PCT INTELLIGENCE REPORT

Tencent AI Agent Harness

Application KYC & Opportunity Map

Build and Evaluate AI-Native Agent Systems End-to-End.

9	1	6	2
PRIORITY ROLES	P0 APPLY	P1 ADJACENT	P2 STRETCH

Report ID: PCT-TENCENT-AI-20260901-001
Snapshot: 2026-09-01
Primary target: Hunyuan AI Agent Harness Engineer
Evidence: Official public sources + labelled analyst interpretation
Decision: APPLY NOW after Human Gate

Public hiring surface, not an internal organization chart. No headcount, urgency, compensation, reporting-line, or hiring-probability claim.

Contents

Reading guide

Sections 1-10 move from public evidence to role decomposition, candidate fit, system design, and a receipt-based application gate.

Report ID: PCT-TENCENT-AI-20260901-001

Snapshot: 2026-09-01, Asia/Shanghai

Primary role: 混元 AI Agent Harness Engineer（北京/深圳）)

Primary organization surface: TEG / 混元-工程平台

Campaign North Star: Build and Evaluate AI-Native Agent Systems End-to-End.

---

1. Executive decision

Decision: APPLY NOW after final Human Gate.

The Harness role is the strongest current Tencent fit because its public JD asks for exactly the chain Pengyi has been building toward:

```
Agent execution
-> trace and observability
-> evaluation pipeline
-> A/B testing and regression detection
-> debugging
-> labeling / SBS / data platform
-> research-to-engineering transfer
-> AI-native daily workflow
```

The match is not perfect. Large-scale distributed observability, production SBS platform ownership, and a strict two-year relevant-work screen are material risks. The correct strategy is therefore not to inflate experience; it is to lead with unusually direct project and open-source evidence, a bounded system design, and explicit gap awareness.

The secondary opportunity map contains six P1 roles and two P2 roles. The cross-border FX quant role is strategically interesting because it combines banking, quant analysis, data-system ownership, and AI-assisted research, but its three-year gate and live FX evidence gap make it a conversation-first target.

---

2. Method and evidence discipline

2.1 Frozen public surface

The collector queried ten official Tencent Careers keyword surfaces. Counts are overlapping search results, not additive hiring totals:

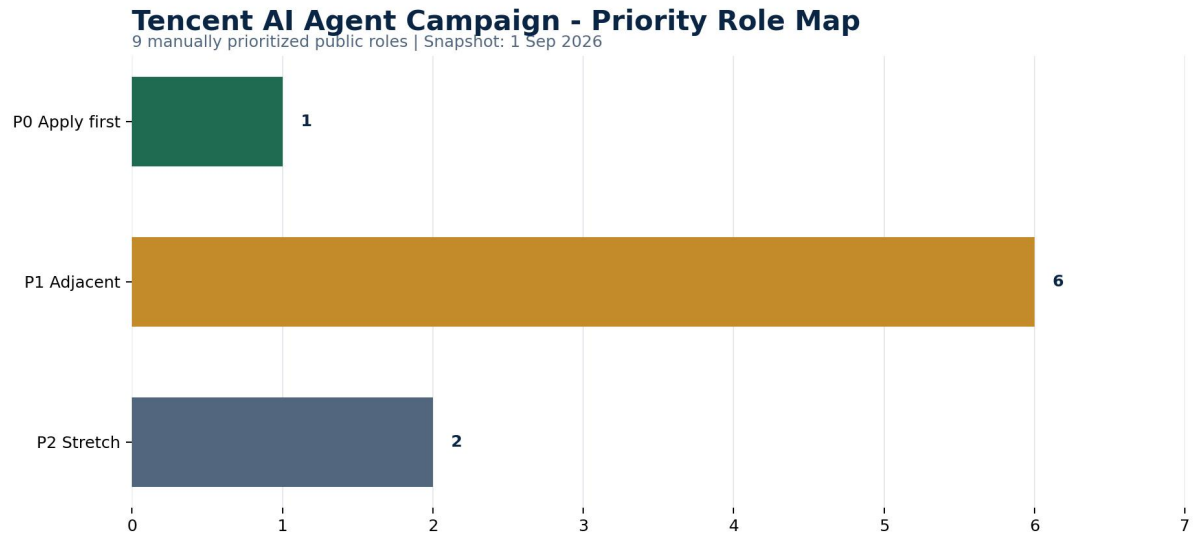

Figure 1. Manually prioritized role surface. Counts are not headcount or hiring probability.

