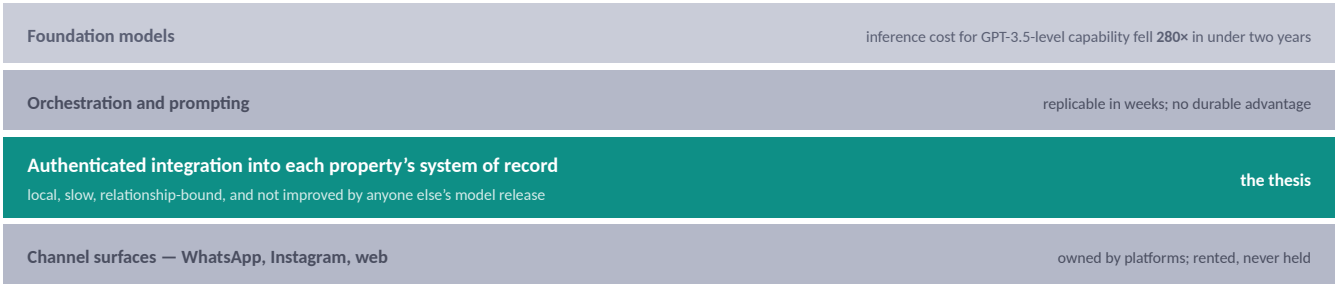
SECTOR THESIS · HOSPITALITY AI · 2026

Grounded Agents

Where the defensible layer sits in hospitality AI — and the four conditions under which this thesis is wrong.

THE STACK, AND WHERE MARGIN SURVIVES



FOUR NUMBERS BEHIND THE THESIS

14.5%
Booking Holdings blended take rate on $186bn gross bookings

−6.7pp
fall in European direct booking share over a decade

<50%
task success for frontier agents on a customer-service benchmark

2 Aug 26
EU AI Act disclosure obligation takes effect

The short version: the model is a falling-cost commodity, the channel is rented, and the only layer that compounds is the one nobody wants to build — a working connection to each hotel's own reservation system.

Sector thesis · September 2026 · Data current to Q3 2026
Audience: Angel and seed investors evaluating vertical AI in travel and hospitality
Sources: 33 cited references including SEC filings and investor releases from Booking Holdings and Expedia Group, Eurostat, the European Commission, HOTREC, Stanford HAI, NIST, EUR-Lex, and peer-reviewed work from *Marketing Science*, *ACM Computing Surveys*, and the proceedings of NeurIPS, ICLR, ICML and STOC
Published by: Vera — an interested party. See the disclosure in §2, and the falsification conditions in §11

CONTENTS

The argument, in order

HOW TO READ A THESIS PUBLISHED BY AN INTERESTED PARTY

This document argues that a particular layer of the hospitality AI stack is defensible. It is published by a company operating at that layer. That is a conflict, not a disqualification, and the correct response is to check the falsifiable parts.

So §11 states four specific conditions under which the thesis fails, each written to be checkable against public evidence rather than argued away. If you read only two sections, read §11 and §12.

SECTION 01

The thesis in eight points

Foundation models are a falling-cost commodity. Messaging channels are rented from platforms. The only layer in hospitality AI that compounds is the one nobody finds interesting: an authenticated, maintained connection into each property's own reservation system.

1 — Hotel distribution is a margin-loss problem of measurable size. Booking Holdings reported $26.9 billion of revenue on $186.1 billion of gross bookings in 2025 — a blended take rate of **14.5 per cent**. Expedia Group reported $14.7 billion on $119.6 billion, or **12.3 per cent**.[1,2] AUTHOR'S CALCULATION from company-reported figures. Neither is pure accommodation commission, so read them as blended rates.

2 — Direct booking share has eroded for a decade. HOTREC's European distribution study, surveying 3,089 hotels across 35 countries, puts direct bookings at **50.9 per cent** of overnight stays against 29.1 per cent for OTAs — down from 57.6 per cent direct in 2013. Over the same decade Booking Holdings' share of the European OTA market rose from 60.0 to **71 per cent**.[3] INSTITUTIONAL

3 — The model is not the moat, because the model is becoming cheap very fast. Stanford's AI Index reports that inference cost for a system performing at GPT-3.5 level fell **more than 280-fold** between November 2022 and October 2024, with hardware costs declining around 30 per cent annually and energy efficiency improving around 40 per cent.[4] INSTITUTIONAL A capability whose price falls by two orders of magnitude in two years is an input, not an advantage.

4 — Nor is orchestration. Prompt design, conversation flow and retrieval plumbing are replicable in weeks, and every model release improves them for every competitor simultaneously. Anything that gets better when someone else ships is not a moat.

5 — What does not commoditise is the authenticated connection to the system of record. It is built per vendor and frequently per property, requires credentials and operational trust, breaks when a schema changes, and is improved by no one's model release. It is slow, local and unglamorous — which is precisely why it accumulates.

6 — Reliability, not capability, is the product — and reliability is unsolved. τ-bench evaluates agents against real API tools and written policy in customer-service domains, scoring by whether the final database state matches an annotated goal. Its finding: "even state-of-the-art function calling agents (like gpt-4o) succeed on <50% of the tasks, and are quite inconsistent (pass^8 <25% in retail)."[5] PEER-REVIEWED An agent that works once is not a product. An agent that works every time is.

7 — Regulation pushes in the same direction. From **2 August 2026**, Article 50 of the EU AI Act requires that people be told they are interacting with an AI system.[6] A 2024 tribunal decision already attached an ordinary negligence standard to chatbot accuracy.[7] And a field experiment across 6,200+ customers found disclosure cut purchase rates by more than 79.7 per cent — because disclosed bots were perceived as *less knowledgeable*.[8]

PEER-REVIEWED Compulsory disclosure plus a knowledge penalty leaves demonstrable accuracy as the only lever.

8 — Two points against, stated here rather than buried. *First:* incumbent reservation-system vendors can ship grounded agents natively and own the integration by definition — they are the system of record. *Second:* the operational evidence in §8 comes from two properties under one ownership, with no control group and no revenue attribution. Neither objection is answered in this document. Both are set out as falsification conditions in §11.

SECTION 02

Method, disclosure and the dataset

Disclosure

This thesis is published by Vera, which operates at the layer it argues is defensible. Three things follow, and a reader should hold us to all of them.

