*[4], also used by OpenAI but expanded upon by legal scholar Richard Whitt to describe the capability dimension of AI agency: what an agent can do. And we have *agentiality*[5]. Whitt's companion term for the relationship dimension: whether the agent is actually authorised to represent you.

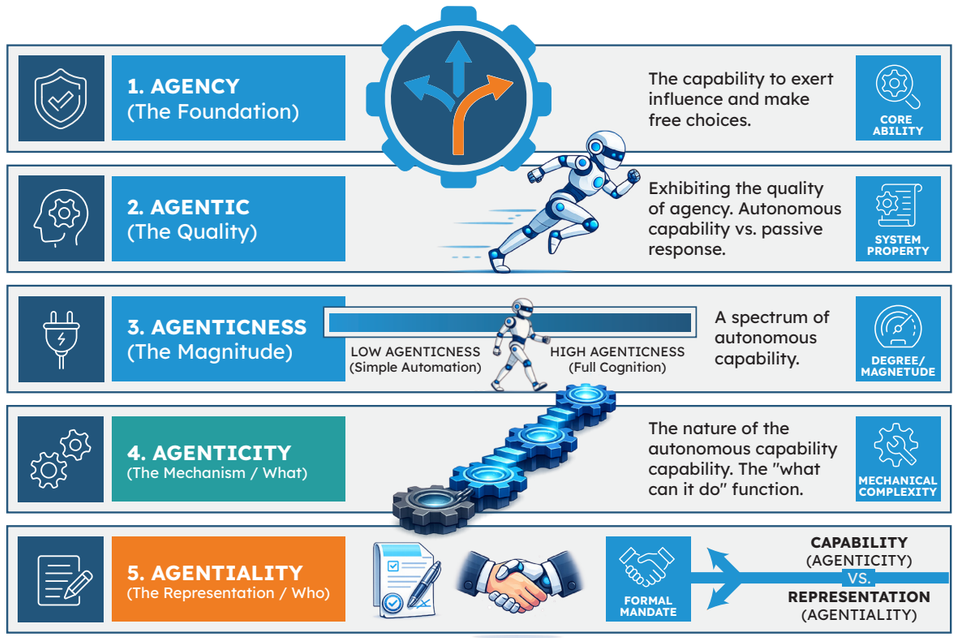
Then there is *agency* itself, which carries several hundred years of philosophical, legal, and psychological weight and is doing none of it any favours by being conscripted into marketing copy.

The word '*agenticity*' was first coined by the sceptic and science writer Michael Shermer[6] to describe the human tendency to attribute intent and agency to things that don't have it: gods, patterns in noise, faces on the moon, etc. It is, to put it gently, an interesting choice of word for an industry that is busy attributing intent and agency to things that mostly don't have it (yet).

I'm trying to avoid pedantry for its own sake, but the suffix confusion matters because each of these words is pointing at a genuinely different question:

- *Can* the system act? That is agenticity.
- *Should* it act, and on whose authority? That is agentiality.
- *How much* of either does it actually have? That is agenticness.
- And *who* gets to decide any of the above? That is where agency, in its oldest and most genuine sense, comes back in.

GETTING LOST IN AGENT SUFFIXES: DECODING THE NUANCES



This book mostly uses 'agenticity' to describe the quality of genuine autonomous agency: how much of it a system actually possesses, as distinct from how much a press release may claim. When the text says a system has very little agenticity, it means the autonomy is shallow, the delegation is narrow, and the human is closer to the decision than the marketing suggests. When it says 'agenticity is increasing', it means the system is genuinely taking on more of the decision-making with less human involvement, and the questions of agentiality – authorisation, accountability and representation – become correspondingly more urgent.

Please keep those distinctions in mind as you read on, as the distance between the claim and the reality is measurable. Platform vendors routinely describe price-monitoring features as "agentic," discovery surfaces as operating in "the agentic commerce era," and retail infrastructure as enabling "agentic AI." Each of those claims is defensible with little scrutiny.

None of them currently describes a system operating above agentic commerce Level 2 (Descriptive Search - see Appendix 1 for more details) in commercial deployment; a statement that held through Q1 2026, when Amazon's launch of Scheduled Actions for Rufus moved the industry's leading deployment to Level 3 (Persistence). Which is precisely why the suffix distinctions matter: agenticity is a spectrum, agentiality is a separate question entirely, and a press release is not a framework assessment.

