FREE GUIDE · HUSNAINBUKHARI.COM

# Pakistan AI Partner Selection Guide

A buyer's guide to hiring an AI expert or automation agency in Pakistan — evaluation criteria, a vetting scorecard, red flags, contracting, and a timezone/compliance checklist for US/UK/UAE teams.

By Syed Husnain Haider Bukhari — AI engineer & founder, Revolutionary Technologies · www.husnainbukhari.com

## Pakistan AI Partner Selection Guide

This guide is the working checklist I use — and hand to clients — when evaluating an AI expert or an AI automation agency in Pakistan. It is written for buyers on US, UK, and UAE teams who want the cost and talent advantage of a Pakistan-based partner without inheriting the risk that comes with an unstructured hire. Everything here expands on the ideas in the companion article; use it as the operational layer that turns a shortlist into a signed, de-risked engagement.

## 1. The Selection Framework at a Glance

Buying applied AI is buying judgment under production constraints. The framework below moves you through four stages in order. Do not skip ahead — most bad engagements come from jumping to price before verifying the fundamentals.

1. **Screen** — filter the market down to a shortlist of three to five partners using coarse signals (production track record, references, university and team pedigree).
2. **Score** — run each shortlisted partner through the weighted scorecard in section 3.
3. **Gate** — apply the three hard gates. Any failure ends the conversation regardless of score.
4. **Scope** — commission a small paid scoping sprint before any large fixed-bid contract.

## 2. The Seven Evaluation Criteria

Weight the first three most heavily — they carry the majority of delivery risk. The last four are important but negotiable.

| Criterion | What good looks like | Weight |
| --- | --- | --- |
| Production track record | Live, referenceable systems in your industry or an adjacent one | High (3x) |
| Evaluation rigor | A repeatable eval harness with accuracy, hallucination, latency, and cost metrics | High (3x) |
| Compliance posture | Clear data residency, PII handling, GDPR/HIPAA readiness, NDA and IP terms | High (3x) |
| Communication cadence | Written weekly updates, shared backlog, defined escalation path | Medium (2x) |
| Timezone overlap | Guaranteed real-time hours with your working day | Low (1x) |
| Delivery transparency | The sales-call senior is the delivery senior; known team seniority | Medium (2x) |
| Commercial clarity | Scope, model, and pricing map to measurable outcomes | Low (1x) |

## 3. The Weighted Scorecard

Score each criterion from 1 to 5, multiply by its weight, and divide by total weight to get a normalized score out of 5. Reject anything below 3.5 or any partner that fails a hard gate. Drop this structure into a spreadsheet or use the JSON version below.

```
{
  "partner": "Example AI Agency",
  "weights": {
    "production": 3, "evals": 3, "compliance": 3,
    "communication": 2, "delivery_transparency": 2,
    "timezone": 1, "commercial": 1
  },
  "scores": {
    "production": 4, "evals": 3, "compliance": 5,
    "communication": 4, "delivery_transparency": 3,
    "timezone": 5, "commercial": 3
  },
  "hard_gates": {
    "signed_nda": true,
    "ip_assignment": true,
    "references_callable": true
  },
  "decision_rule": "reject if any hard_gate is false OR weighted_score < 3.5"
}
```

## 4. The Three Hard Gates

These are non-negotiable. A failure on any one ends the evaluation immediately, no matter how strong the rest of the scorecard looks.

- **Signed NDA before any confidential detail is shared.** A partner who resists this at the earliest stage will resist accountability later.
- **Explicit IP assignment.** The work product — code, prompts, fine-tunes, and data pipelines — must be unambiguously yours in writing.
- **At least one callable reference.** Not a testimonial quote. A real customer you can speak to about how the partner handled a problem under pressure.

## 5. Interview Questions That Separate Shippers From Demo-Builders

Ask these in a live call and listen for specificity. Vague, buzzword-heavy answers are the signal.

1. Walk me through a production AI system you shipped. What broke in the first month, and how did you fix it?
2. How do you measure whether an LLM feature is getting better or worse over time?
3. Describe a time you told a client not to build something they asked for. What did you recommend instead?
4. What is your retrieval strategy for keeping answers grounded, in plain language?
5. How do you keep PII out of third-party model APIs?
6. Who exactly will write the code, and what is their seniority?

> The single best predictor of a successful engagement is whether the partner pushes back thoughtfully during scoping. Someone who agrees to everything is either not listening or planning to renegotiate later.

## 6. Compliance and Contracting Checklist

Settle every one of these before code is written. For regulated industries — healthcare, finance — none are optional.

