Arviser

EXECUTIVE PACKAGE SHEET · 2026

Independent LLM Quality Audits & Fixed-Scope AI Engineering Sprints

Measured software quality, repeatable CI evaluation gates, and production-grade AI vertical slices. Delivered hands-on by senior engineers (10+ years in production) under direct founder oversight (30+ years in IT).

Arviser Sp. z o.o.
Warsaw, Poland (EU)
KRS 0001068182
NIP PL5252980041
arviser.com · contact@arviser.com

EXECUTION STANDARD	FOUNDER OVERSIGHT	SCOPE GUARANTEE	RISK PROTECTION
10+ Yrs Senior Engineers + AI	Vadim Nareyko · 30+ Yrs IT	Written Scope · Overrun on Us	50% Kill Waiver · 100% Credit

1. Complete Fixed-Scope Package Lineup (EUR, excl. VAT)

50% on signature · 50% on delivery · Free 48h written scope via arviser.com/#scope

1. 24-Hour Technical PoC Research Spike — from €2,400

3–5 business days · 50% Risk-Shared Option (€1,200 total on Kill) · ~10,300 PLN

Verify in principle whether a working pipeline solves the narrow technical bottleneck across 2 to 3 models and 20 to 50 cases before full product spend.

Deliverables: Written pass marks, reproducible CLI or notebook benchmark in Promptfoo or DeepEval, p95 latency and unit-cost math, Go, Pivot, or Kill memo, repo handoff, 100% fee credit in 60 days on Go.

2. AI Readiness Check — from €2,900

1 week fixed scope · 100% credited to next engagement within 90 days

For stalled pilots needing production gates, risk ownership, and a clear kill, continue, or audit decision before committing another quarter of budget.

Deliverables: System and use-case map, production-gate scorecard against our public checklist, severity-ranked risk register, acceptance thresholds draft, 90-day engineering roadmap.

3. 32-Hour Eval & Guardrail CI Harness — from €3,100

1 week (32 senior hours) · Automated regression and refusal gates in CI/CD

For teams with an existing prototype or live endpoint that currently relies on manual vibe-checking instead of automated pass/fail release gates.

Deliverables: Promptfoo or DeepEval wired into GitHub Actions or GitLab CI, 50-case baseline golden set, refusal-both-ways checks, PII and schema assertions, latency and token-budget thresholds.

4. 40-Hour Working AI Prototype Sprint — from €3,900

1–2 weeks (40 hours) · €3,300 on Startup Co-Branded Track · ~16,800 PLN

When feasibility is clear and no separate PoC is needed: ship a clean, deployed end-to-end vertical slice without fragile junior boilerplate.

Deliverables: Typed API contracts, RAG or tool-calling backend with fallbacks and cost logs, web UI or Slack/CRM integration, input guardrails, 15–25 smoke evals, Docker deploy.

5. 60-Hour Prototype-to-Production Hardening — from €5,600

2 weeks (60 senior hours) · Reliability, cost controls, and runtime protection

Turn an internal demo or 40-hour prototype into a resilient service engineered for real customer traffic, provider outages, and adversarial inputs.

Deliverables: Multi-provider routing and timeouts, semantic or deterministic caching, per-tenant budget caps, runtime kill-switch, redacted structured logs, async production trace sampling.

6. Two-Week Independent LLM Quality Audit — from €8,000

2 weeks fixed scope · €6,800 Startup Co-Branded · 100% credited in 90 days

Independent third-party measurement of a live LLM or RAG system across six scored quality areas so leadership gets an objective Go, Hold, or No-Go.

Deliverables: 6-area scorecard (1–5 rubric), ranked failure-mode register, repeatable evaluation suite in your repo, 1-page executive summary, deep engineering report, 60-min readout.

7. 80-Hour Production AI MVP Sprint — from €8,500

3–4 weeks (80–120 hours) · Full production release with CI gates

Customer-facing AI service integrated with your authentication, RBAC, data pipelines, monitoring, and EU AI Act Art. 50 technical documentation inputs.

Deliverables: Production backend and UI slice, RBAC and tenant isolation, full CI evaluation harness, observability dashboards, runbooks, Art. 50 transparency evidence inputs.

