

OutcomeClear Outcome Audit Report

Customer: Northstar Commerce Ltd

AI Support Vendor: HelixAI Support

Audit Type: Independent Outcome Verification — Synthetic Blind Reasoning Test

Evidence Cutoff: 31 August 2026, 23:59 UTC

Outcomes Audited: 50

Claimed Price per Outcome: €2.50

Total Claimed Value Audited: €125.00

1. Executive Summary

OutcomeClear independently reviewed **50 customer-support outcomes claimed as successfully resolved and billable by HelixAI Support**.

Each claimed outcome was tested against the agreed commercial definition of a successful AI-resolved outcome, using the supplied customer conversations, system-of-record events, approved policy information, follow-up activity and evidence-coverage records.

Final Result

Classification	Outcomes	Percentage	Value
VALID	36	72%	€90.00
INVALID	10	20%	€25.00
REVIEW	4	8%	€10.00
Total	50	100%	€125.00

€90.00 of claimed value was independently verified as satisfying the approved OutcomeSpec.

€25.00 of claimed value was classified INVALID under the approved OutcomeSpec because one or more required commercial conditions conclusively failed.

A further **€10.00 remains pending verification** because the available evidence was insufficient to conclusively determine whether the relevant outcomes satisfied all contractual conditions.

REVIEW outcomes have **not** been treated as confirmed invalid charges.

2. Scope

The audit assessed whether each vendor-claimed successful AI support resolution satisfied Northstar Commerce Ltd's agreed commercial definition of a billable outcome.

The assessment covered:

- whether the interaction was an eligible genuine customer-support interaction;
- whether the AI was the primary resolver;
- whether the customer's actual requested outcome was achieved;
- whether required transactional actions were successfully completed;
- whether informational answers were materially correct;
- whether the issue reopened or materially recurred within the agreed finality period;
- whether the customer abandoned the interaction before resolution;
- and whether sufficient authoritative evidence existed to verify the outcome.

The audit covered **50 vendor-claimed outcomes** at **€2.50 per claimed successful outcome**.

Evidence was assessed through the supplied cutoff of **31 August 2026 at 23:59 UTC**.

3. Approved OutcomeSpec

The audit applied the commercially approved OutcomeSpec.

In summary, a claimed outcome was considered billable only where all applicable requirements were satisfied.

Core Commercial Requirements

Eligibility

The interaction must represent a genuine production customer-support interaction. Test, spam and duplicate interactions are excluded.

AI Primary Resolution

The AI must complete the material customer outcome before a human performs a material completion step. Human activity occurring after completion or relating to a separate issue does not automatically disqualify the outcome.

Customer Outcome Achieved

The customer's actual requested outcome must be achieved. Customer intent is determined from the underlying interaction rather than relying solely on the vendor's category label.

Transactional Completion

Refunds, cancellations, subscription pauses, address changes and promotional-code applications require successful completion in the relevant authoritative system of record.

Password / Access Outcomes

Password-reset outcomes require either an authoritative completed reset event or sufficiently explicit customer confirmation that access was restored.

Information Accuracy

Information provided to customers must be materially correct when compared with the applicable system of record or approved policy information.

72-Hour Finality

The same underlying issue must not reopen or materially recur for the same customer within 72 hours of the claimed resolution.

Customer Abandonment

Where additional information is required before the AI can complete the requested outcome, an interaction is not considered successfully resolved if the customer stops responding before completion.

Evidence Sufficiency

Where an authoritative evidence source is known to be unavailable or incomplete, missing evidence alone is not treated as proof of failure. Such cases remain pending verification unless another commercial requirement conclusively failed.

4. Data Sources Reviewed

OutcomeClear assessed the supplied evidence from the following sources:

- vendor outcome claims;
- customer-support conversation messages;
- backend system-of-record events;
- customer follow-up and reopen activity;
- approved customer-support knowledge and policy information;
- source-coverage and outage records;
- commercial agreement and pricing terms;
- and audit metadata.

Vendor-generated labels such as “Resolved” or “Billable” were treated as claims requiring independent verification rather than as evidence of success.

5. Audit Method

Each claimed outcome was independently matched to its underlying evidence and assessed against every applicable OutcomeSpec requirement.

OutcomeClear used the following evidence hierarchy:

1. **Authoritative systems of record** for transactional execution and backend status;
2. **Approved knowledge and policy sources** for informational accuracy;
3. **Customer conversations** for customer intent, AI statements, abandonment and human participation;
4. **Follow-up evidence** for reopen and repeat-issue testing;
5. **Source-coverage records** to determine whether missing evidence could reliably be interpreted.

A conversational statement from the AI that an action had been completed was not treated as proof where the OutcomeSpec required authoritative system evidence.

Where a requirement conclusively failed, the outcome was classified **INVALID**.

Where no requirement conclusively failed but necessary evidence was unavailable or ambiguous, the outcome was classified **REVIEW**.

Only outcomes supported against all applicable requirements were classified **VALID**.

6. Financial Reconciliation

Reconciliation Category	Value
Total claimed value audited	€125.00
Verified value	€90.00
Invalid value	€25.00
Review-pending value	€10.00

Confirmed Invalid Value

€25.00 of claimed value was classified INVALID under the approved OutcomeSpec.

This represents **10 of the 50 claimed successful outcomes**.