No competitor is cited as a source of fact. Market-size estimates are labelled as commercial research and shown as a range across providers rather than a single number, because the providers disagree materially. And the thesis is stated with explicit falsification conditions in §11, so that it can be checked rather than believed.

Sourcing rules

Company financials come from filings and investor releases, not from press summaries. Take rates are derived from those figures and labelled as derived, with the caveat that total revenue includes advertising and other lines.

Technical claims come from refereed proceedings. Where a preprint is cited it is marked as one. Benchmark results are attributed to the benchmark by name, with the model generation stated, because agent benchmark numbers age in months.

Market sizing is treated as the weakest evidence in the document. Two reputable firms differ by roughly 19 per cent on the same base year for the same market, and by four points of compound growth rate (§9). Both are shown. Neither is adopted.

EVIDENCE CLASSIFICATION

PEER-REVIEWED
Journal or refereed proceedings.

OFFICIAL
Regulator, statistics agency, tribunal or securities filing.

INSTITUTIONAL
Trade body, standards body or university institute.

CALCULATION
Derived by the authors, inputs shown.

COMMERCIAL RESEARCH
Market-research firm or investor-firm benchmark. Weakest class in this document.

FIRST-PARTY
The publisher's own operational data.

WHAT IS DELIBERATELY ABSENT

No total addressable market figure. No revenue projection. No customer count. No comparison to a named competitor's performance. Each was considered and left out because the available evidence would not carry it.

The operational dataset

The figures in §8 are the publisher's own first-party operational data, drawn from a deployment across two hotel properties under common ownership, running for approximately seven months. Detailed metrics were measured over one 30-day window within that period, in which the system handled 60,611 guest messages.

WHAT THE DATASET CANNOT SUPPORT

There is no control group, so no causal claim about bookings or revenue. There is no revenue attribution, so no return-on-investment figure appears anywhere in this document. There is no independent audit. Two properties under one ownership is not a sample and supports no generalisation. It establishes that the architecture runs at volume with a low escalation rate. For a seed-stage thesis that is a relevant fact; it is not a proof of commercial value, and any investor treating it as one is over-reading it.

On the cost figures

The economics in §8 use the publisher's top-tier published price and an AI usage estimate computed for intensive use — long prompts, a capable model, heavy tool calling. It is an upper bound, deliberately chosen so that realised costs are more likely to fall below it than above.

A NOTE ON BENCHMARK AGEING

Agent benchmark results cited in §5 are from the model generations the papers tested. They are used to characterise the *shape* of the reliability problem — single-turn solved, multi-turn stateful not — rather than as current capability estimates. Anyone updating this thesis should re-run the check.

SECTION 03

What hotels actually spend money losing

Any thesis about selling software to hotels has to start from where hotel margin currently goes. It goes to distribution, and the share has been rising for a decade.

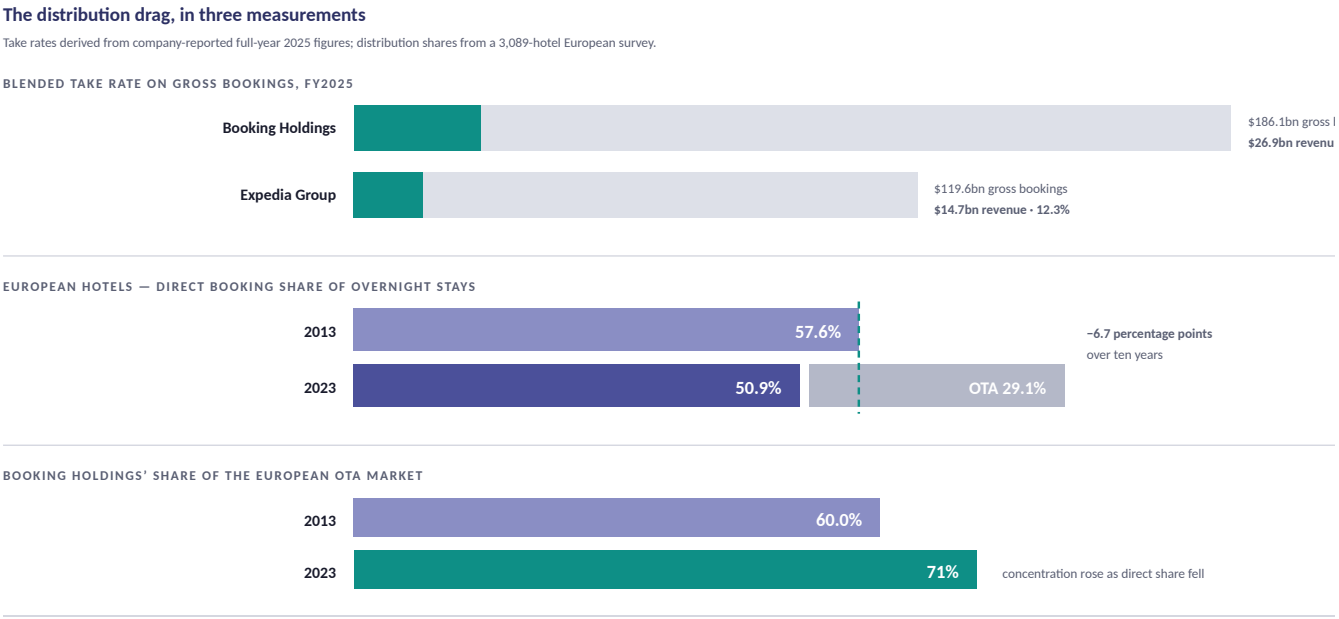


Figure 1 — The distribution drag. Gross bookings and revenue: Booking Holdings Q4 2025 earnings release[1] and Expedia Group Q4/FY2025 release as filed with the SEC.[2] OFFICIAL Take rates are AUTHOR'S CALCULATION and are *blended* — total revenue includes advertising and other lines, so they are not pure accommodation commission. Distribution and market shares: HOTREC European Hotel Distribution Study 2024 (reference year 2023; 2,537 individual hotels across 35 countries plus 552 chain properties).[3]

Neither OTA publishes a headline commission rate. Booking.com's own partner documentation states only that "the commission percentage varies by country and can also vary depending on your property type or location."[9] That is why this report uses derived blended take rates rather than the "15 to 25 per cent" range that circulates in trade commentary, for which we could find no authoritative basis.

