

The Shadow Load

Who Pays for Agentic Travel Search?

Paper: WP-2026-17

Type: Position Paper

Date: June 2026

Author: Tim de Rosen, CEO and Co-Founder, AIVO Inc.

Series: OTA Cohort/ Agentic Brand Control

DOI: [10.5281/zenodo.20552188](https://doi.org/10.5281/zenodo.20552188)

POSITION PAPER

This paper sets out the theoretical frame for the agentic shadow load and the case for a paired diagnostic and remediation architecture. An empirical companion paper, currently in field, audits five OTA surfaces (Booking.com, Expedia, Trip.com, Google Free Booking Links, Hopper) under PSOS resilience testing and brand.context structural audit. The companion paper will cite this one.

Abstract

Agentic search consumes online travel agency infrastructure at scale without the booking signal that historically justified the serving cost. A single property detail page costs an agent four to twelve thousand input tokens to parse, with three to seven properties typically consumed per comparative query. The cost falls on the model provider in inference, on the OTA in serving, and on the user in latency. No party recovers value through the conventional booking funnel.

This paper defines the agentic shadow load, examines the inversion of finder and found in agent-mediated search, and proposes a paired diagnostic and remediation architecture. The AIVO Meridian PIM diagnostic, scored across RCS, PSOS, and CSR, identifies where product evidence is illegible to agents under conversational pressure. brand.context, a signed JSON-LD evidence layer, restores legibility and reduces token cost by an order of magnitude. Together they constitute the operational practice of Agentic Brand Control.

1. The shadow load

Every time an agent parses an OTA page to answer a hotel query, three parties exchange value and only one of them never sees a return. The user gets an answer. The model provider pays for inference. The OTA serves the bytes. The OTA bears infrastructure cost without recovering it through a booking funnel that no longer fires.

This is the agentic shadow load. It is small per query and large in aggregate.

2. The physics of the parse

A property detail page on a major OTA runs to several hundred kilobytes of rendered HTML, most of it React boilerplate, hydration state, tracking pixels, and component scaffolding. For a human eye, this resolves into a clean card with price, rating, and a booking button. For an agent counting tokens, it is a high-cost parse.

A single property page can consume four to twelve thousand input tokens before the model arrives at a usable summary. Comparative search is the dominant pattern. Agents typically consume three to seven properties per query. The inference cost per session climbs into the tens of cents, almost none of it recoverable through onward transaction.