- Signed mutual NDA in place.
- IP assignment clause covering code, prompts, models, and data artifacts.
- Data-handling agreement: where data lives, who accesses it, retention, and deletion.
- PII redaction or tokenization before data reaches any external model API.

- GDPR / HIPAA applicability confirmed and addressed where relevant.
- Sub-processor list disclosed (which model providers and infra vendors are used).
- Clear payment terms tied to milestones, not calendar dates alone.

## 7. Timezone Overlap Map

Pakistan Standard Time is UTC+5. This gives a workable real-time window with all three target markets.

| Your market | Overlap with PST (UTC+5) | Best meeting window |
|---|---|---|
| United Kingdom | ~4-5 hours | UK morning / Pakistan afternoon |
| United Arab Emirates | Near full working day (1 hour offset) | Any time |
| US East Coast | ~2-3 hours | US morning / Pakistan evening |

## 8. The Paid Scoping Sprint Template

Never open with a large fixed-bid contract. Commission a two-to-three week paid scoping sprint first. It tells you more than any sales process, and it de-risks the big commitment for both sides.

1. **Week 1 — Discovery.** Map the target workflow, define success in numbers, and design the eval criteria.
2. **Week 2 — Thin vertical slice.** Ship one narrow end-to-end path through the real system, instrumented with the evals.
3. **Week 3 — Roadmap and decision.** Deliver a prioritized build plan sequenced by ROI and risk, plus a go/no-go recommendation.

Define success concretely before the sprint starts — for example, "deflect 40% of tier-one support tickets at above 90% resolution accuracy, under a target cost per conversation." That specificity forces the eval design up front and gives you an objective way to judge delivery.

## 9. University Signal Cheat Sheet

Treat the university as one weak signal among many, never the deciding factor. It narrows the funnel; it does not pick the hire.

- **FAST-NUCES** — strongest pure CS and competitive-programming culture; clean algorithmic thinking.
- **LUMS** — strong research and data science; graduates who communicate well with Western stakeholders.

- **PIEAS** — deep math and ML fundamentals; engineers who understand models rather than treating them as black boxes.
- **NUST** — broad and large; a big share of the market's senior software and applied-research engineers.

## 10. Pricing Models and What Each One Buys You

Once a partner clears the gates and the scoping sprint, you will land on one of three commercial structures. Each has a place; the trick is matching the structure to how well-defined the work is.

| Model | Best when | Watch out for |
|---|---|---|
| Fixed-scope project | The work is well-defined and the evals are agreed | Scope creep disputes; padding to cover uncertainty |
| Monthly retainer | Ongoing iteration, maintenance, and a live roadmap | Coasting; insist on a visible backlog and outcomes |
| Time and materials | Genuine R&D where the path is unknown up front | Open-ended billing; cap it and review weekly |

My general advice: use fixed-scope for the first real build after a successful scoping sprint, then move to a retainer once trust and a roadmap exist. Reserve time-and-materials for genuinely exploratory work, and always attach a spending cap and a weekly review so it cannot drift.

## 11. A 30-Day Onboarding Plan

Choosing well is only half the job; the first month sets the tone for the whole engagement. Use this sequence to get to real, shipped value fast.

1. **Days 1-3:** Sign the NDA and IP terms, grant scoped access, and confirm the data-handling agreement is live.
2. **Days 4-7:** Agree the single success metric and the eval harness that measures it.
3. **Days 8-14:** Ship the thin vertical slice end-to-end, instrumented with evals from day one.
4. **Days 15-21:** Review real numbers against the target; adjust architecture before scaling scope.
5. **Days 22-30:** Lock the prioritized roadmap and decide on the commercial structure for the next phase.

If a partner cannot get you to an instrumented, working slice inside the first three weeks, that is data. It usually means either the problem is less defined than you thought, or the partner is slower than they appeared in sales.

## 12. Red Flags Catalog

- Portfolio full of demos and screenshots, no live referenceable systems.
- No answer to how quality is measured — no evals, just confidence.
- Reluctance to sign an NDA or assign IP.
- The senior from the sales call disappears once delivery starts.
- Suspiciously cheap pricing with no scope discipline (juniors learning on your budget).
- High buzzword density, low architectural detail.
- No pushback during scoping.

## Work With Us

If you want a straight, no-hype assessment of whether an AI automation project is worth building — and whether a Pakistan-based partner is the right way to build it — work with Syed Husnain Haider Bukhari. Bring the actual problem, not a vague AI ambition, and you will get an honest view of what to build, how, and what it should cost. If you want to see the kind of production software behind the advice, take a look at ProLeads, the lead-generation platform I build and run.

---