

AI 工具箱

客戶服務





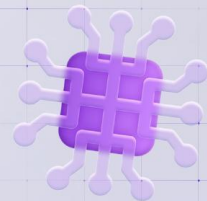
PRACTICAL AI IMPLEMENTATION

The Three Walls Every AI Project Must Knock Down

10+ years of production AI in Hong Kong — what it actually takes to go live

Rick Chan · Senior Project Manager, Mindlayer

Date: 6 October 2026





ABOUT MINDLAYER



Enterprise Class AI Solution Provider

Founded in 2015 by the first Google Developer Expert (GDE) in Machine Learning in Hong Kong



Delivered 100+ projects

Successfully completed 100+ projects across enterprise and government sectors.



Resolved 3M+ enquiries / year

AI/NLP-powered Chatbot and Voicebot resolving end-user enquiries at scale.



15+ Trusted Clients from Enterprise & Government

Tailored AI solutions delivered across multiple HK Government departments and enterprises.



Strong Track Record

Cyberport Incubatee (2018), CityU HK Tech 300 Angel Fund, TSSSU.

Built to meet the highest security and scalability demands of government and large enterprises



Build Your Own Virtual Assistant

Tailor-made Virtual Assistant for your organisation — automate business processes, increase productivity and enhance user experience with our **all-in-one Conversational AI Platform**.

All-in-one Conversational AI Platform



Conversational Platform

Understand what the user is saying and respond fast and precisely.



Data Crowdsourcing Platform

Enable high-quality data collection from your target group at scale.



Call Center as a Service (CCaaS)

Your call centre and IVRS provisioned in just a few clicks.



Live Agent Platform (TeamLayer)

Instant communication with customers across different channels.



Text, Voice & Data Analytics

Powerful analytics on conversations and customer data.



ML Professional Service

Tailor-made solutions for your end-to-end production system.



PROVEN EXCELLENCE

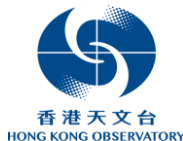
Trusted Across Government & Enterprise



Years of successful collaboration with our clients, delivering:

- Accurate, domain-tuned models
- Scalable, reliable systems
- High user satisfaction
- Recognised industry expertise

We empower our clients with proven, productive AI solutions.





Every AI Project Faces Three Walls

1



Knowledge & Data

Tacit know-how and messy real-world data are rarely ready to use

2



Environment

Limited on-prem GPU infrastructure, no external API, data cannot leave the site

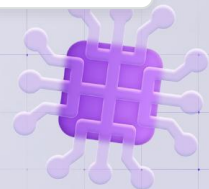
3



Behaviour

A general-purpose model will confidently answer anything — including what it shouldn't say

Each wall needs to be knocked down one-by-one.





Knowledge and Data Are Rarely Ready to Use

THE CHALLENGE



Frontline know-how isn't written down



Prompts need domain expertise



Real-world data is messy & mixed-language

HOW WE SOLVE IT

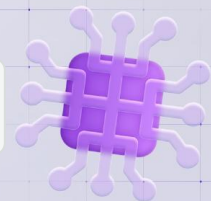


Interview frontline teams early to capture tacit know-how



Dedicated data-analysis phase with AI — efficient, at-scale review & grouping of real data

Note: building the knowledge base is not a one-off task — it needs ongoing upkeep.





Built for a Locked-Down Environment

THE CHALLENGE

- Limited GPU resources — no external API calls
- Real data must stay on-site; limits testing speed

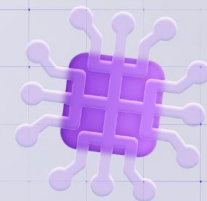
HOW WE SOLVE IT

- Benchmarked smaller / quantised open-source models
- On-site development & testing (or data-masking)

MODEL BENCHMARK — before vs. after

Model	Accuracy	Trade-off	Stage
Qwen (3B)	77%	Fast, low memory — least accurate	BEFORE
Qwen (14B)	78.41%	Fast, higher memory — forgets on long context	BEFORE
Qwen (27B)	85%	Slower, higher memory — most accurate, best on long context	AFTER

Chosen: 27B — best accuracy and long-context handling despite the higher resource cost.





Controlling What the Model Can Do

A general-purpose LLM will confidently answer anything — including what it shouldn't.

We call this discipline **Harness Engineering** — building guardrails around a model, not just training the model itself.



Input gate

Only answers questions inside the defined scope



Generation gate

RAG — only answers from verified documents



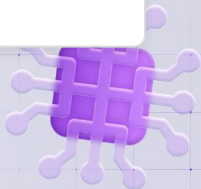
Output gate

Confidence routing + human approval before sending



Continuous monitoring

Every answer logged, auditable; corrections feed back





Five Questions Before You Start

1

What if it's wrong?

Set the acceptable error rate first — then pick the technology

2

Who checks the work?

Start with an internal use case with human approval

3

Where's the data?

Verified documents and real-life examples drive accuracy

4

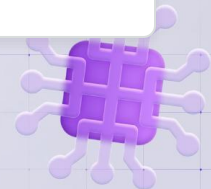
What must it connect to?

Identify existing systems; integration is usually the long pole

5

Where does it run?

On-premises, private cloud, or a local model for sensitive data





PRODUCT LIST



The Mindlayer Product Suite

When the technology barrier gets lower, real experience becomes the differentiator.



AI Virtual Assistant Platform

LLM · NLP · Text & Voice



Teamlayer — Live Agent

AI Copilot · Compliance Check



AI Knowledge Base

RAG · Multi-format Docs



AI Email Automation

Auto-classify · CRM Sync



AI Voicemail Automation

STT · Auto-summary



Voice Analytics

Sentiment · Trends · NER



Hireably

Digital Human · AI Personal Coach



AI Infrastructure

Lightweight LLM · 24/7 Monitoring



Knowledge · Environment · Behaviour

Knowledge

Build the base before you build the model

Environment

Work inside the constraint, be honest about the trade-off

Behaviour

Four gates — Harness Engineering, not an open-ended chatbot

hello@mindlayer.io · +852 3500 3910 · <https://www.mindlayer.io>

Thank you