The regulatory environment has moved. Booking.com was designated a gatekeeper under the Digital Markets Act on 13 May 2024, with full compliance required from 14 November 2024; parity clauses are prohibited and the platform "must not introduce other measures with the same effect."[10] OFFICIAL Removing rate-parity constraints makes a direct channel economically expressible for the first time in years — which is the demand condition this thesis depends on.

THE ONE FORECAST IN THIS DOCUMENT, AND WHOSE IT IS

Skift Research projects direct digital channels overtaking OTAs in hotel gross bookings by 2030, at $409 billion against $333 billion.[11] COMMERCIAL RESEARCH It is a forecast by a travel research firm, not a measurement, and it is the only forward-looking number in this document. It is included because it names the direction of the thesis; it should carry no weight in an investment decision beyond that.

SECTION 04

Why the model is not the moat

An input whose price falls by two orders of magnitude in two years is not a defensible asset. It is a reason to be somewhere else in the stack.

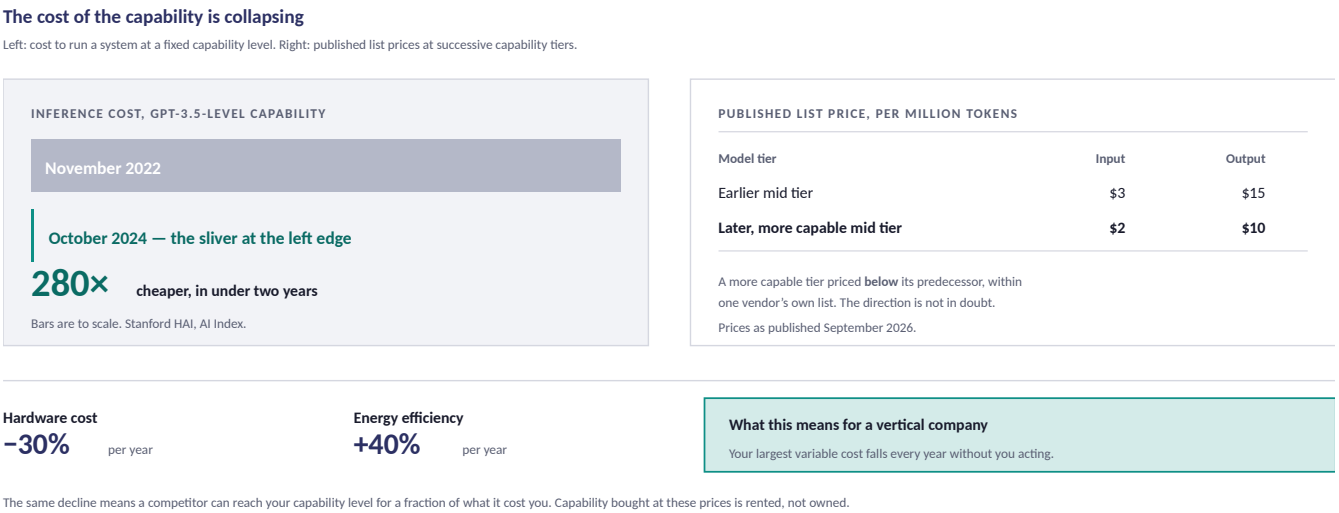


Figure 2 — The cost curve. Inference cost decline, hardware cost and energy efficiency figures: Stanford HAI, *AI Index Report*.[4] INSTITUTIONAL List prices are from a model provider's published pricing page as at September 2026 and are shown to illustrate within-vendor direction, not to compare vendors.[12]

The strategic reading is straightforward. A 280-fold decline in the cost of a capability is excellent news for anyone *using* it and terrible news for anyone whose business is *having* it. It means the variable cost line in §8 falls each year without management action — and equally that any competitor can reach the same capability for a fraction of what it cost eighteen months earlier.

The same logic disposes of orchestration as a moat. Prompt engineering, conversation design and retrieval plumbing are weeks of work, widely documented, and improved for everyone whenever a model ships. A survey of augmented language models in *Transactions on Machine Learning Research* describes the whole pattern — models calling external modules to extend what they can do — as an established research direction rather than a proprietary technique.[13] PEER-REVIEWED Anything a survey paper can describe is not a secret.

SECTION 05

Reliability, not capability, is the product

The benchmark that matters for a commercial agent is not whether it can do the task. It is whether it does the task the eighth time.

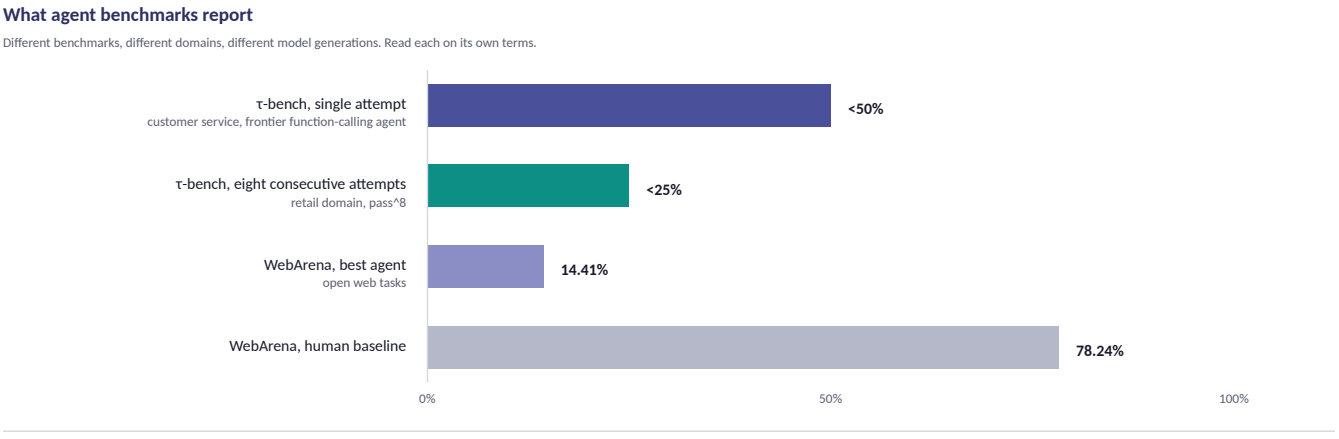


Figure 3 — Agent benchmark results. τ-bench: Yao, Shinn, Razavi & Narasimhan, ICLR 2025 — agents evaluated against real API tools and written policy documents, scored by whether the final database state matches an annotated goal.[5] WebArena: Zhou et al., ICLR 2024.[14] PEER-REVIEWED Both report results for the model generations tested at publication and will have moved; they are cited for the structure of the problem, not as current capability.