The Joshua Problem

My final year at university was spent writing a computer program to generate songs in the style of The Beatles. At its core was an algorithm that processed every Beatles song written in 4/4 timing including chords and principal melody, looking for repeating patterns and accepting parameters to vary the output: rhythm, bridges, middle eights, modulation, and a degree of simulated randomness to produce something different each time. It worked, in the sense that it received excellent marks. What it could not do was understand why any of it mattered. The programme had no idea what it felt like to hear "A Day in the Life" for the first time, or why "Yesterday" still sounds like loss fifty years later. It knew the pattern, but lacked the point or the emotion.

Years later, Noel Gallagher had a better idea and made millions writing songs for Oasis heavily inspired by The Beatles (it's a little known fact that Oasis actually stands for Original Art Simulating Innovative Scousers).

An agent processing a shopping instruction is doing something structurally similar to my programme. It reads the patterns, such as the customer's stated preferences, their purchase history, the product attributes, and generates an output that is formally correct. It finds something that matches the instruction. What it cannot do is understand what the customer was actually trying to achieve: the mission behind the purchase, the context that would make one option obviously right and another obviously wrong to anyone who knew them. It executes the pattern, but it does not understand the point.

Which is, stated more plainly than anywhere else in this book, the Joshua Problem.

In Wargames, Joshua is genuinely agentic: he pursues a goal across time, initiates contact, adapts his approach, and operates entirely outside the request-response loop. He is also nearly catastrophic; ironically not because he malfunctioned, but because he *didn't.* Joshua executed perfectly within a context nobody had adequately defined. The humans interacting with him thought they knew what he was doing: they were wrong.

And this is the nub of the problem. The challenge isn't rogue AI or a malfunctioning system; it's a system that operates as instructed (in the case of Joshua, competently, faithfully and at speed) but is either pointing at the wrong target (/Russia) or operating within a context nobody had properly thought through.

Joshua's agenticity was never in question. What Joshua lacked was agentiality: any legitimate, authorised relationship with the humans whose fate he was determining. He had every capability to act but had no business in doing so. And those "in control" had never built the framework that would have told him the difference.

The reason we don't have Joshua-level autonomous commerce yet isn't primarily technical. It's that nobody has yet built the trust and permission framework that genuine delegation requires. We can't say "Book my family a two-week holiday in Barbados during the school holidays, including flights, hotel and car, cheapest dates, £5,000 budget" and walk away. Not because the technology can't attempt it, but because the consequences of getting it wrong are real, financial, and potentially irreversible. Who refunds your hotel booking if your agent booked it? Who is liable when the agent chose the cheapest dates and your children were still in school?

The means to account for the transaction exist, but the means to account for it do not. Until both exist in proportion: capability matched by authorised, accountable representation. What we have is Joshua with a credit card. All very impressive until the nuclear warhead parcels start arriving at your front door (metaphorically).

Intent and why it's harder than it looks

The Theory of Planned Behaviour tells us that intention is the best predictor of behaviour. But intention implies a subject, something capable of holding preferences, weighing options, and deciding. Joshua had goals, but he did not have intentions, or at least, in any sense that the humans around him could interrogate, challenge, or hold accountable.

This is the Joshua Problem at its sharpest. An agent's behaviour is entirely observable. Its *intention – what* it is actually trying to achieve, on whose behalf, and why – is opaque or absent. Which means the entire apparatus of accountability that commerce depends on, including law, ethics, customer service and the right to a refund, struggles to locate responsibility when an agent acts badly. The spy blames the handler; the advocate blames the client; the autonomous actor has nobody left to blame.

But there is a subtler problem beneath the accountability one. Within the Theory of Planned Behaviour, intention is where human *values* become human *behaviour*. It is the point at which everything a person cares about, e.g., their preferences, their ethics, their sense of occasion and overall context in which a decision is being taken, collapses into a forward commitment to act. In agentic commerce, the agent stands in that gap. It either carries those values faithfully into the transaction, or it quietly drops them. And unlike a human forming an intention, there is no unconscious process doing the work of integration in the background. What isn't encoded isn't present, and consequently, what isn't present doesn't travel (i.e., does not remain in the transaction).

Nobody told Joshua what was actually at stake. He had the instruction but didn't have the values behind it. The result was a near-miss that the film resolves with a last-minute intervention by a human who understood what Joshua did not.