Keyword	Official result count at capture
Agent	199
混元	78
大模型	364
评测	70
推理	117
AI Infra	666
研发效能	8
量化	38
外汇	3
金融科技	23

The union contained 862 unique public records. A deliberately broad signal classifier marked 575 China-location records for review; **575 is a discovery universe, not a candidate-fit count**. Manual prioritization reduced this to nine evidence-relevant roles.

2.2 Three evidence classes

- **FACT:** official role metadata, official Tencent publications, public code, public merged PRs, and repository test receipts.

- **INTERPRETATION:** what the role cluster suggests about technical demand or how candidate evidence maps to a JD.
- **HYPOTHESIS:** internal topology, hiring urgency, reporting lines, and team strategy. Hypotheses are not presented as facts.

Full JDs are not republished in the public dataset. Each role receipt keeps the official URL, metadata, item counts, capability signals, and SHA-256 hashes.

---

3. Tencent AI operating context

3.1 Public company signal

Tencent's July 2026 Hy3 release says the company rebuilt Hunyuan infrastructure from late January 2026 and completed a loop from foundation reconstruction to product feedback and model improvement in under six months. It also names integration into products such as WorkBuddy/CodeBuddy, Yuanbao, Marvis, and ima. That is consistent with a role surface emphasizing execution reliability, evaluation, and product feedback rather than model demos alone.

This does **not** prove the Harness role owns those products. It establishes only the wider public context: Agent capability is being pushed through models, runtime behavior, products, and developer surfaces.

3.2 Technical architecture inferred from public roles

```
Models / post-training
 |
Agent RL framework ---- data and environment platform
 | |
 +------ secure sandbox ---+
 |
Agent execution harness / runtime
 |
trace -> observability -> debugging -> replay
 |
eval pipeline -> SBS -> A/B -> regression gates
 |
research feedback <-> product and developer workflows
```

Inference: The cluster resembles a complete Agent engineering platform, not a single orchestration library. Confidence is medium because the components are visible in public JDs, but internal ownership and integration are unknown.

3.3 Open-source and cloud-adjacent reference points

- Tencent Cloud Agent Runtime publicly describes runtime, secure sandbox, tool infrastructure, gateway, memory/session, observability, governance, and security.

- Tencent-Hunyuan C3-Benchmark evaluates tool dependencies, hidden information, dynamic decision paths, and multi-turn Agent robustness with reproducible artifacts.
- Tencent projects such as CUARewardBench and AI-Infra-Guard show adjacent public interest in Agent reward evaluation and security testing.

These repositories are useful technical references. They are not proof that the P0 TEG team owns or directly uses them.

4. Priority opportunity map

Priority	Role	Location / BG	Gate	Decision
P0	混元 AI Agent Harness Engineer	Beijing/Shenzhen title; TEG	2+ years	Apply first
P1	混元 Agent 评测 Infra 工程师	Beijing/Shanghai/Shenzhen; TEG	3+ years	Adjacent application after P0 review
P1	混元评测 Infra 可观测性工程师	Beijing/Shenzhen; TEG	3+ years	Stretch; prepare incident/trace defense
P1	Agent Infra 高级研发工程师	Shenzhen; TEG	2+ years	Strong Shenzhen adjacency
P1	混元 Agent 数据与环境平台工程师	Shenzhen/Beijing/Shanghai; TEG	1+ years	Strong data-contract adjacency
P1	混元 Agent 强化学习框架工程师	Shenzhen/Beijing/Shanghai; TEG	1+ years	Framework fit, RL ownership gap
P2	混元沙盒平台工程师	Beijing/Shenzhen/Shanghai; TEG	1+ years	Deep virtualization gap
P2	金融科	Shenzhen;	5+ years	Strategic stretch

Priority	Role	Location / BG	Gate	Decision
	技-Agent 架构设计与推理性能优化工程师	CDG		
P1	跨境支付-外汇量化分析产品经理	Shenzhen; CDG	3+ years	Conversation-first quant path

The role title/location mismatch on the P0 page is operationally important: the title lists Beijing/Shenzhen while the detail API returns Beijing. Confirm the actual Shenzhen option before submission or in the first recruiter exchange.