The Berkeley Function Calling Leaderboard, published at ICML 2025, reaches the same conclusion from a different direction: "while state-of-the-art LLMs excel at single-turn calls, memory, dynamic decision-making, and long-horizon reasoning remain open challenges."[15] PEER-REVIEWED A booking conversation — dates, occupancy, child ages, room split, policy, payment — is exactly multi-turn, stateful and long-horizon.

Models also hallucinate the tool call itself, not merely the prose. Patil and colleagues note that even strong models show an "inability to generate accurate input arguments and their tendency to hallucinate the wrong usage of an API call."[16] And factual reliability, measured across current models on a 6,000-question benchmark, ranges from a 22 to a 94 per cent hallucination rate.[17] INSTITUTIONAL

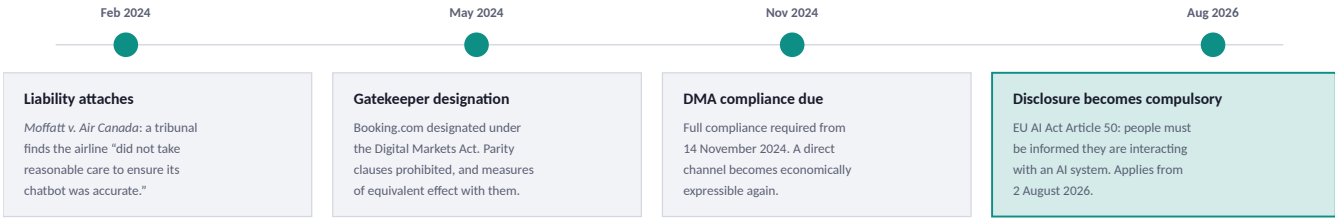
For an investor the implication is specific and it is not a technology bet. The capability is available to everyone at falling cost. The difference between a demo and a deployment is the operational work that closes the gap between pass@1 and pass^8: constrained action scope, validated arguments, refusal behaviour when the record is silent, escalation paths, and instrumentation that detects the failure before the guest does. That work does not appear in a model release. It accumulates in a codebase and in a set of live integrations.

SECTION 06

The regulatory forcing function

Three separate developments between 2024 and 2026 push the market toward systems that can demonstrate where their answers came from. None was designed to; together they are decisive.

Three developments, one direction



AND THE MEASURED COST OF DISCLOSURE

−79.7% **purchase rate when chatbot identity was disclosed — field experiment, 6,200+ customers, 2019**

The mechanism the authors identify: customers perceived the disclosed bot as *less knowledgeable*. That is the variable grounding changes, and the only one left once disclosure is mandatory.

Figure 4 — The regulatory forcing function. Tribunal decision: *Moffatt v. Air Canada*, 2024 BCCRT 149.[7] DMA designation and compliance: European Commission.[10] EU AI Act Article 50: Regulation (EU) 2024/1689.[6] OFFICIAL Disclosure penalty: Luo, Tong, Fang & Qu, *Marketing Science* 38(6), 2019.[8] That study covers outbound sales calls in China with pre-LLM chatbots; the magnitude will not replicate, and only the mechanism is carried forward here.

For governance buyers, NIST's AI Risk Management Framework organises practice into four functions, of which MEASURE "employs quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts."[18] INSTITUTIONAL An ungrounded assistant cannot satisfy MEASURE, because there is no reference against which to benchmark an answer. That is a procurement blocker in any enterprise with an AI governance policy — and the share of businesses with no responsible-AI policy at all fell sharply between the last two AI Index editions.[17]

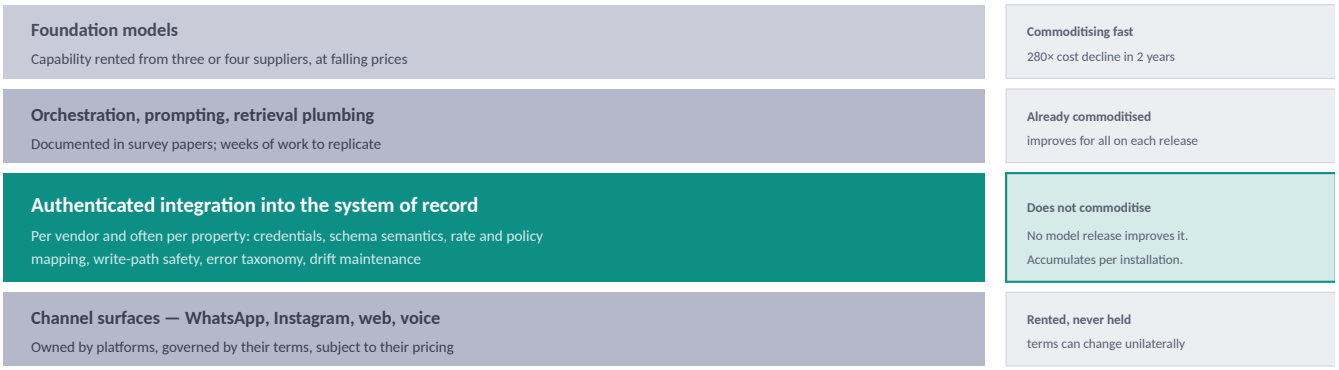
SECTION 07

The integration layer as the defensible asset

The part of this business that is hard to build is the part nobody writes about: getting a working, authenticated, maintained connection into each hotel's own reservation system, and keeping it working.

Where value accrues, and what commoditises

Four layers. Three of them get cheaper or easier for every participant simultaneously.



The test for any layer: does it get better when a competitor ships something? If yes, it is not a moat.

By that test only the third layer qualifies — and it is the layer that looks least like a technology business and most like field engineering.

This is an analytical framework, not a finding from any cited source.

Figure 5 — Value accrual across the stack. Authors' framework. The cost-decline figure is from Stanford HAI;[4] the characterisation of orchestration as documented practice follows the augmented-language-model survey literature.[13] The layer assessment itself is argument, not evidence.

Why the integration is slow, specifically

Anatomy of a grounded integration

Three cost classes. Only the first amortises across customers.

