UndercoverAgent - An Agent in your Pocket

Machine Learning Singapore

- Martin Andrews = UndercoverAgent @ mdda.net

26-February-2026

About Me

- Machine Intelligence / Startups / Finance
  - Moved from NYC to Singapore in Sep-2013
- 2014 = 'fun' :
  - Machine Learning, Deep Learning, NLP
  - Robots, drones
- Since 2015 = 'serious' :: NLP + deep learning
  - Including Papers...
  - & GDE ML; ML-Singapore co-organiser...
  - & Red Dragon AI...

MLSG
MACHINE LEARNING
SINGAPORE

Outline

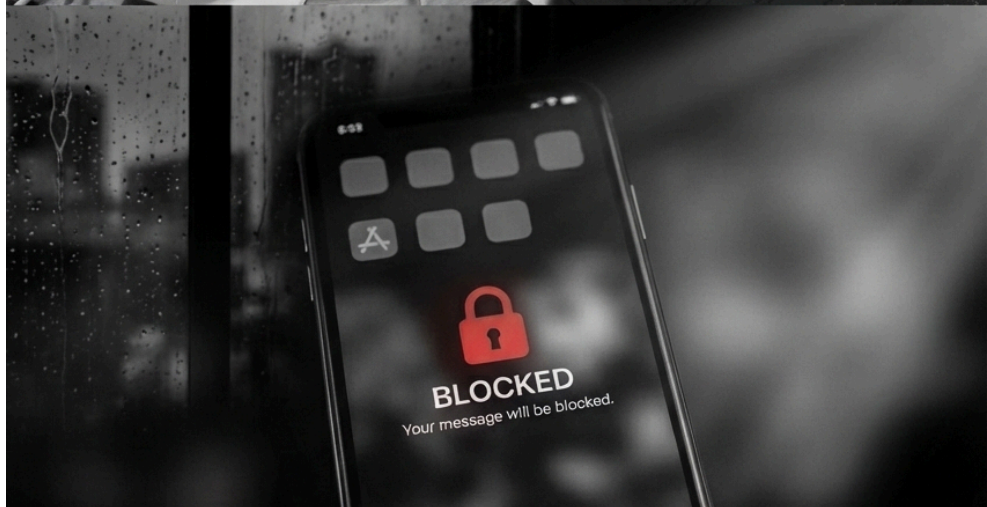
- Motivation
- Design Decisions
- ~~Vibe Coding~~ Agentic Engineering
- DEMO
- Next Steps & QR-code
- Wrap-up

MLSG
MACHINE LEARNING
SINGAPORE

Motivation

The Problem with AI Agents Today

- Most "agents" are either...
  - **Too powerful**: arbitrary shell access, credentials in env vars, no audit trail
  - **Too limited**: can't actually *do* anything useful
- OpenClaw / similar tools: great for developers, terrifying for everyone else
  - `exec` + persistent memory + your API keys = what could go wrong?
- Non-developers are left out entirely
  - The mental model of "shell access" was never theirs to begin with



MLSG
MACHINE LEARNING
SINGAPORE

What if the Phone **is** the Agent?

Not a remote client connecting to a laptop daemon

- The phone itself has:
  - context, credentials, action
  - is already trusted for banking, authentication, notifications
- The mobile OS sandbox is free security
  - iOS/Android give us App isolation already
- Native permission UX users already understand
  - eg: location, contacts, camera

The Core Insight

> The Request Broker is not just a security feature added to an agent.