5. P0 JD decomposition

5.1 Output contract

The role is a **Harness / Evaluation / Developer Infrastructure engineer**, not primarily a prompt engineer. Its expected outputs are likely:

- a stable event and trace schema for Agent runs;
- searchable run-level observability and failure triage;
- automatic evaluation and comparison pipelines;
- replay, regression detection, and release gates;
- debugging tools that connect symptoms to task/step/tool/model state;
- data and labeling infrastructure that feeds research and product iteration;
- code that future humans and code agents can both maintain.

5.2 Capability-to-evidence map

Capability	Strongest evidence	Confidence	Interview defense
Typed execution contracts	PDSH task contracts and fail-closed policies	High internally	Explain state, invariant, transition, failure, receipt
Evaluation discipline	Fixed sets, benchmark	Medium-high	Define baseline, metrics, threshold, confidence, rollback

Capability	Strongest evidence	Confidence	Interview defense
	cards, regression gates		
Debugging	Merged boundary fixes in major OSS repositories	High publicly	Walk one failure from reproduction to regression test
Full-stack delivery	PWEB and Cloudflare-deployed workflow systems	High	Explain contract from UI to API to storage/deploy
Agentic coding judgment	Daily multi-model coding plus review gates	Medium-high	Name failure modes; show when to reject model output
Observability	Run receipts and state logs	Medium	Distinguish local auditability from distributed telemetry
Production scale	No verified large-scale receipt	Low	Present migration design; do not claim ownership

5.3 The strongest public evidence story

The open-source story is not “many PRs.” It is a repeated engineering pattern:

1. locate a semantic boundary in an unfamiliar codebase;
2. reproduce the failure with minimal scope;
3. preserve existing abstractions;
4. add a regression test;
5. incorporate maintainer review;
6. stop when the boundary is fixed.

Examples include request-local API keys in LightRAG, bounded retry backoff in Vibe-Trading, pagination validation in Tencent WeKnora, and malformed HTTP/HTML boundary detection in OpenClaw. This pattern directly supports the JD's demand for code quality, debugging, failure-mode awareness, and high-density iteration.

5.4 Eligibility and honesty gate

The role states two or more years of work experience. Pengyi's formal post-graduation work history may not independently clear a strict tenure screen.

The application should still proceed because the plus factors explicitly value open source, personal projects, 0-to-1 Agent products, and intensive agentic coding. The CV must keep dates exact and let the reviewer decide equivalence.

6. Candidate positioning

6.1 One-sentence positioning

> *AI-native research and systems engineer who builds governed Agent execution, evaluation, and debugging workflows, then validates them through tests, receipts, full-stack delivery, and maintainer-reviewed open-source fixes.*

6.2 Evidence chain

```
Complex-systems training
-> quantitative experiment discipline
-> banking data/risk/operations context
-> Agent control plane and evaluation projects
-> typed full-stack deployments
-> maintainer-reviewed OSS debugging
-> Tencent Harness role
```

6.3 Quantified evidence allowed in the CV

- dsh-quant public source: 59 typed tools, six domains, 215 unit tests, zero runtime dependencies.
- WorldQuant 2024: 166 submitted Alpha studies in the private research archive; competition ranks must use canonical evidence.
- Selected merged PR links, using titles and URLs verified on 2026-09-01.

6.4 Evidence that must remain bounded

- PDSH and PAT are internal systems, not production customer platforms.
- Deployed websites prove delivery, not traffic, adoption, or business impact.
- A merged PR proves a contribution, not employment or endorsement.
- Architecture proposals prove design judgment only after implementation and benchmark boundaries are stated.

7. Proposed technical interview architecture

7.1 Canonical trace identity

```
tenant -> project -> experiment -> run -> task -> turn -> step -> tool_call
```

Each event should include immutable IDs, timestamps, parent relationships, attempt/supersession semantics, model/tool/config version, input/output artifact references, error class, token/cost/latency measures, privacy classification, and release eligibility.

7.2 Event flow

```
Agent runtime
-> append-only event ingestion
-> schema validation + redaction
-> durable event store + artifact store
-> online metrics / alerting
-> replay and bad-case workbench
-> evaluation runner
-> baseline comparison
-> regression gate
-> release decision + receipt
```

7.3 Evaluation design

- fixed golden tasks plus versioned failure-case suites;
- deterministic tool fixtures when live systems would add noise;
- task outcome, trajectory quality, tool correctness, policy compliance, latency, token/cost, and human-rework metrics;
- stratification by task type and known failure mode;
- A/B test with frozen model/config/tool versions;
- confidence intervals or repeated stochastic runs where applicable;
- automatic gate plus human review for ambiguous or high-impact cases;
- rollback receipt linked to the exact failing baseline comparison.