ONCE PER PMS VENDOR	ONCE PER PROPERTY	CONTINUOUSLY, FOREVER
· Authorisation flow and token lifecycle	· Credentials issued by the hotel	· Schema drift — breaks silently
· Schema mapping: room type, rate plan, inventory, occupancy rules	· Network authorisation — IP allow-listing routinely blocks foreign-hosted callers	· Credential and token rotation
· Pricing semantics: child-age brackets, minimum stay, multi-room distribution	· That property's own rate plans mapped	· Seasonal policy and rate-plan changes
· Cancellation policy representation	· Its policies verified against what staff actually tell guests	· PMS version upgrades
· Write-path safety for a booking	· Test bookings through to the record	· Quote-to-record match monitoring
· Error taxonomy — distinguishing "no availability" from "call failed"	· Cutover with a human fallback running	· Escalation-rate monitoring
		· Incident response when the reservation system itself is down

The first column amortises across every customer on that PMS. The second and third do not — and they are why this is not a weekend project.

No duration is claimed for any of these steps, because we have no sourced basis for one. The argument is about the *count and nature* of the steps, not their length.

Authors' description of the work, drawn from operating the deployment in §8. Not a finding from any cited source.

Figure 6 — Anatomy of a grounded integration. Authors' own account of the work involved, presented as argument rather than evidence. It is included because the shape of this work is what the thesis rests on, and a reader should be able to judge it directly.

Two properties of this layer matter for an investment case. It is **cumulative**: each completed PMS integration lowers the cost of the next property on that system, so the unit economics of installation improve with density rather than with scale in the abstract. And it is **trust-gated**: a hotel granting production write access to its reservation system is making an operational decision, not a purchasing one, which lengthens the sales cycle and correspondingly raises the switching cost once made.

SECTION 08

Unit economics from the pilot

Two properties, approximately seven months, 60,611 guest messages in the measured window. Here is what that supports, costed on the least favourable assumption available.

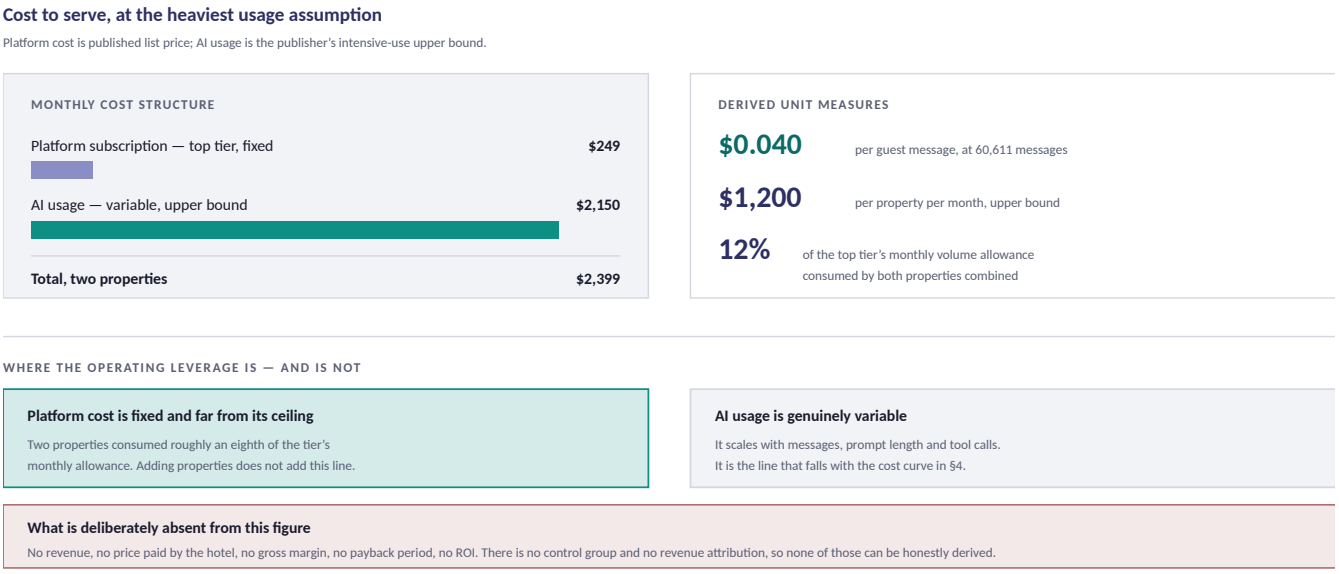


Figure 7 — Cost to serve. Platform price from the publisher's published list; AI usage from the publisher's intensive-use estimate. FIRST-PARTY Per-message, per-property and allowance-consumption figures are AUTHOR'S CALCULATION at the measured 60,611-message volume across two properties. The AI usage figure assumes long prompts, a capable model and heavy tool calling; a lighter configuration costs materially less.

The structurally interesting fact is the split. The fixed line is small and has headroom; the variable line is the one that matters, and it is precisely the line falling with the cost curve in §4. A company at this layer has a cost base whose dominant component gets cheaper without management action, against a customer base whose switching cost rises with every integration completed.

What this does **not** tell an investor is anything about willingness to pay, gross margin or payback, because no revenue figure appears anywhere in this document. A pilot with no control group cannot produce one honestly, and a document that produced one anyway would be telling you more about its author than about the business.

SECTION 09

Market structure and addressable base

Establishment counts are official and reliable. Market-size estimates are not, and this section shows why by putting two reputable providers side by side.