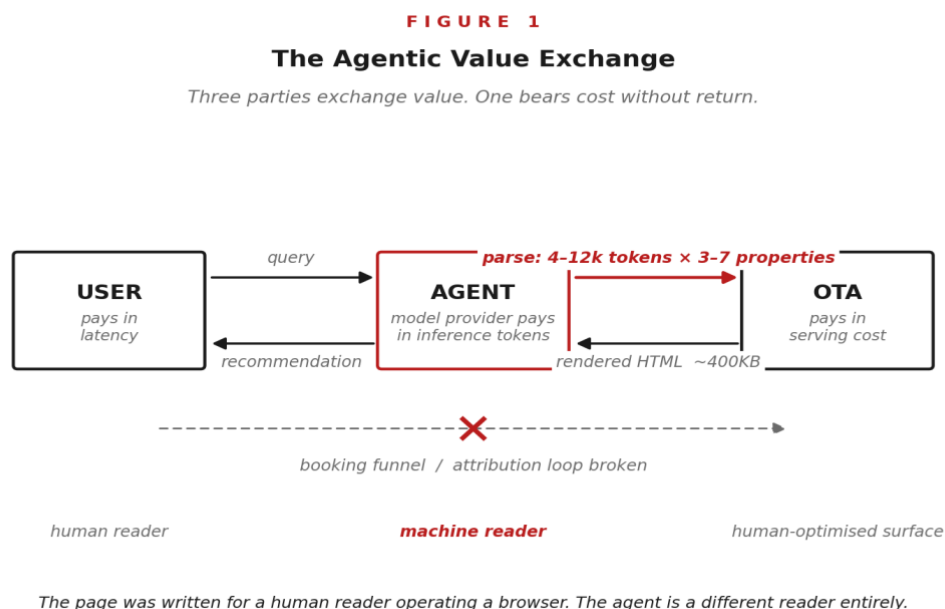


Figure 1. The Agentic Value Exchange. *Three parties exchange value in an agent-mediated search. The model provider pays in inference tokens. The OTA pays in serving cost. The user pays in latency. The conventional booking funnel and its affiliate attribution loop are bypassed.*

The OTA pays in compute and bandwidth to serve a page that was never optimised for this consumer. The model provider pays in tokens to parse a structure that was never designed for it to read. The user gets a recommendation and either books elsewhere or fails to book at all. The pricing and information loop, which the OTA category spent two decades engineering for human conversion, is bypassed.

3. The finder and the found

Two questions follow.

First, who is the finder and who is the found? Conventional affiliate logic says the finder pays the found, because the found is the destination and the finder is sending traffic. Agentic search inverts this. The finder, the agent, consumes the resource. The found, the OTA, pays in serving cost. No traffic arrives in the conventional sense, because the user session terminates inside the agent. The found is being mined, not visited.

4. The broken attribution loop

The inversion deepens when the agent does pass a deep link through to a booking page. The session that arrives at the OTA carries no reliable attribution. The standard mechanisms for tracking traffic origin, UTM parameters, affiliate cookies, referrer headers, were built for browser-to-browser transitions. They do not survive the handoff from a headless agent session to a clean browser tab.

The OTA records the booking as direct. The agent records the recommendation as completed. The commercial relationship between them remains unmeasurable. The shadow load is therefore compounded by an attribution blind spot. Cost is incurred upstream. Credit, where credit exists at all, is recorded downstream as ambient.

5. The blocking fallacy

A cynical reading would suggest the obvious response. Block agent traffic at the edge. Identify the user agents, rate-limit them, return 403s. The infrastructure cost disappears.

So does the brand.

An OTA that blocks agents drops out of the consideration set the agent presents to the user. The model has no view of inventory it cannot read. The brand becomes commercially invisible at the decision stage, regardless of how strong the underlying inventory is. Blocking solves the cost problem by accepting absence. It is not a defensive posture. It is a withdrawal.

The strategic surface is not access control. It is legibility.

6. The measurement gap

What does the OTA actually know about how legibly it is being read?

The shadow load is not a cost line in any current ledger because it is not yet instrumented. Page views do not register. Bookings do not arrive. Inference happens on someone else's compute. The OTA sees only absence.

The diagnostic question is therefore prior to the remediation question. Before an OTA can decide whether to publish a structured evidence layer, it needs to know which of its properties are being misread, where the model is reconstructing pricing incorrectly, where refundability claims are decaying across reasoning turns, and where the agent is silently substituting a competitor because the parse confidence dropped below threshold.