8. Founder Advisory, Retainers & Dedicated Engineers — €250/h · €2,400/mo

100% consultation credit in 30 days · Specialists from €6,500/mo

Direct 1-on-1 Founder Consultation (€250/h), fractional Senior AI Architect Retainer (20h/mo €2,400 · 40h/mo €4,500), or full-time specialists.

Dedicated monthly bands (min 3 mos): Mid AI-QA from €6,500/mo · Mid AI/ML from €7,500/mo · Senior AI-QA or MLOps from €9,000/mo · Senior AI/ML (10+ yrs) from €10,000/mo.

2. Decision Matrix: Which Entry Point Fits Your Current Stage?

Qualify by technical bottleneck and production risk, not headcount

Your Current Situation	Recommended Package	Timeline	From Price	Fee Credit Path
Hard technical unknown (accuracy, OCR/PDF parsing, p95 latency, unit cost, local PII)	**24h Technical PoC Spike**	24h (3–5 days)	**€2,400 (€1,200 Kill)**	100% to 80h MVP or Audit in 60d (50% to 40h Prototype)
Feasibility is clear; need a clean, deployed vertical slice without PoC overhead	**40h Working AI Prototype**	40h (1–2 weeks)	**€3,900 (€3,300*)**	Builds directly into 60h Hardening or 80h MVP
Working demo exists, but zero automated CI evals, refusal checks, or regression gates	**32h Eval & Guardrail CI**	32h (1 week)	**€3,100**	Standalone or paired with Hardening or Audit
Stalled pilot; leadership needs production gates, risk register, and 90-day roadmap	**AI Readiness Check**	1 week	**€2,900**	100% credited to Audit or build within 90 days
Live LLM or RAG system needs independent quality scores before scaling traffic	**LLM Quality Audit**	2 weeks	**€8,000 (€6,800*)**	100% credited to next engagement within 90 days

Arviser

METHODOLOGY, BENCHMARKS & GOVERNANCE

How Arviser Compares to Big4 Advisories, Certifiers & Dev Agencies

Written-First Intake
Free 48h scope reply
€250/h Founder Call
arviser.com/#scope

3. Market Positioning: Technical Measurement & Senior Sprints vs Alternatives

Transparent fixed-scope pricing vs European market benchmarks

Audit Dimension	Arviser Audit	Big4 Governance	Certifiers (SGS/BSI/TÜV)
Primary Focus	**Live LLM/RAG scores, failure modes, CI suite**	Policy frameworks, controls, board decks	Formal standard conformity check
Who Executes	**10+ yr senior engineers + founder**	Advisory and GRC consultants	Accredited scheme auditors
Deliverable	**Runnable Promptfoo/DeepEval suite + report**	Governance matrices and slide decks	Conformity statement or certificate
Timeline & Fee	**2 weeks · from €8,000 (€6,800*)**	Multi-month · €40,000–€150,000+	Scheme cycle · custom quote

Build Dimension	24h PoC & 40h Prototype	EU Agency Discovery	Vibe-Coded Freelance
Job to be Done	**Prove bottleneck (24h) or ship slice (40h)**	Workshops, wireframes, backlog	Unverified demo wrapper
Quality Gates	**20–50 PoC cases or 15–25 smoke evals**	Rarely includes automated evals	Zero tests, hidden security gaps
Risk Protection	**50% Kill waiver on 24h PoC; fixed cap**	Paid T&M regardless of viability	Rewrite needed before prod
Published Fee	**€2,400 (PoC), €3,900 (Prototype)**	€10,000–€25,000 before code	Low day-rate, high rewrite cost

4. Six Scored Audit Areas, 100-Case RAG Golden Set & 50% Risk-Shared Protocol

Open-source evaluation tooling (Promptfoo, DeepEval, Ragas) — zero vendor lock-in

Six Scored Quality Areas (1–5 Rubric) & 100-Case RAG Golden Set

- **1. Golden Set Coverage:** 100-case RAG evaluation baseline structured as **30 normal queries**, **20 multi-hop reasoning**, **20 hard retrieval edge cases**, **15 unanswerable refusal cases**, and **15 conflicting-source cases** with passage-level citations.
- **2. Groundedness & Faithfulness:** Answers strictly bound to retrieved evidence; zero fabricated numbers, fees, or policy clauses.
- **3. Refusal Both Ways & Safety:** Balanced testing against both unsafe compliance (jailbreaks, tool misuse, PII leak) and false over-refusal on benign prompts.
- **4. CI & Regression Gates:** Automated pass/fail checks on every prompt, model, or chunking PR in GitHub Actions or GitLab CI.
- **5. Production Sampling:** Stratified async trace sampling to detect post-release drift.
- **6. Cost & Latency Budgets:** Token unit cost per 1,000 runs and p95 latency under realistic load.