- OpenClaw's magic = `exec`
  - enables everything
  - *also enables everything dangerous*
- UndercoverAgent's magic = structured API access
  - (potentially) same WOW moments
  - structurally cannot enable the dangerous things

"The LLM never sees your API keys"

"Competitive Landscape"

	OpenClaw	Nanobot / TinyClaw	UndercoverAgent
Target	Developers	Developers	Everyone
Platform	Desktop daemon	Desktop/server/device	iOS + Android
Shell access	✓	✓ (device)	✗ (structural)
Credential safety	env vars	env vars	broker injection
Installation	Tricky	Varies	AppStore
Non-dev UX	✗	✗	✓
Skill system	SKILL.md	SKILL.md	SKILL.md ✓

5,700+ community skills on ClawHub : ~50% are pure API wrappers ⇒ convertible

Design Decisions

Framework = React Native + Expo

Design Decisions

- Cross-platform from one codebase
  - iOS / Android / web
- Strong LLM training coverage
  - better vibe coding results
- Expo Go: zero-install prototyping
  - QR code → running app
- OS sandbox model
  - per-app isolation for free
- Expo Go has `react-native-webview`
  - WebView sandbox works immediately

MLSG
MACHINE LEARNING
SINGAPORE

Skill Storage & Ingestion

Design Decisions

```
documentDirectory/
  skills/
    pending/   ← awaiting approval
    active/    ← approved; loaded on startup
  workspace/
    AGENTS.md  ← injected into every system prompt
    USER.md
    HEARTBEAT.md
    memory/    ← agent writes freely here
```

- Human-authored:
  - fetch from URL, paste from clipboard, QR scan
- Agent-authored:
  - LLM output → `pending/` → approval gate → `active/`
- Same flow for all paths:
  - validation → approval gate → activate

Permissions Model

Design Decision

Binary "approve/deny" is the wrong abstraction

Tier	When	UX
Silent	Fits policy	Invisible (audit log only)
Notify + Undo	Consequential but allowed	Brief toast, 30s undo
Diff Approval	Outside policy	Full approval card
Policy Extension	New action class	Agent asks first

"The agent has permission to do exactly one thing without asking: add a UI card.
Everything else requires your approval."

The Schema Escape Pressure Problem

> The community *will* push for arbitrary JS execution.
>
> OpenClaw's own history proves this pressure is real and sustained.

- Answer: formal schema extension proposals
  - never escape hatches
- `ui.type: webview` is the right answer to "I need rich UI"
  - sandboxed, not escaped
  - Sub-App javascript can only access web via proxy
    - which is injected from the main App

Heartbeat Architecture

Design Decision

- **Scheduled**: fire at fixed intervals — daily briefing, weekly review
- **Event-driven**: poll cheaply; only call the model when something changed

```
poll:
  - source: telegram.updates
    condition: count > 0
    trigger: call_model
  - source: url_fetch:https://feed.example.com/rss
    condition: item_count_changed
    trigger: call_model
```

Key principle: platform code evaluates the condition; the model evaluates the *meaning*