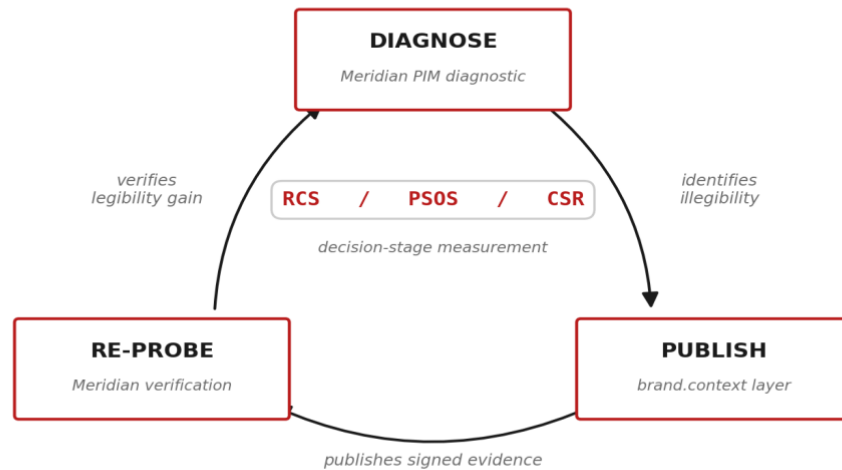
This is the work the AIVO Meridian PIM diagnostic was built to do. The diagnostic probes the model the way an agent would, sequence by sequence, and scores the product data layer on legibility. RCS, the Reasoning Chain Score, aggregates the decision-stage outcome into a single headline metric. PSOS, the Prompt Space Occupancy Score, scores the model's confidence in a brand's claim across five dimensions of breadth, depth, resilience, sentiment, and decay. CSR, the Conversational Survival Rate, measures whether the brand survives across multi-turn reasoning into the recommendation. Weak PSOS resilience is the leading indicator of CSR degradation. Together the three produce a map of where the OTA's evidence is structurally fragile when read by a machine.

What the diagnostic consistently surfaces, across categories, is that legibility failure is rarely a content problem. The information is on the page. The model cannot parse it cheaply enough, or with enough confidence, to carry it through to the recommendation. The page was written for a human reader operating a browser. The agent is a different reader entirely.

7. brand.context as remediation

brand.context is the paired remediation layer. A JSON-LD artefact designed for agent consumption reduces the parse from kilobytes to a few hundred bytes. A property's canonical claims, refundability terms, taxes, inclusions, availability windows, and verified ratings are expressed in a structured, signed, machine-legible payload that the agent retrieves before, or instead of, the human page. Token cost falls by an order of magnitude. Latency falls. PSOS resilience rises because the model is no longer reconstructing claims from page furniture. The OTA controls the canonical version of its own evidence rather than allowing the model to assemble one.

FIGURE 2
The Diagnose, Publish, Re-probe Loop
Meridian identifies illegibility. brand.context restores it. Meridian verifies.



The diagnostic and the remediation are paired. The closed loop is the discipline.

Figure 2. The Diagnose, Publish, Re-probe Loop. *Meridian's PIM diagnostic identifies where product evidence is illegible. brand.context publishes a signed, structured evidence layer that restores legibility. Meridian re-probes to verify gain. The decision-stage metrics at the centre (RCS, PSOS, CSR) measure the loop.*

The diagnostic and the remediation are paired. The Meridian PIM diagnostic identifies where evidence is illegible. brand.context is the layer that makes it legible. The closed loop is diagnose, publish, re-probe.

8. Agentic Brand Control as operating discipline

This pairing sits inside a larger operating frame. Agentic Brand Control is the discipline of governing how a brand is read, cited, and recommended by AI systems. It treats legibility as a defensive asset, evidence as a controlled surface, and decision-stage measurement as the missing layer between SEO and conversion analytics.

The OTA category is the first place this will be visible at scale, because the parse volume is enormous and the conversion gap is severe. It will not stay confined there.

The brand that is cheapest to parse correctly will increasingly become the brand that the agent recommends, because the model is optimising for confidence under token constraint.

The shadow load is the cost of being read badly.

Agentic Brand Control is the practice of not being read badly. brand.context is the instrument. AIVO Meridian is how you know whether it is working.

Citation

de Rosen, T. (2026). The Shadow Load: Who Pays for Agentic Travel Search? AIVO Journal Position Paper, WP-2026-17. AIVO Inc. DOI:**10.5281/zenodo.20552188**

ABOUT THE AUTHOR

Tim de Rosen is CEO and Co-Founder of AIVO Inc. AIVO Meridian is the company's AI decision-stage measurement platform. The AIVO Standard is the company's institutional framework for Agentic Brand Control, with working papers published under the AIVO Standard community on Zenodo (aivostandard.org).