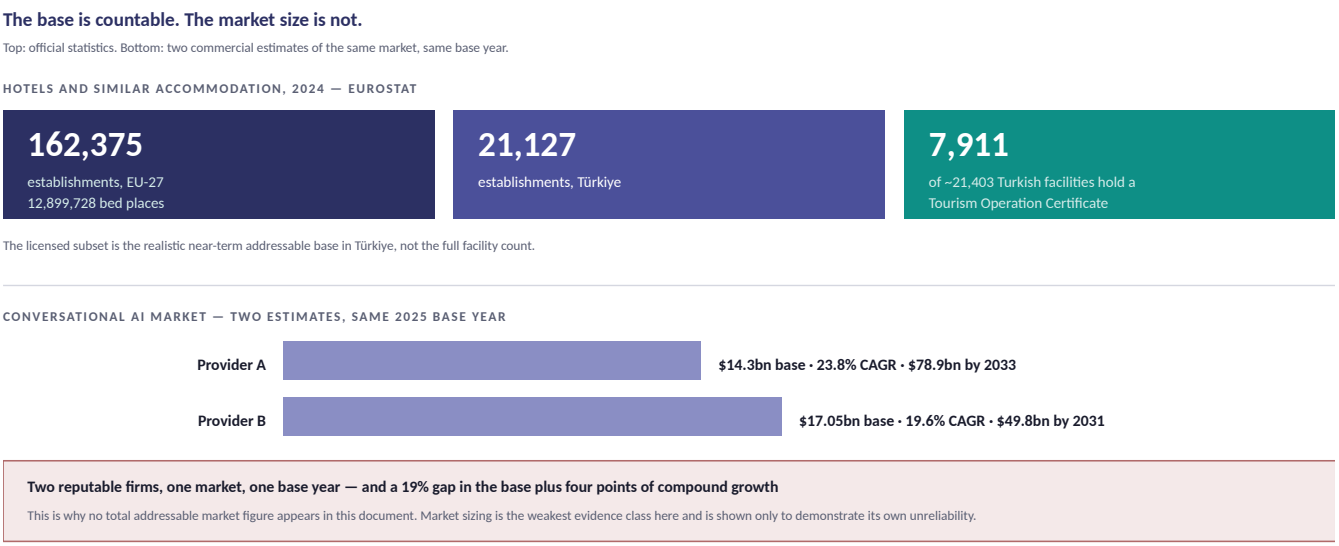


Figure 8 — Addressable base and market-sizing divergence. Establishment and bed place counts: Eurostat dataset `tour_cap_nat`, NACE I551, 2024.[19] OFFICIAL Turkish licensed-establishment split: trade press reporting a finalised official list; the issuing body was not confirmable from the page, so it is attributed to the publication.[20] Market estimates: two commercial market-research providers, both for a 2025 base year.[23,24] COMMERCIAL RESEARCH

For context on the operator base, HOTREC characterises European hospitality as roughly two million companies, 99 per cent of them SMEs, employing 11.5 million people and contributing about 3 per cent of EU GDP — a figure that covers hotels, restaurants and cafés together, and which the source publishes without a reference year.[21] INSTITUTIONAL The relevant structural fact for a software business is the SME concentration: a market of small independent operators is one where per-property integration cost dominates, and where density on a given reservation platform matters more than headline market size.

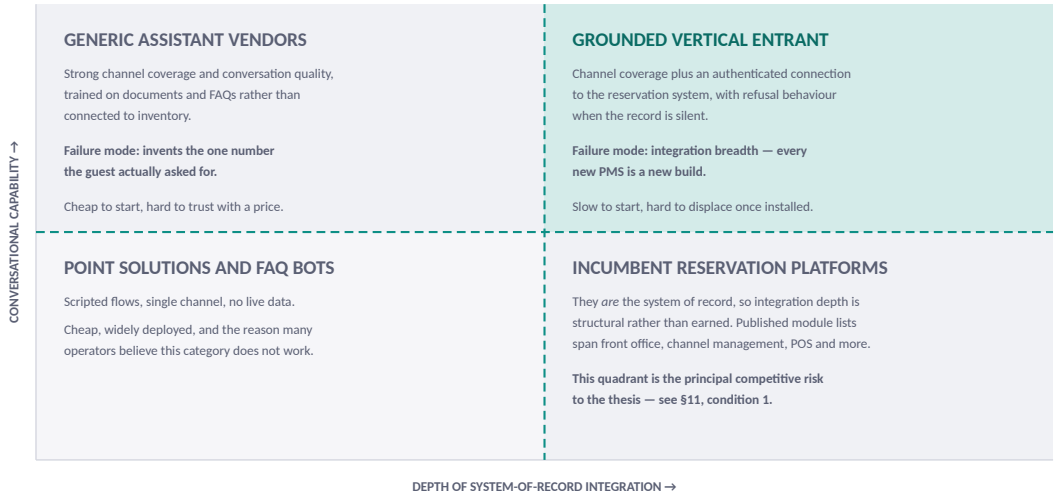
For the hotel software market specifically, one provider puts hotel and hospitality management software at $4.2 billion in 2025 growing to $8.4 billion by 2033 at a 9.8 per cent compound rate.[22] That is a single estimate from a single provider and should be treated accordingly.

SECTION 10

Competitive structure

Where a vertical entrant sits

Horizontal: depth of integration into the property's system of record. Vertical: conversational and channel capability.



Authors' framework. Quadrant placement is analytical, not measured, and no company is positioned by name.

Figure 9 — Competitive map. Authors' framework; no company is placed on it. Module scope for incumbent platforms is drawn from vendor self-description: one Turkish platform publicly claims more than 64,000 accommodation partners across 100+ countries,[25] another claims more than 5,000 references and describes itself as market leader.[26] VENDOR CLAIMS, UNAUDITED Neither claim is verified here and neither is used as evidence for anything.

The competitive question a seed investor should ask is not whether a vertical entrant can build better conversational quality than a generic vendor — it can, and so can the next entrant. It is whether the incumbent reservation platforms will ship this natively before the entrant accumulates enough integration density to matter. That is a race, not a moat, and §11 treats it as the primary risk.

What is observable today: the incumbent platforms publish broad module lists spanning front office, reservations, channel management, restaurant and spa, accounting and procurement — and conversational AI is not prominent among the capabilities they advertise.[25,26] That is evidence about their marketing, not about their roadmap, and it should be weighted as such.

The benchmark reality check on any growth assumption

Private B2B SaaS growth has compressed. Median growth across more than 1,000 private B2B SaaS companies was **25 per cent in 2024**, down from 30 per cent the prior year, and falls with size: 40 per cent below $1 million ARR, 24 per cent at $3–10 million, 20 per cent above $10 million.[27] INVESTOR-FIRM RESEARCH Net revenue retention correlates with contract value: at $25,000–$50,000 annual contract value the median is **102 per cent**.[28] A hospitality SME product will sit at or below that band, so a plan modelling 120 per cent-plus retention is modelling a different business. A separate benchmark set puts median gross margin at 77 per cent and median net revenue retention at 101 per cent, with customer acquisition cost payback lengthening.[29]

SECTION 11

Risk register and falsification

Four conditions under which the thesis in §1 is wrong. Each is written to be checked against public evidence rather than argued about.