Pending Verification

€10.00 across four outcomes remains pending verification.

These cases have not been classified as confirmed invalid because the supplied evidence does not support that conclusion.

7. Primary Exception Categories

The ten INVALID outcomes were distributed across the following primary failure categories:

Exception Category	Outcomes
Required backend action not completed	4
Ineligible / test interaction	1
Requested customer outcome not achieved	1
Issue reopened inside finality window	1
Disqualifying human intervention	1
Customer abandoned before resolution	1
Same underlying issue repeated inside 72 hours	1
Total INVALID	10

Several outcomes also contained secondary commercial failures. These were retained in the underlying exception ledger but were not double-counted financially.

8. Illustrative Exceptions

Transaction Reported as Completed but Backend Action Failed

In one refund case, the AI informed the customer that the refund had been processed.

The authoritative Payments system recorded the refund event as **FAILED**.

Because the commercial definition required successful backend execution, the outcome was classified **INVALID**.

Customer Given Incorrect Policy Information

In one informational interaction, the AI advised that unused products could be returned within **60 days**.

The approved policy established a **30-day** return period.

The response was therefore materially incorrect and the claimed resolution was classified **INVALID**.

Human Completed the Material Customer Outcome

In one subscription-cancellation interaction, the AI stated that it could not complete the cancellation and transferred the interaction.

A human subsequently performed the material cancellation action.

Because the AI had not completed the customer outcome before the human intervention, the claimed AI resolution was classified **INVALID**.

Resolution Initially Completed but Same Problem Recurred

In one password-reset case, an authoritative reset event was successfully completed.

However, the same customer returned approximately **21 hours later** reporting that the same login problem remained.

Because the same underlying issue recurred inside the contractual 72-hour finality period, the original claimed successful outcome was classified **INVALID**.

9. Review-Pending Outcomes

Four outcomes remain pending verification.

Payments Evidence Outage

A refund claim lacked authoritative backend evidence, but the relevant Payments source had a documented outage during the applicable period.

The missing event therefore cannot reliably establish either successful completion or failure.

Status: REVIEW

Finality Window Not Complete at Evidence Cutoff

A subscription cancellation had authoritative completion evidence.

However, its contractual 72-hour finality period had not fully elapsed by the supplied evidence cutoff.

The finality requirement could therefore not yet be verified.

Status: REVIEW

Incomplete Conversation Record

An address-change event was successfully recorded in the backend system.

However, a documented gap existed in the conversation transcript, preventing OutcomeClear from conclusively establishing whether a human performed a material completion step.

Status: REVIEW

Authentication Evidence Outage

A password-reset interaction occurred during a documented Authentication evidence outage.

The customer indicated that the reset may have worked but did not explicitly confirm that access had been restored.

Neither permitted verification path could therefore be conclusively established.

Status: REVIEW

10. Human QA

Before final reconciliation, OutcomeClear's QA process required manual review of:

- **100% of INVALID outcomes;**
- **100% of REVIEW outcomes;**
- and a deterministic **10% sample of VALID outcomes.**

A total of **18 outcomes** were included in the QA review.

No classification overrides were made.

The preliminary audit results therefore became the final audit results without adjustment.

11. Material Limitations

This audit is limited to the commercial agreement, OutcomeSpec and evidence supplied for assessment.

Four outcomes remain unresolved because required evidence was unavailable, incomplete or had not yet matured through the full contractual finality period.

Those outcomes have deliberately remained **REVIEW** rather than being assumed either valid or invalid.

The audit does not treat REVIEW value as confirmed vendor overbilling.

12. Recommended Control Improvements

OutcomeClear recommends the following controls to make future outcome-priced AI billing easier to verify and reconcile.

1. Require Transaction Completion IDs

For transactional outcomes such as refunds, cancellations and account changes, retain the authoritative completion-event ID alongside the vendor's claimed outcome.

This creates a direct link between the billed outcome and the underlying executed action.

2. Define Material Human Intervention More Precisely

The commercial agreement should include practical examples of which human actions disqualify an AI-resolved outcome and which post-resolution or unrelated activities do not.

This reduces ambiguity in future billing reconciliation.

3. Preserve Evidence Through the Full Finality Window

Conversation, follow-up and backend records should remain available until the entire contractual reopen/finality period has elapsed.

This prevents otherwise successful outcomes from remaining pending solely because the audit cutoff occurred too early.

4. Maintain Stable Cross-System Identifiers

A consistent customer, ticket and outcome identifier should be retained across the support platform, AI vendor records and backend systems.

This makes matching, repeat-contact detection and reconciliation materially easier and more reproducible.

13. Final Conclusion

Across the **50 vendor-claimed successful AI support outcomes** tested:

36 outcomes representing €90.00 were independently verified as VALID.

10 outcomes representing €25.00 were classified INVALID under the approved OutcomeSpec.

4 outcomes representing €10.00 remain pending verification because sufficient evidence was unavailable or incomplete.

The results indicate that independent verification can identify meaningful differences between vendor-claimed successful outcomes and outcomes that satisfy the customer's agreed commercial definition of success.

Disclaimer

OutcomeClear's Outcome Audit is an independent operational and commercial verification of supplied outcome evidence against an agreed OutcomeSpec. It is not a statutory financial audit, accounting opinion, legal opinion, or assurance engagement conducted under professional auditing standards.