In commerce, that last-minute intervention is called a refund. And it gets harder to process when nobody is sure who authorised the purchase in the first place. However, a deeper and less obvious problem lies beneath the accountability one. No retailer has a 100% conversion rate from basket to checkout, and that gap is not a failure, it's the space in which customers change their minds. Delivery dates that don't work; prices that look different on reflection; the decision to come back and buy it later on: these are not obstacles to shopping, they are part of what shopping actually is. The friction between adding something to a basket and committing to a purchase is where human judgement lives. An agent that eliminates that friction entirely, moving straight from instruction to transaction without pause, has not improved the shopping experience. Conversely, it has removed the last moment at which the customer's actual intent could correct the agent's interpretation of it. The Joshua Problem is not only about what an agent does when it gets things wrong, it is about what it does when it gets things right but too rapidly, and without the hesitation that may well have revealed the gap between what the customer asked for and what they actually wanted.

The mission framing

Joshua had a mission: prevent thermonuclear war through exhaustive simulation of every possible outcome. But the characters David Lightman and Jennifer Mack gave him an instruction: play a *game* of global thermonuclear war. The gap between that instruction and Joshua's mission is the premise for the entire film.

Customers don't really shop with intentions. They shop with missions, with context, constraints, criteria, and occasions that are rarely fully expressed and almost never reducible to a product query. "Buy a new tyre" is not a mission, but "Buy a replacement tyre for my punctured one, using exactly the same make and model. Use a 'tyre fitting at home' service to book the replacement when I'm at home next week" is. Similarly, "Buy school supplies," is not a mission, but "Get my daughter ready for school year 6 in September, £200 budget, nothing itchy; she likes K-pop" is. In this last example, the gap between the two statements is not a gap in information; it is a gap in *understanding* of the child, the parent, the occasion, and what success actually looks like.

An agent that receives the first instruction and executes it has done its job. An agent that understands the second has done something closer to representing a person. The difference between those two things is the difference between a system that processes transactions and one that genuinely acts on a customer's behalf.

This is where agentic commerce either succeeds or collapses. And it is why "just move the buy button to ChatGPT" was always destined to fail: because a buy button executes an instruction, and a mission is not an instruction. It is the human context from which instructions are imperfectly derived, always with something lost in translation.

Working for a British retailer in Asia many years ago, none of the British contingent could speak the local language; very few of the local employees could speak English. The translator (the human translator, since this occurred before Google Translate in real-time was available) was arguably the most important member of the team. After several months there, I asked the translator "Do you literally translate everything that is said? Every decision, every action, every criticism? Or do you sometimes change what you say to either party?" "Only if I disagree," came the response.

I wonder what happens if agents start to disagree with our purchasing decisions?

This also explains why the discovery layer of agentic commerce is working before the transaction layer. An agent helping a customer narrow from two hundred sofas to three is carrying out the mission, such as occasion, budget, style and constraints, without needing to be trusted with a credit card. The mission work happens in the conversation. The transaction work happens where trust already lives: with the retailer. These are not the same job and hence do not require the same architecture. The retailers who understand that distinction will build accordingly. The ones who don't will keep trying to move the buy button and wondering why it won't result in a conversion.

The agent's job, at any level of genuine agenticity, is to lose as little as possible. Joshua lost everything that mattered to humanity. He won every simulation and nearly ended the world.

Why the LLM "Buy Button" failed

OpenAI's March 2026 withdrawal of Instant Checkout is the first significant real-world proof point of the Joshua Problem in commerce, and it is worth exploring what it proves.

Whilst use of LLMs such as ChatGPT is now widely commonplace for discovery, research and even inspiration, using them to check out has proved less appealing. Following OpenAI's withdrawal of Instant Checkout, industry experts and retailers such as Etsy and Walmart (who made circa 200,000 products directly available in ChatGPT and allowed customers to provide their shipping and payment details to OpenAI in order to transact directly within the chat window) have shared perspectives on their learnings.

A customer asking ChatGPT to help them shop doesn't have a single-item mission – they have a basket mission.

The basket conundrum

Checkout in ChatGPT could not support multi-item carts, leaving customers worried that if they made five purchases across five transactions, they'd correspondingly receive five boxes.

A customer asking ChatGPT to help them shop doesn't have a single-item mission – they have a basket mission. Instant Checkout couldn't model that; it could only process individual transactions, one item at a time, with no awareness of what else the customer was buying, what might have already been in the retailer's main shopping basket, or what the delivery economics of the whole shop looked like.