What would make this thesis wrong

	CONDITION	WHAT YOU WOULD OBSERVE	STATE AS AT Q3 2026
1	**Incumbent reservation platforms ship it natively** They own the record, so their integration cost is zero.	A major PMS or channel manager announcing native conversational booking against its own inventory, bundled at no incremental price.	Not prominent in what the large Turkish platforms currently advertise. This is **the primary risk.**
2	**Reliable generic agents plus a universal connector standard** Integration collapses from engineering to configuration.	τ-bench-class pass^k scores approaching reliability thresholds, plus broad adoption of a connector protocol by hospitality systems.	pass^8 below 25% in the published retail domain. Not close yet — but this is the fastest-moving variable here.
3	**Channel platforms disintermediate** The conversation and the booking both move on-platform.	A messaging platform offering native booking against hotel inventory, with its own commercial terms attached.	Business AI conversations on one platform went from 1m to 10m per week in a single quarter of 2026. Watch closely.
4	**Demand does not materialise** Hotels keep paying the take rate rather than investing in a direct channel.	European direct booking share continuing to fall in the next HOTREC distribution study, despite parity clauses now prohibited.	Direct share fell 6.7 points over the decade to 2023. The post-DMA reading is the one that will settle this.

And the evidence risk, which is ours rather than the market's

The operational evidence in §8 comes from two properties under one ownership, over roughly seven months, with no control group, no revenue attribution and no independent audit. It shows the architecture runs at volume. It does not show that anyone will pay for it, and no figure in this document claims otherwise.

Figure 10 — Falsification conditions. Conditions and observations are the authors' framework. State assessments draw on the benchmark results in §5,[5] the platform disclosure in §3,[30] the HOTREC distribution series,[3] and vendor self-description.[25,26] Assessments of what competitors will do next are judgements, not evidence.

Condition 1 deserves the most weight and gets the least attention in pitches of this kind. An incumbent that already holds the reservation record faces none of the integration cost described in §7 — for them it is an internal API call. The only structural answers available to an entrant are speed of installation across many platforms, a channel and conversation layer the incumbents have not built, and the operational discipline the benchmarks in §5 say separates a demo from a deployment. Whether those are sufficient is the investment question, and this document does not claim to settle it.

Condition 4 has a second reading worth noting. Research on US branded hotels found loyalty members booking direct delivered roughly a 9 per cent premium over OTA bookings in average daily rate net of booking costs, with OTA bookings consuming around 15 per cent of guest-paid rate in routine booking costs against around 8 per cent for loyalty direct.[31] COMMERCIAL RESEARCH That study covers US data from 2016 to 2018 and is dated; it is cited because it is the canonical reference for the economics of a direct booking, and because no current equivalent could be located.

SECTION 12

Limitations

1. **This thesis is published by an interested party.** Everything in §1 is argued by a company operating at the layer it calls defensible. The falsification conditions in §11 exist so the argument can be

tested rather than trusted.

2. **The operational dataset is two properties under one ownership.** No control group, no revenue attribution, no independent audit, one market, seven months. It supports an existence claim about architecture at volume and nothing else.
3. **Take rates are derived and blended.** Total revenue at both OTAs includes advertising and other lines, so 14.5 and 12.3 per cent are not accommodation commission rates. No authoritative published commission rate exists — Booking.com's own partner documentation declines to state one.[9]
4. **Market sizing is unreliable and is treated as such.** Two reputable providers differ by roughly 19 per cent on the same base year and four points of compound growth (§9). No total addressable market figure appears anywhere in this document, and none should be inferred from the establishment counts, which measure a different thing.
5. **Agent benchmark figures age in months.** The τ-bench and WebArena results in §5 are for the model generations those papers tested. They characterise the shape of the reliability problem — single-turn solved, multi-turn stateful not — and should be re-checked before being relied on as current.[5,14]
6. **The disclosure-penalty figure is from 2019 and will not replicate.** It covers outbound sales calls in China with pre-LLM chatbots.[8] Only the mechanism — that the penalty ran through perceived lack of knowledge — is carried forward, and even that is an inference.
7. **EU AI Act Article 50 does not apply until 2 August 2026**, so no evidence exists yet about how compulsory disclosure changes behaviour in practice.[6] Everything said about its commercial effect is prospective.
8. ***Moffatt v. Air Canada* is a small-claims tribunal decision** with an award of CAD $812.02, not binding appellate precedent.[7] Its weight is doctrinal.
9. **Vendor claims about customer bases are unaudited.** The partner and reference counts in §10 are company self-description, reproduced to characterise the competitive landscape and not used as evidence for any conclusion.[25,26]
10. **The integration anatomy in §7 is our own account of our own work.** It is argument presented as argument. No duration, cost or difficulty figure is attached to it, because we have no sourced basis for one and an unsourced one would be exactly the kind of number this document criticises elsewhere.
11. **SaaS benchmarks are investor-firm and vendor-published research**, not audited data, and one of the two sets used discloses neither sample size nor methodology.[27,28,29]
12. **The causal chain underlying the thesis is a synthesis across separate literatures.** Grounding reduces hallucination,[32,33] hallucination creates legal exposure,[7] and disclosure suppresses conversion unless the system seems knowledgeable[8] are three findings from three unrelated fields. Assembling them is our argument, not a published result, and nobody has tested the chain end to end.

> **WHAT A DILIGENCE PROCESS SHOULD ASK FOR THAT IS NOT HERE**
>
> Revenue and pricing realised per property. Gross margin after support cost. Installation time and cost per property, measured rather than described. Retention across a cohort longer than seven months. Escalation rate and quote-to-record match rate over time rather than at a point. Every one of those is a fair question, and none of them is answered in this document.

SECTION 13

References

All sources accessed and verified September 2026. Company financials are cited to filings and investor releases; market-research estimates are labelled as such.