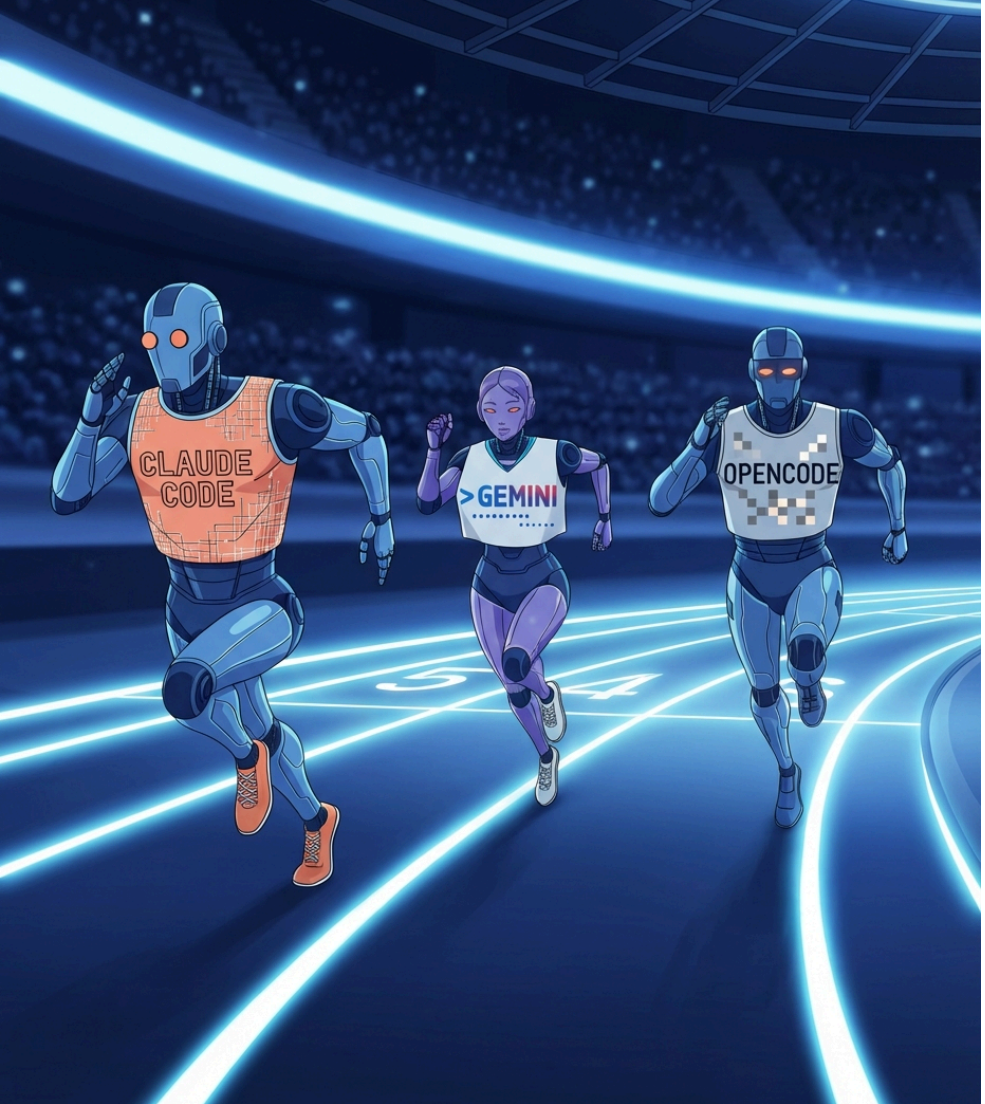
iOS caveat: silent push is advisory. Graceful stateless catch-up on next wake.

MLSG
MACHINE LEARNING
SINGAPORE

~~Vibe Coding~~

Agentic Engineering

MLSG
MACHINE LEARNING
SINGAPORE

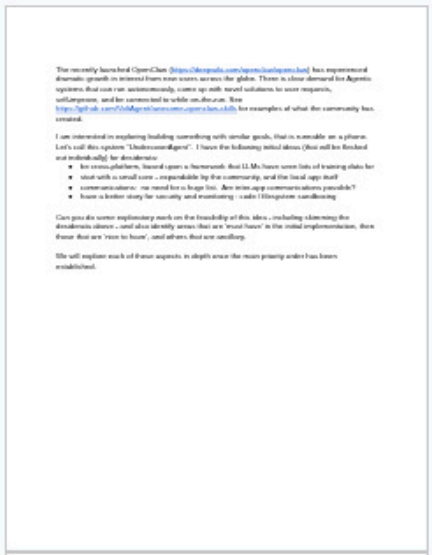


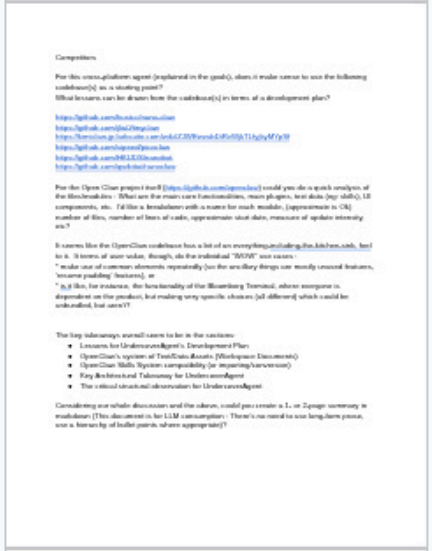
The last few weeks...