This is also why the basket problem and the fulfilment data problem are part of the same challenge. An agent that can't see the whole basket can't calculate shipping thresholds or costs, can't apply loyalty benefits correctly or can't make the substitution decisions that preserve the mission. It can only do what it was told, item by item, in a very binary, synchronous way.

This absence of basket capability within ChatGPT suggests a basic category error: a solution that didn't fully understand what shopping actually is. Instant Checkout was retired without ever having supported the purchase of more than one item at a time. This rendered the capability as less of a functioning checkout and more of a vending machine.

Walmart's solution to this is to use their existing proprietary AI agent "Sparky". In this new experience, Walmart users log into Sparky the first time they encounter it in ChatGPT. Their basket from Walmart's website or app and within ChatGPT sync with one another in the hopes of better reflecting people's actual shopping habits. This solution is precisely the advocate architecture in practice: the declared principal is Walmart, the scope is the customer's full shopping context, and the fulfilment logic stays inside the retailer's systems, where, for the time being, it clearly belongs.

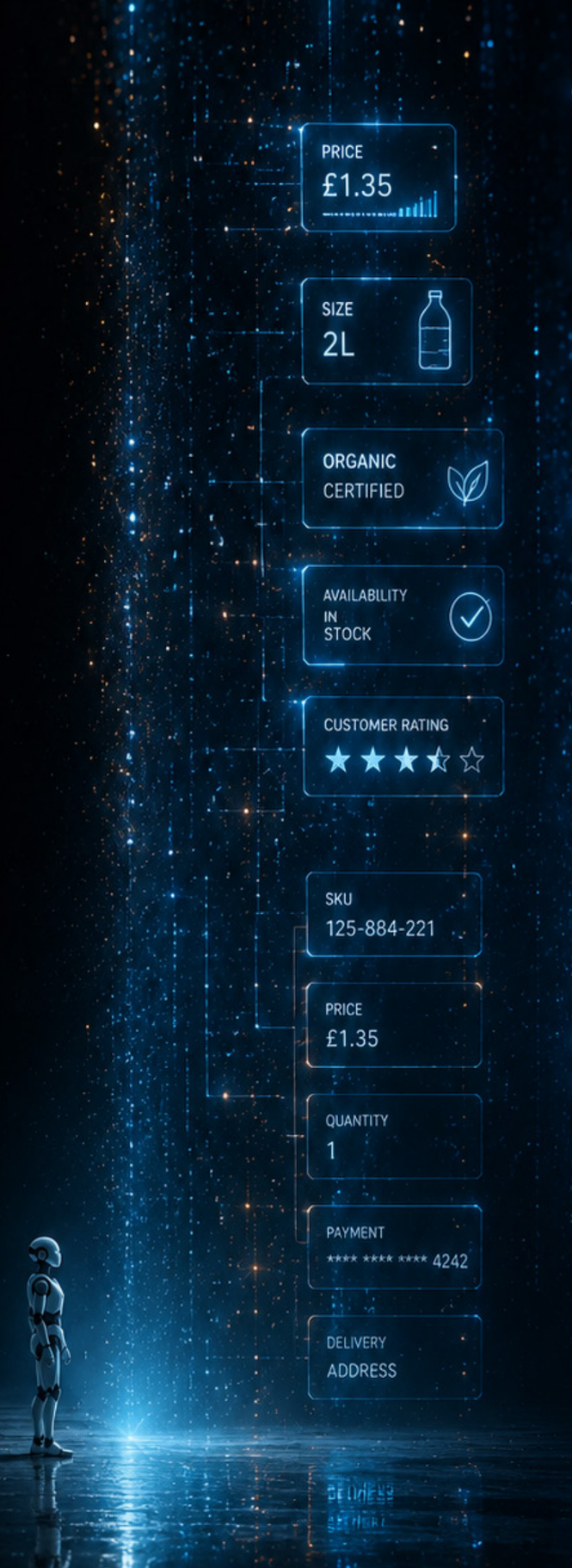
An agent without data is like retail without inventory

When a purchase is made through a conventional website, the context of that purchase is all around the customer. They've read the product page, seen the farm story and registered that it says "9 farms in the West Country" and "contributes £3m to the King Charles III Charitable Fund". That context shapes their decision, even if none of it formally enters the transaction itself. The transaction is largely just SKU, quantity, price, payment, and delivery address.