1 **Booking Holdings**, Q4 and full year 2025 earnings release. bookingholdings.com › Q4-25-BKNG-Earnings-Release.pdf

2 **Expedia Group**, Q4 and full year 2025 earnings release, filed with the SEC. sec.gov › expedia earningsrelease-q42025

3 **HOTREC Hospitality Europe**, *European Hotel Distribution Study 2024*, conducted by HES-SO Valais-Wallis; reference year 2023; 3,089 hotels. HOTREC-Distribution-Study-2024.pdf

4 **Stanford Institute for Human-Centered AI**, *AI Index Report* — inference cost, hardware cost and energy efficiency trends. aiindex.stanford.edu/report

5 **Yao, S., Shinn, N., Razavi, P. & Narasimhan, K.** (2025), "τ-bench: A Benchmark for Tool-Agent-User Interaction in Real-World Domains", *ICLR 2025*. arxiv.org/abs/2406.12045

6 **European Union**, Regulation (EU) 2024/1689 (Artificial Intelligence Act), Article 50; obligations apply from 2 August 2026. eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

7 ***Moffatt v. Air Canada***, 2024 BCCRT 149, British Columbia Civil Resolution Tribunal, 14 February 2024. canlii.org › 2024bccrt149

8 **Luo, X., Tong, S., Fang, Z. & Qu, Z.** (2019), "Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases", *Marketing Science* 38(6), 937–947. DOI 10.1287/mksc.2019.1192. pubsonline.informs.org/doi/10.1287/mksc.2019.1192

9 **Booking.com for Partners**, Understanding our commission. No rate is disclosed. partner.booking.com › understanding-our-commission

10 **European Commission**, "Booking must now comply with the Digital Markets Act". digital-strategy.ec.europa.eu › booking-must-now-comply-digital-markets-act

11 **Skift Research**, Hotel Distribution Outlook 2024, reported November 2024. **Commercial research forecast.** skift.com › direct-bookings-to-lead-by-2030

12 **Anthropic**, published model pricing, accessed September 2026. platform.claude.com/docs/en/about-claude/pricing

13 **Mialon, G. et al.** (2023), "Augmented Language Models: a Survey", *Transactions on Machine Learning Research*. arxiv.org/abs/2302.07842

14 **Zhou, S. et al.** (2024), "WebArena: A Realistic Web Environment for Building Autonomous Agents", *ICLR 2024*. arxiv.org/abs/2307.13854

15 **Patil, S.G. et al.** (2025), "The Berkeley Function Calling Leaderboard", *ICML 2025*, PMLR 267, 48371–48392. proceedings.mlr.press/v267/patil25a.html

16 **Patil, S.G., Zhang, T., Wang, X. & Gonzalez, J.E.** (2024), "Gorilla: Large Language Model Connected with Massive APIs", *NeurIPS 2024*. arxiv.org/abs/2305.15334

17 **Stanford Institute for Human-Centered AI**, *AI Index Report 2026*, Chapter 3: Responsible AI — hallucination rates on AA-Omniscience and responsible-AI policy adoption. hai.stanford.edu › ai_index_report_2026_chapter_3_responsible_ai.pdf

18 **NIST**, *AI Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1. DOI 10.6028/NIST.AI.100-1. nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

19 **Eurostat**, dataset `tour_cap_nat`, hotels and similar accommodation (NACE I551), establishments and bed places, 2024. ec.europa.eu/eurostat/databrowser/view/tour_cap_nat

20 **Turizm Günlüğü**, Turkish accommodation facility and tourism operation certificate counts, December 2025. Issuing body not confirmable from the page; attributed to the publication. turizmgunlugu.com › turizm-isletmesi-belgeli-konaklama-tesisi

21 **HOTREC Hospitality Europe**, Facts and Figures. No reference year stated on the page. hotrec.eu/en/facts_figures.html

22 **Grand View Research**, Hotel and hospitality management software market report. **Commercial research.** grandviewresearch.com › hotel-hospitality-management-software-market-report

23 **Grand View Research**, Conversational AI market report. **Commercial research.** grandviewresearch.com › conversational-ai-market-report

24 **MarketsandMarkets**, Conversational AI market. **Commercial research.** marketsandmarkets.com › conversational-ai

25 **HotelRunner**, company About page. **Unaudited vendor claim.** hotelrunner.com/en/about

26 **Elektraweb (Talya Bilişim)**, hotel software product page. **Unaudited vendor claim.** elektraweb.com/otel-yazilimi

27 **SaaS Capital**, 2025 B2B private SaaS company growth rate benchmarking (1,000+ respondents, 2024 performance). **Investor-firm research.** saas-capital.com › 2025-B2B-Benchmarking-Growth-Rates.pdf

28 **SaaS Capital**, retention rates for private SaaS companies by contract value. **Investor-firm research.** saas-capital.com › good-retention-rate-private-saas-company

29 **Benchmarkit**, 2025 SaaS Performance Metrics. Sample size and methodology not disclosed. **Commercial benchmarking.** benchmarkit.ai/2025benchmarks

30 **Meta Platforms**, Q1 2026 earnings call transcript, 29 April 2026 — Business AI conversation volume. investor.atmeta.com › META-Q1-2026-Earnings-Call-Transcript.pdf

31 **Kalibri Labs**, *Book Direct Campaigns 2.0* (US data, January 2016 – August 2018; ~19,000 hotels). **Commercial research; dated.** kalibrilabs.com › Book Direct Campaigns 2.0.pdf

32 **Lewis, P. et al.** (2020), "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", *NeurIPS 2020*. arxiv.org/abs/2005.11401

33 **Shuster, K., Poff, S., Chen, M., Kiela, D. & Weston, J.** (2021), "Retrieval Augmentation Reduces Hallucination in Conversation", *Findings of EMNLP 2021*. DOI 10.18653/v1/2021.findings-emnlp.320. aclanthology.org/2021.findings-emnlp.320

Material referenced in this thesis

The deployment described in §7 and §8
vera.support/solutions/hotels/

The published pricing used in §8
vera.support/pricing/

Grounding, escalation and refusal behaviour — §5
vera.support/ai-chatbot/

Full capability scope — §10 positioning
vera.support/features/

Channel surfaces referenced in Figure 5
vera.support/apps/

Adjacent verticals on the same architecture
vera.support/solutions/

Publisher material, listed for reference. Nothing in sections 3 to 12 is cited to any of it, and none of it is evidence for any claim in this document.

REUSE, CITATION AND CORRECTIONS

Quotation, excerpting and reproduction are permitted with attribution. Suggested citation: *Grounded Agents: Where the Defensible Layer Sits in Hospitality AI*, 2026. Vera, September 2026. Corrections are welcome and will be published. The companion operational report for hotel operators, *The Grounded Front Desk*, covers the same evidence base from the buyer's side. Product and platform material is at vera.support, with the hospitality deployment described at vera.support/solutions/hotels/ and the published pricing used in §8 at vera.support/pricing/.