KILLGATE | EVIDENCE-FIRST PUBLIC RESEARCH

Vehicle Audio Diagnostics

VERDICT More public evidence is needed — Test a critical capability in parallel CONFIDENCE 0.91

The market problem is credible and buyers demonstrably spend money on diagnosis. The idea you tested still fails the research gate because its only material differentiation—identifying the actual failing component from 30 seconds of consumer smartphone audio with 95% accuracy—has not been established at that scope. This is not ≥ KILL; it means the technical premise should be falsified before buyer interviews or product build.

The idea you tested — KG-2026-08-28-VAUDIO-001

Field Locked definition

target customer U.S. owner of an out-of-warranty gasoline passenger vehicle who hears an unfamiliar mechanical noise and is considering a mechanic visit

Problem Owner cannot determine what component is failing before visiting a mechanic

Offer Smartphone app records the vehicle for 30 seconds and identifies the failing mechanical component

Accuracy 95% top-1 component identification accuracy ≥

Price $19.99 per diagnosis

Current alternatives Mechanic diagnosis, OBD scanner/app, asking forums/friends/mobile mechanics, existing AI sound-diagnostic apps

Important constraints Smartphone/audio variability, open-ended fault taxonomy, ground-truth labeling, safety/trust, substantiation of the 95% claim

Change control Lowering accuracy, returning “possible causes,” adding OBD as a required input, changing price, or narrowing faults materially requires a new contract

For this contract, “95% accurate” cannot be rescued by reporting 95% accuracy only on cases where the model feels confident. The in-scope population and abstention rules must be frozen before testing.

What was checked

Gate Result Evidence / threshold

Rules fixed before research PASS KG-2026-08-28-VAUDIO-001

3 DIRECT public sources ≥ PASS 10+ directly relevant sources reviewed

2 independent source groups ≥ PASS Reddit consumers, repair pricing, app stores/vendors, OEM, academic research, FTC

Customer pain PASS Owners seek help for noises, fear breakdowns, and struggle to reproduce intermittent problems for mechanics.

Current workaround / alternatives PASS Mechanics, crowdsourced recordings, mobile mechanics, OBD products, and sound-analysis apps are already used/offered.

Existing spend / budget proxy PASS RepairPal estimates general diagnosis at $122–$179; Pep Boys lists $110. Consumers describe $150–$257+ diagnostic charges.

Active disconfirming search PASS Accuracy limitations, free/cheap substitutes, capture failures, and legal/trust constraints searched.

2 supporting claims ≥ PASS Pain/spend exists; acoustic signal contains useful diagnostic information.

2 constraints surfaced ≥ PASS Technical generalization + low-priced competition; additionally safety/substantiation.

Load-bearing technical feasibility FAIL / Needs a real-world test Public evidence does not establish 95% ≥ exact-component identification over the locked population.

Research gate overall FAIL Test a critical capability in parallel blocks Public evidence supports buyer testing.

Qualified buyers NOT YET TESTED 0 / 8

Strong pain direct buyers NOT YET TESTED Public posts cannot count.

$19.99 price-positive buyers NOT YET TESTED 0 / 4

Verified paid pilots NOT YET TESTED 0 / 2

Source diversity is not being confused with evidence independence here. Competitor/App Store pages establish features and prices; they are not independent proof of customer demand.

Mechanism scoreboard

Mechanism Role Status Assessment

30-second phone recording contains useful mechanical information Required part of the offer Possible, not proven Strong evidence for bounded fault classes

Detect abnormal vs normal operation Central to why someone would pay submechanism Possible, not proven Technically credible

Classify a broad fault family Central to why someone would pay submechanism Possible, not proven Research reaches roughly 90%+ in relevant settings

Identify the exact failing component at 95% across ≥ consumer vehicles Central to why someone would pay Needs a real-world test Unresolved load-bearing claim

Guided recording / vehicle-profile context Helpful extra Already common Existing apps already do this

Consumer pays $19.99 once Central to why someone would pay economic mechanism Needs a real-world test Mechanic spend supports value, but exact digital substitutes are substantially cheaper

Uncertainty / safety / abstention handling A limit on the offer Backed by public evidence Existing products consistently limit diagnostic claims; cannot silently add this to rescue the locked offer

The decomposition matters. A paper showing 95% accuracy distinguishing six fault categories does not prove 95% accuracy identifying the actual failed component.