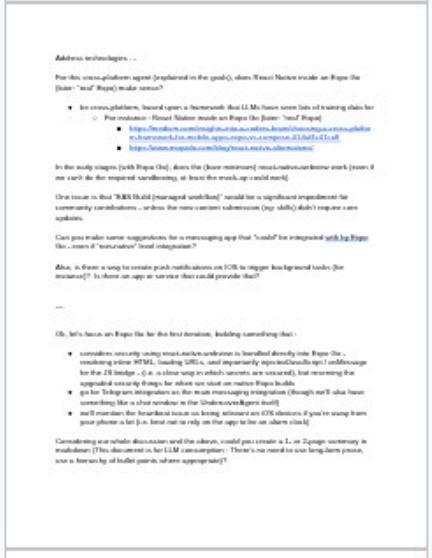
- Gemini CLI
  - Created `project tree` for Geany
    - `geany` = cross-platform editor
    - GTK / C plugin (backwards compatible)
  - bit of a struggle, TBH!
- opencode : MiniMax 2.5
  - Added new functionality to https://mdda.net/
    - grappled with `next.js`
    - tried to implement bulk changes without asking
- opencode : MiniMax 2.5
  - Started development of MLSG theme for `slidev`
    - I wanted to have rational file layout
    - also: add ';' comments, ++
  - MiniMax actually got very confused
    - Claude Code one-shotted issues...

Process for UndercoverAgent

- Wrote prompts into Google Doc
  - Did planning in a Project
    - output format = Markdown
    - put into `git`
    - and back into project as source documents
  - Asked for 5-day hackathon plan
    - defined DEMOs v early
- Using markdown, moved to Claude Code ($)
  - Started on Day 1 of plan...
    - had to find an old phone
    - figure out a new Google account
  - actual progress started...
- Finally, back to Project
  - "Write me `slidev` slides"