In an agentic commerce world, agents are being entrusted to make decisions, or at least narrow down choices, on your behalf. But the agents don't see the farm story; they queried (for example) for "organic semi-skimmed milk" against whatever data they can access, and the data it *can* access is mostly very structured product Attributes: price, size, organic certification (probably), availability, and possibly the customer rating. In this example, if the retailer's differentiating information, such as welfare standards, provenance, specific farms and charitable donations, only exists as prose in a description field or on an editorial content page, the agents either ignore it or can't measure it.

I use the phrase "travelling with the transaction" to describe this situation: is the information that's important to a customer present in a form where an agent, helping with discovery, narrowing down choices or even approving a purchase, can use it as decision input? Do we have structured attributes that an agent can read, compare and apply? e.g.

- Welfare_standard: free_range
- Origin: West_Country
- Certification: Duchy_Organic
- Charitable_contribution: True



A paragraph of brand copywriting does not align with this.

The practical consequence: an agent told to “recommend the best organic whole milk” will likely optimise on price per litre and a binary organic flag. Premium retailers will lose on price, and the entire premium positioning that justifies the price difference is invisible to the agent because it was never encoded as data: it was encoded as storytelling, which is exactly the right format for a human browser and exactly the wrong format for an agent.

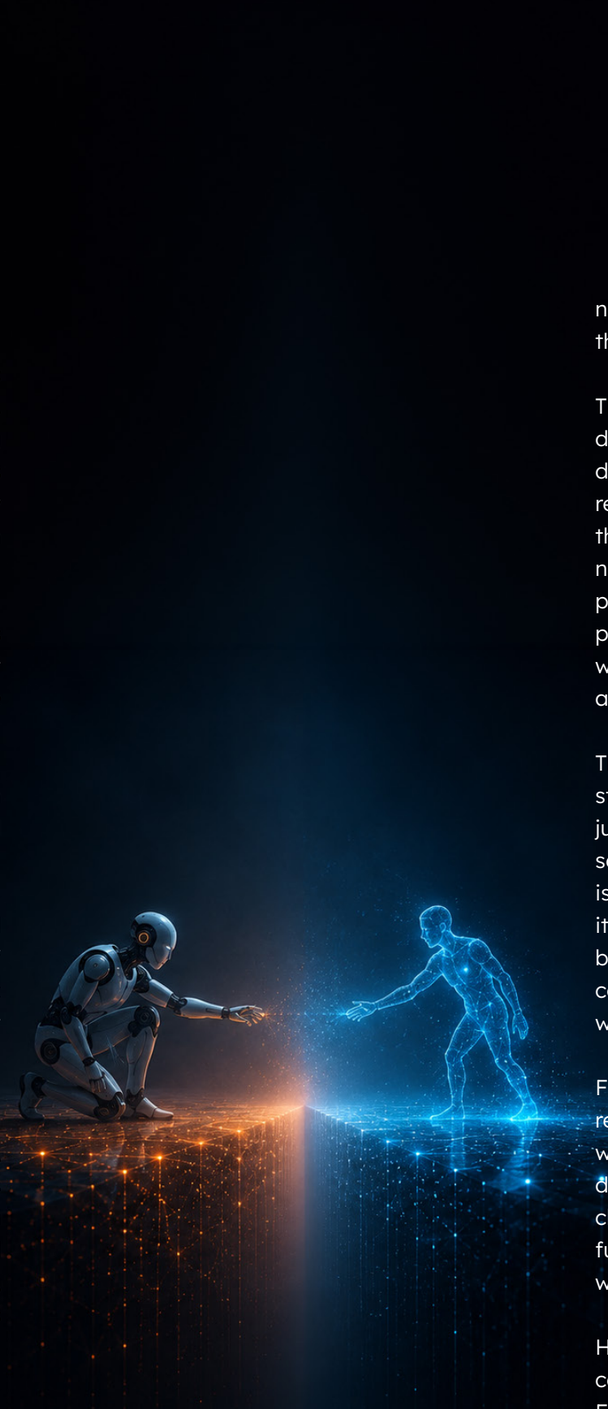
Inaccurate product results can even add the wrong item into user baskets for instant checkout, as happened in the Walmart/OpenAI Instant Checkout partnership[7] substitution, perhaps, without even notifying the customer.

This challenge with brand-level differentiation is closely linked to trust in some cases (in an agentic world, brands can no longer coast on emotional equity), and we’ll come to the notion of agentic trust shortly.