Strongest evidence FOR There is a genuine economic job. RepairPal’s August 2026 estimate puts general automotive diagnosis at $122– $179, while Pep Boys lists $110 for engine diagnostic service. Public consumer examples include diagnostic charges of roughly $150, $160, $225 and $257. A $19.99 product therefore addresses an existing paid job rather than inventing one.

The consumer behavior already resembles the proposed workflow. Owners upload audio recordings, ask strangers to identify noises, hire mobile/Facebook mechanics, and describe difficulty reproducing intermittent noises once they reach a shop. That is directly relevant observed workaround behavior.

Acoustic diagnosis itself is technically real. Škoda’s Sound Analyser used ordinary smartphones/tablets and reported over 90% accuracy on ten learned patterns, including steering, compressor and DSG-clutch problems. A 2024 study recorded actual Ford/Toyota vehicles for exactly 30 seconds using a cellphone and achieved about 92.2% accuracy across broad fault categories.

Tightly bounded experiments can exceed 95%. Smartphone-based automobile power-seat fault classification reported approximately 98–99.6%. That strengthens the case that phones can capture mechanical signals—but it simultaneously illustrates the scope problem: high accuracy is easiest on a narrow machine/fault taxonomy.

Strongest evidence AGAINST

Closest consumer substitutes are dramatically below $19.99. AutoDiag offers the first diagnosis free and credits from $0.99; EngineScan sells its premium report for $2.99; DriveVerse lists a complete analysis for $4.99. That does not disprove $19.99 willingness-to-pay, but the premium must come from a measurable advantage—most plausibly much higher accuracy/trust—not merely the convenience of recording a noise.

The strongest directly comparable research misses the locked accuracy claim. The 2024 30-second cellphone study achieved about 92.2% while identifying relatively broad categories such as spark-plug, airflow, electrical, engine/turbo and front-end problems—not individual failed parts. A later vehicle-sound framework reported approximately 95% classification performance, again on bounded categories rather than arbitrary component identification.

Existing products themselves avoid the locked semantics. EngineEar reports “possible mechanical causes ranked by likelihood” and explicitly says it is not a substitute for professional diagnosis. EngineScan describes its output as a “first look” and likewise says it does not replace inspection or safety decisions. Operators actually shipping this mechanism generally position it as triage/probabilistic guidance, not “we identified the failing component with 95% accuracy.” ≥

Input reliability is nontrivial. A current DriveVerse reviewer reported completing the workflow only to have the application fail to analyze the audio; a MechanicAdvice user also noted that phone recordings materially changed how the noise sounded. These examples are weak individually, but directionally consistent with domain-shift risk.

The 95% claim carries substantiation risk. A 95% accuracy claim is an objective advertising claim. FTC guidance requires a reasonable evidentiary basis for objective product claims and notes that performance/safety claims can require strong substantiation.

Evidence audit

Category Assessment

Known facts Consumers experience uncertain vehicle noises; professional diagnosis costs materially more than $19.99; consumers already record noises; multiple audio-diagnostic apps exist; smartphone acoustic fault classification works in bounded settings.

Inference A credible 95%-accurate solution could command a premium over $0.99–$4.99 competitors because it substitutes partly for a $110–$179 professional diagnostic.

Major unknown Whether a 30-second consumer recording contains enough information to identify the exact component across a commercially useful range of failures at 95%. ≥

Major unknown Whether consumers will actually pay $19.99 when cheaper/free sound-analysis alternatives exist.

Major unknown What percentage of real mechanic-bound complaints are acoustically diagnosable at all.

Major unknown Data acquisition cost: diverse phones × vehicles × operating conditions × confirmed component-level labels.

Contradiction Market/OEM implementations generally provide likely causes or limited-pattern diagnosis, while the contract promises definitive component identification.

Contradiction Professional diagnosis establishes substantial spend, but digital sound-diagnostic pricing clusters far below $19.99.

Excluded evidence Competitor claims such as “98% accuracy” were treated as seller assertions, not proof of technical performance.

Excluded evidence App ratings with only a handful of reviews were not treated as demand validation.

Excluded evidence No public comment, download, signup, or “sounds useful” reaction was promoted into price-positive or payment evidence.

Why the verdict follows

This is not a pain-market failure. Pain, current behavior, alternatives and spend all clear the public research minimums.