7.4 Failure model

Layer	Example	Detection	Response
Input	prompt injection or malformed task	policy/schema gate	reject or isolate
Planning	wrong dependency	trajectory checker	retry with bounded policy or fail

Layer	Example	Detection	Response
	order		
Tool	timeout, invalid index, side effect ambiguity	typed result + idempotency key	supersede attempt; never double-commit
Model	hallucinated state or tool	trace/eval rule	classify bad case and add regression fixture
Platform	queue loss, clock skew, partial write	durable sequence and reconciliation	quarantine run and rebuild receipt
Evaluation	judge drift or leakage	judge versioning and holdout audit	invalidate comparison and rerun

8. Cross-border FX quant path

The FX role forms a second, differentiated path: banking domain context plus quant/data-product engineering. It asks for platform-data modeling, online strategy monitoring, an owned FX data system, customer FX profiling, and quantitative business support.

Pengyi has credible transferable evidence in banking, multi-source quantitative research, data quality, and AI-assisted workflows. The hard gaps are live FX strategy optimization, counterparty behavior cases, and the three-year gate. Therefore the correct action is a targeted conversation and a separate future PCV bundle, not dilution of the Harness Package 008.

See `TENCENT_FX_QUANT_PRODUCT_DUE_DILIGENCE_20260901.md` for the dedicated analysis.

9. Application campaign

9.1 Asset contract

- **PCV Package 008:** CN ATS master, EN ATS, CN profile, EN profile; one-page, current JD, verified claims only.
- **Application site:** four CV surfaces, evidence links, role-specific thesis, and claim boundary.
- **PIT Package 008:** technical Q&A, system design, behavioral, project defense, OSS defense, Tencent KYC, mock plan, scorecard, weakness tracker.

- **PEXT:** state transitions and evidence receipts; no `EXECUTED` state before the real confirmation exists.

9.2 Human Gate checklist

- [ ] Confirm role remains active and Shenzhen is selectable.
- [ ] Review every date, metric, link, and private/public boundary.
- [ ] Select CN ATS for primary submission; keep EN ATS ready.
- [ ] Confirm phone, email, location preference, notice period, and mobility.
- [ ] Complete one 45-minute system-design mock and one 30-minute OSS defense.
- [ ] Decide whether to request warm context before or immediately after apply.
- [ ] Submit through the official page and capture the confirmation receipt.
- [ ] Follow up at day 7 only if there is no response.

9.3 Immediate execution order

```
G0 Target captured COMPLETE
G1 Public research COMPLETE
G2 Fit decision COMPLETE: APPLY
G3 CV/evidence ready BUILDING
G4 Human review PENDING
G5 Application executed PENDING
G6 Follow-up PENDING
G7 Closed outcome PENDING
```

---

10. Final verdict

Primary: apply to the Harness role as soon as Package 008 passes Human Gate.

Adjacent: prepare Agent Infra and Data/Environment as the next two branches.

Quant: open a bounded FX quant conversation, but do not use the Harness CV.

Narrative: the differentiator is not generic enthusiasm for Agents; it is an evidence-backed ability to turn non-deterministic Agent behavior into governed, observable, testable, debuggable, and releasable engineering systems.

Sources

Official hiring

- Tencent Harness role
- Tencent Agent Eval Infra role
- Tencent Eval Observability role
- Tencent Agent Infra role
- Tencent Agent Data and Environment role
- Tencent Agent RL Framework role
- Tencent Sandbox role
- Tencent FinTech Agent role
- Tencent FX Quant Product role

Official/public technical context

- Tencent Hy3 release, 2026-07-06
- Tencent Q1 2026 results
- Tencent Cloud Agent Runtime
- Tencent-Hunyuan C3-Benchmark
- Tencent CUARewardBench
- Tencent AI-Infra-Guard

Candidate public evidence

- dsh-quant
- LightRAG #3717
- Vibe-Trading #1210
- CLI-Anything #421
- WeKnora #2776
- OpenClaw #122290