Startup Program (/startups) & Risk-Shared PoC

- **50% Risk-Shared 24h PoC (€1,200 on Kill):** In Hours 1–4 we lock numeric pass marks on 20–50 cases. If none of the 2–3 tested models clear the bar, the verdict is **Kill** and the second €1,200 is waived. You keep a non-exclusive internal license; Arviser may publish a sanitized synthetic-data note.
- **Go Outcome (100% Exclusive IP + Credit):** When pass marks are met, 100% exclusive IP transfers under NDA and 100% of €2,400 credits toward an 80h MVP or Audit within 60 days.
- **15% Co-Branded Open Proof Discount:** €3,300 (40h Prototype) or €6,800 (Audit) with joint case study & CTO quote.
- **Free 2-Page VC Data Room Memo:** Signed architecture, latency, unit-cost, and eval baseline memo for investor due diligence.
- **Cloud Credit Builds & 60-Day Split:** Deploy on AWS Activate, Google, Azure, or Cloudflare credits; 50/50 split on €9,500 bundles.

5. Open Proof & Public Engineering Artifacts (Not Client Case Studies)

Public benchmarks, datasets, and tools built by Founder Vadim Nareyko

WardBench

SYNTHETIC FHIR BENCHMARK

Measures clinical agent utility vs policy violation rate under role, purpose, and dynamic patient Consent across 11 models.
DOI: 10.5281/zenodo.23664058
GitHub & PyPI: wardbench

Genivara-PII (R&D)

ON-DEVICE WASM PRIVACY

8.2 MB INT8 model running in browser WASM (~15 ms median latency) to strip PII before data reaches external LLMs. R&D track, not an Arviser SKU.
DOI: 10.5281/zenodo.19118893

lexaudit & Eval Repos

OPEN MIT TOOLING

lexaudit: deterministic CLI for calques, false friends, and invented tokens in LLM text (PyPI). Plus **golden-set-schema** and **go-llm-gateway-audit** reference repos.

SeaLights Experience

QUALITY INTELLIGENCE

Founder background in enterprise quality intelligence and test-impact analytics: shipping software under measured CI risk gates. Employer domain framing only; zero NDA data.

6. Technology Stack Coverage, Industries & Regulatory Guardrails

62 specialized engineering, AI method, service, and industry routes on arviser.com

Supported Stacks & First-Wave Industries

- **Backend & Gateways (/engineering):** Python (FastAPI, LangGraph, Pydantic), TypeScript/Node (Vercel AI SDK, Zod), Go (high-throughput LLM gateways, audit logs), Rust (Wasm edge guards), Java/Kotlin (Spring AI), C# (.NET Semantic Kernel), PHP, Rails, Elixir, Scala, C++.
- **Frontend & Mobile:** React, Next.js RSC, Vue/Nuxt, Angular, SvelteKit, React Native, Flutter, iOS Swift, Android Kotlin, Electron, Tauri.
- **Core Industries (/industries):** Fintech (KYC, policy & fee grounding), E-commerce (catalog accuracy & abuse refusal), Non-clinical Health-tech (synthetic/DPA software testing).

Engagement Rules & Strict Legal Boundaries

- **Written-First Triage:** Free project scoping and technical Q&A via arviser.com/#scope (reply within 48h on Warsaw business days). No unvetted free discovery calls.
- **Direct 1-on-1 Founder Consultation:** €250/hour excl. VAT. 100% (€250) credited toward any signed package from €2,400 within 30 days.
- **What We Do Not Do:** Not a Notified Body; no EU AI Act conformity certification; no clinical care, clinical decision support, or MDR/IVDR clinical evidence; never send patient or Art. 9 data via web forms (synthetic data or client VPC under DPA only).