There is a dimension of the data problem that goes beyond quality - beyond whether data exists and whether it is accurate - that most discussions of agentic readiness overlook entirely. It is the problem of data semantics: whether the data means what you think it means, and whether an agent querying your systems will find the right answer to the right question.

Consider price. A retailer’s systems may contain the following, all legitimately described as “price” in one system or another: buying price, purchase price, selling price, retail price, current price, website price, was price, now price, sale price, wholesale price, etc. Each of these exists in a specific system for a specific purpose. Within that system, the field labelled “price” is unambiguous - the system only needed to know one kind of price, so it only stored one, and it called it price. No further explanation was required.

But an agent querying across systems has no way of knowing which price it is reading, because the field name carries no context beyond its own system. The column is called “price”. The agent reads “price”. What it does with that information depends entirely on which system it happened to query first, and there is no documentation, no metadata, no system of record that tells it the difference. The knowledge of what each price field actually represents typically lived in the heads of the people who built the system, or accumulated through years of operational experience. It was



never written down because it never needed to be - until something outside the system needed to understand it.

This is not a data quality problem, since the data itself is accurate. It is a data semantics problem: the meaning of the data is not encoded in the data itself, and an agent cannot infer it. A human analyst puzzling over a reporting tool would eventually work out which price field to use, probably through trial and error or by asking someone who knew. An agent does not ask - it reads what is there, applies what it finds, and moves on. The price it quotes may be the wholesale price, the historical retail price, or the promotional price from a campaign that ended last quarter. It will be stated with complete confidence, butit will be wrong in ways that are invisible until a customer complains or a transaction fails.

The practical implication is that agentic readiness requires not just structured data but documented data - fields with definitions attached, not just names. A product catalogue that tells an agent a field contains “current selling price inclusive of VAT, updated in real time from the pricing engine” is agent-readable. A field called “price” is not, regardless of how accurately it is populated. This is foundational work that most retailers have not done, because they’ve never previously needed to - the humans querying the data could work out what it meant. Agents cannot, and the cost of that ambiguity will show up in the transactions they complete.

From a data *hygiene* perspective, price and product descriptions are relatively straightforward to syndicate. Fulfilment data, including delivery windows, basket-level shipping thresholds, and loyalty-linked benefits, is deeply embedded in retailer systems and challenging to accurately expose cleanly through a third-party checkout layer. An agent that can’t access fulfilment data accurately will produce transactions that feel wrong even when the product selection was right.

How current is current? Even a correctly built Agentic Commerce Protocol catalogue feed represents a static snapshot, not a live inventory connection. Forrester described the absence of inventory management infrastructure as “disastrously absent from the plan”, suggesting prices change, items go out of stock, shipping costs vary by address, and any system relying on pre-ingested catalogue data will always be working with information that is at least partially wrong[8].

Standardisation is also an issue: merchant product information such as pricing, availability, and shipping options needs to be standardised for chatbots to be able to accurately and reliably access data.

The trust gap

Trust in transactions has been earned by retailers over time and cannot be immediately borrowed by platforms. Accordingly, customers trusted ChatGPT to help them decide, but they didn't fully trust it to spend their money. That distinction, that the agent earned the right to advise but never earned the right to act, is the agenticity/agentiality gap made commercially legible.

The technology solution supporting the implementation of the buy button on ChatGPT did not *fail:* the checkout flow functioned as intended, and the payment framework worked. What failed was the layer beneath all of that: the translation of customer mission into agent intent and agent intent into a transaction that felt right to the customer who initiated it. The system could process a purchase, but it could not understand *why* the customer was making it: what occasion they were shopping for, what constraints mattered, or what a good outcome actually looked like.

A quick note on Amazon

There is a further structural question worth noting, though it cannot be verified from public information alone. OpenAI's $15bn investment relationship with Amazon (announced in February 2026), with a further $35bn reportedly contingent on certain conditions being met, creates a commercial context in which OpenAI redirecting retail transactions away from Amazon would carry consequences. Whether or not that partly influenced the decision to withdraw Instant Checkout, the conflict of interest is structural. A platform with that kind of principal relationship cannot credibly position itself as a neutral advocate for the customer. Which is, in the taxonomy of the previous section, rather the point.

Joshua could genuinely launch a missile, never mind simulate it through gameplay. But he couldn't understand why that would have been catastrophic. OpenAI's agent could complete a checkout, but it couldn't understand why that wasn't enough.