The blocker is much narrower and more serious: the contract's economic differentiation depends heavily on 95% exact-component accuracy. Public evidence shows useful acoustic information and encouraging bounded classifiers, but it does not establish the locked mechanism.

Because cheap competitors already provide probabilistic sound diagnosis, dropping from "95% exact component" to "likely causes" would materially alter why someone pays $19.99. Killgate therefore does not permit quietly weakening that mechanism and declaring success.

Result: More public evidence is needed, Test a critical capability in parallel.

Build-risk memo

The expensive part is unlikely to be the mobile UI or audio model API. It is the ground-truth dataset.

A credible system needs repair-confirmed component labels across makes, engines, mileage, microphones, phone positions, background noise, temperatures and operating conditions. Worse, the problem is open-world: two acoustically similar noises can originate from different components, while many failed components generate no distinctive audible signature.

The second commercial risk is acquisition economics. This is potentially a low-frequency, one-shot $19.99 purchase competing against free search, Reddit, $0.99–$4.99 AI products and reusable OBD hardware. Even if gross margin is excellent, CAC could kill the DTC model.

Neither risk changes today's verdict; they become relevant only if the acoustic mechanism survives.

Minimum next action

Do not build the app yet. Do not start broad buyer interviews yet.

Run the cheapest serious component-level acoustic falsification test.

Use real vehicles with known repair outcomes. Before the mechanic diagnoses each vehicle, capture exactly the input the proposed app would receive: one 30-second owner-smartphone recording plus whatever vehicle metadata was already included in the frozen test protocol. Ground truth should be the confirmed failed/repaired component, not the mechanic's initial guess.

Freeze the component taxonomy and test set beforehand. Score top-1 exact-component match; "one of these three things," fault-family-only answers, and abstentions cannot count as correct for this contract. Separate vehicles/shops used to develop the model from the final holdout, and report component-level sensitivity as well as aggregate accuracy.

For a cheap first falsification, a roughly 100-case blind holdout is enough to discover whether the system is nowhere near the claimed regime. Prelock 95% top-1 accuracy as the contract threshold; anything materially ≥ below it means this contract cannot proceed. Even hitting 95% on 100 cases would only justify a larger external ≥ substantiation study—it would not yet justify advertising "95% accurate."

Compare the model against at least one competent mechanic given the same limited audio/input, so you know whether the algorithm is extracting signal rather than merely rediscovering obvious noises.

What would change the verdict

A blind external test demonstrating the locked 95% exact-component result across a frozen, commercially useful ≥ fault population would remove the main research blocker. At that point the hypothesis could move to Public evidence supports buyer testing and direct buyer validation should test the exact $19.99 offer.

If the system instead performs well only at "likely fault family / possible causes / urgency," that is evidence for a potentially useful product—but it is a PIVOT, requiring a new contract rather than retroactively changing this one.

Even after technical success, final GO would still require the locked human gates: at least 8 qualified buyers, 30% ≥ strong pain from recent real examples, 4 price-positive buyers at exactly $19.99 unless payments supersede that ≥ gate, and 2 verifiable paid pilots/deposits. ≥

Source references used in the research run

■ RepairPal — General Diagnosis Cost Estimate: https://repairpal.com/estimator/general-diagnosis-cost ■ Pep Boys — Engine Diagnostic Service: https://www.pepboys.com/auto-service-repair/engine-diagnostics

■ Škoda Storyboard — AI Sound Analyser: https://www.skoda-storyboard.com/en/press-releases/skoda-auto-uses-artificial-intelligence-for-even-more-accurate-car-diagnostics-2/ ■ MDPI Applied Sciences 2024 — vehicle fault diagnosis with cellphone recordings: https://www.mdpi.com/2076-3417/14/15/6532 ■ FTC — Policy Statement Regarding Advertising Substantiation: https://www.ftc.gov/legal-library/browse/ftc-policy-statement-regarding-advertising-substantiation ■ BlueDriver — diagnostic scan tool: https://us.bluedriver.com/products/bluedriver-scan-tool ■ EngineEar — Google Play listing: https://play.google.com/store/apps/details?id=com.engineear.ai ■ DriveVerse — App Store reviews/listing: https://apps.apple.com/us/app/driveverse-ai-car-diagnostics/id6744003712

Decision semantics: public research can produce Public evidence supports buyer testing, More public evidence is needed, or Change the idea before more testing only. A final GO requires direct buyer/economic gates and verified paid evidence.