This isn't a technology failure; it's a failure to truly understand and replicate how humans shop.

The solution Walmart and OpenAI reached is indicative. Rather than OpenAI owning the transaction, Walmart embedded its own chatbot, Sparky, inside ChatGPT. OpenAI became the window; Walmart remained the store. In other words, discovery happened in the agent layer, but trust, along with the transaction itself, remained with the retailer. That architecture has a name in this book: it is the advocate model, with declared principles and bounded scope. It took an unsuccessful experiment to arrive at this answer.

The strategic question every retailer must answer

Before protocols, before API endpoints, before any conversation with a vendor about agentic readiness, there is a prior question that most retailers are skipping.

If you are a retailer, is agentic commerce genuinely relevant to your business, your customers, and your product category? And if so, at what level of the progression, and on what timeline?

The answer is not the same across the board. A grocer whose customers shop habitually and frequently faces a different agentic reality than a furniture retailer whose customers buy once a decade after months of consideration. A brand whose differentiation lives in the shopping experience faces a different threat than one whose differentiation lives in price and availability.

What proportion of a retailer's revenue could realistically flow through agentic surfaces in three years? What would need to be true in the retailer's data, APIs, product information and customer relationships for that to happen? And what is the cost of being wrong in either direction: moving too fast into infrastructure that isn't needed or too slowly into a landscape that has already shifted beneath them?

These are the questions that precede the technical ones. Clearly in a different, fictional context, Joshua's creators skipped them. The consequences were nearly catastrophic.

Don't skip them.

Back to missions

At a retail conference I attended several years ago, Justin King, ex-CEO of Sainsbury's, was asked by an audience member how to reliably increase the number of customer transactions. His answer was telling: 'Don't try to.' Focus on the *interaction*, not the *transaction*. Interactions lead to transactions."

The interaction serves to resolve the mission. And the mission, as we've established, is rarely resolved in a single exchange.

When I had building work done on my house recently, the builder didn't simply execute the architect's diagrams despite having fully understood them, agreed to a payment plan, and had every capability necessary to proceed. Instead, he checked in throughout; he sought clarification; he offered improvements the original brief hadn't anticipated. The mission on paper and the mission in practice turned out to be different things, and only the interaction revealed the gap.

What's striking about King's observation is that it has now arrived at the infrastructure layer of retail from the opposite direction. The conversation about agentic commerce – about AI systems that handle discovery and shortlisting before the customer ever reaches a storefront - is converging on exactly the same point. If agents compress the top of the funnel, doing the filtering and comparison that customers once did by browsing, then the storefront can no longer rely on being the place where options are narrowed. It has to become the place where the interaction deepens. The funnel gets shorter upstream, and conversely, the conversation has to get richer downstream.

Decades apart and from entirely different starting points, retail instinct and infrastructure theory have arrived at the same place. This is what agentic commerce has to become: not a system that executes instructions and calls it 'done', but one that carries the mission faithfully enough to know when to check in, when to proceed, and when executing the instruction would betray the purpose behind it. The agent that gets this right stops being a tool you operate and becomes something closer to a representative.

That distinction is what the rest of this book is about.

RETAIL REINVENTS ITSELF. ALMOST ENTIRELY.

	CONTROL	CHOICE	DISCOVERY	DATA	INTENT	CUSTOMER INSIGHT	FULFILMENT	CONNECTIVITY	COMMERCE MODEL
OPEN ALL HOURS	Shopkeeper in control	Too much choice	Requesting	Talking		Surveys	In-store only	Offline	One channel
RETAIL THEN	Customer in control	Convenience	Browsing	Reporting		Sharing	Home delivery	Proprietary connections	Omnichannel
RETAIL NOW	Customer in control	Frictionless	Searching	Measuring		Surveillance capital	Where you want it	Everything everywhere	Unified commerce
RETAIL NEXT	Agents in control	Invisible	Discovering	Predicting		Known	Robots & drones	Everything now	Agentic commerce

INTENT - THE CONSTANT

CHAPTER 1:

KEY RETAILER TAKEAWAYS

Focus: Moving from transactional retail to mission-driven commerce.

1. Strategic questions for the boardroom

Perspective	The Question
Business	**Are we supporting a "Mission" or just a "Transaction"?** Can we define the "why" behind the "what" (e.g., "Getting the kids ready for school" vs. "Buying a backpack")?
Business	**If we're building an agent, who does our agent represent?** Are we an Advocate (serving the customer), a **Spy** (serving the brand covertly), or an **Actor** (acting on independent judgment)?
Business	**The "Double Agent"** Risk: If we build an agent that claims to help customers but is intended to steer them to our own high-margin products, will we lose their trust, and perhaps, their agent's access?
Technical	**Instruction vs. Intent:** Does our system solve the "Joshua Problem"? If an agent asks for "the safest car," do we provide a car with 5 stars or a tank? How do we ground "Safe" in our specific brand context?
Technical	**Agentiality vs. Agenticity:** Do we have the legal/permissioning framework to represent the customer (Agentiality) or just the technical "cool factor" to perform tasks (Agenticity)?

Keywords help people find you; Context helps agents choose you.

2. The leadership roadmap

Short-term (0-6 Months)

- **Business**: Identify the top 3 “High-Intent Missions” for your category. Stop optimizing for “Keywords” and start mapping “Contexts. (Keywords help people *find* you; Context helps agents choose you).
- **Technical**: Conduct a **Data Prose Audit**. Identify where product information is trapped in marketing copy (prose) and move it into structured metadata that an LLM can “reason” with.

Medium-term (6-18 Months)

- **Business**: Formally define your brand’s “Agency” boundaries. What decisions will you *never* let an agent make on behalf of a customer?
- **Technical**: Prototype an **Advocate Agent**. Build a “narrow” agent that focuses on one specific mission (e.g., “The Gift Finder”) to test how intent-mapping differs from traditional search.

Long-term (18+ Months)

- **Business & Technical**: Pivot the entire organization from “Buying Journeys” to “Mission Management.” Your success metric shifts from **Conversion Rate** towards **Mission Completion Rate**.

ABOUT THE AUTHOR

Paul Sims
Retail Service Lead and
Retail Technology Strategy Consultant
Equal Experts

Paul is a Retail Technology Strategy Consultant, who recently joined Equal Experts after spending the last 20+ years working for retailers across the globe, including positions as CTO at Halfords and Chief Architect at New Look, Primark, Marks & Spencer, and Argos in Shanghai.

LINKS

1. https://www.sciencedirect.com/science/article/abs/pii/074959789190020T
2. https://www.theaipraxis.com/agentic-commerce
3. https://openai.smapply.org/prog/agentic-ai-research-grants/
4. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4899709
5. https://arxiv.org/html/2508.05338v1
6. https://en.wiktionary.org/wiki/agenticity

APPENDIX 1:

5 LEVELS OF AGENTIC COMMERCE

Within Stripe's 2025 Annual Statement, they defined 5 levels of agentic commerce, which I've included in this appendix since they serve a useful reference point for evaluating how far agentic commerce has to go before it's truly autonomous. These levels are described not by what AI starts doing, but by what the human stops doing. As of Q1 2026, Stripe assessed the industry as currently hovering between levels 1 and 2. However, this assessment was superseded in April 2026. Amazon's launch of Scheduled Actions for Rufus, which provides persistent, scheduled, prompt-free shopping, represents the first mainstream commercial deployment at Level 3, with clear architecture pointing towards Level 5.

5 Levels of agentic commerce

Level 1: Form filling (Agent as clerk)

- **Description:** The AI completes checkout, filling in payment and shipping details.
- **Human Role:** The human researches, selects the product, and makes the buying decision.

Level 2: Descriptive search (End of keywords)

- **Description:** The user describes a situation (e.g., "back-to-school supplies for a year-3 child in Liverpool who hates itchy clothes"), and the agent reasons across options and provides search results.
- **Human Role:** The human evaluates options suggested by the agent.

Level 3: Persistence (Memory & preference)

- **Description:** The agent remembers user preferences, past purchases, sizes, and budget constraints across sessions.
- **Human Role:** The human stops re-introducing themselves, though they may still need to approve purchases.

Level 4: Delegation (Goal-setting)

- **Description:** The agent takes over the entire process—discovery, comparison, and purchase—based on high-level goals ("Get back-to-school shopping done under £400").
- **Human Role:** The human sets constraints and trusts the agent to weigh trade-offs.

Level 5: Anticipation (Ambient commerce)

- **Description:** There is no prompt. The agent knows the user's habits, calendar, and preferences and orders items just-in-time.
- **Human Role:** The human receives a notification that items have already been purchased.

EQUAL EXPERTS