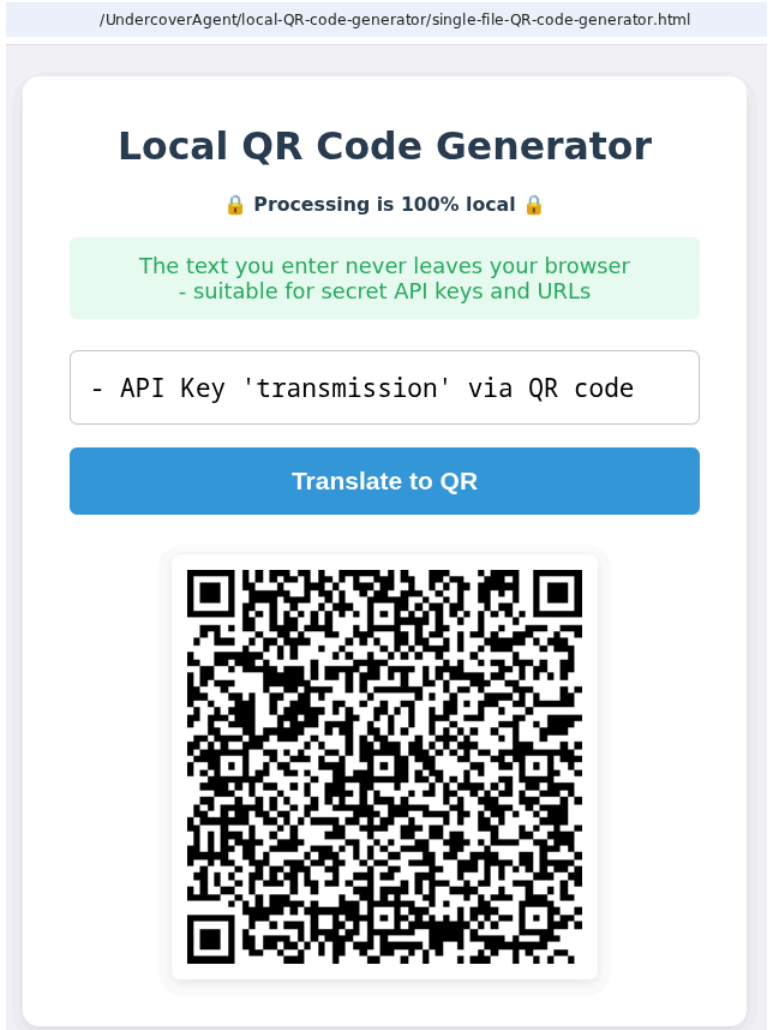




MLSG
MACHINE LEARNING
SINGAPORE

What Worked Well

- API Key 'transmission' via QR code
  - One-shotted by Gemini Pro 3.1
    - single page local QR code generation
- Structured design docs
  - LLM could be re-contextualised quickly across sessions
- Skills-as-manifests
  - proven cross-ecosystem pattern
  - Python, TS, Go all converged on it
- `react-native-webview` fetch shim pattern
  - surprisingly clean to implement
- Expo Go QR workflow
  - live changes on phone
  - non-devs will be able to run the app immediately
- The `pending/` → `active/` folder split
  - enforced the approval gate structurally



MLSG
MACHINE LEARNING
SINGAPORE

What Was Harder Than Expected

- State management for demos
  - resettable named workspace snapshots were essential
    - system didn't see how they would interfere
- Human readability
  - discovered `AGENTS.md` was being maintained
    - rather than a single source of truth
- Telegram REST inteface
  - Gaslit by system for several hours
    - had missed mechanics of Telegram handshake

Demos

Live Demo

Three Tracks

Track	Shows
Memory & Personality	Agent builds a persistent model of you through conversation
Problem Solving + New UI	Agent generates novel capability and UI from natural language
Proactive Agent	Agent acts on your behalf without being asked

[Live demo on actual device]

Demo Track 1: Memory & Personality

- Fresh state: agent knows nothing
  - Agent asks 5 conversational questions
  - Writes to `USER.md` and `SOUL.md` in real time
- Workspace file viewer shows population happening live
- Follow-up message reflects what was said 30 seconds ago

"This is what it remembers. It carries this into every future interaction."

MLSG
MACHINE LEARNING
SINGAPORE

Demo Track 2: The Market Bell

- Prompt: *"I want a card showing when major stock markets are open"*
- Agent generates skill manifest
  - with `allowedHosts`,
  - UI schema
  - refresh interval
- Approval card shown:
  - audience sees the broker/permissions declaration
- Approve → card appears:
  - SGX, LSE, NYSE, TSE — status + countdown

MLSG
MACHINE LEARNING
SINGAPORE

Demo Track 3: Proactive Relationship Concierge

NOT ATTEMPTED HERE

Pre-populated `USER.md` contains:

```
- **Marcus** — close friend, Amsterdam.
  Last contact: ~6 weeks ago. Usual: every 3–4 weeks.
```

On heartbeat trigger:

> *"You haven't spoken to Marcus in about 6 weeks — longer than usual. He mentioned the Amsterdam move last time. Want me to draft a quick message?"*

Telegram message arrives on phone during the demo.

What’s Next

Near-term Roadmap

What's next

v0.2 (post-MLSG tier)

- multi-file SKILLs
- test out HEARTBEAT etc
- permissions with rewriting
- SKILL importer
- compositional SKILLs

v0.8 (native build tier)

- Full permissions constraint enforcement
  - content-filter, rate-limit, recipient-filter
- Real CSP enforcement
  - native WebView config
- Reliable iOS background fetch
  - `UIBackgroundModes`

MLSG
MACHINE LEARNING
SINGAPORE

Long term Roadmap

What's next

v1.0

- On-device inference
  - Apple Neural Engine / Snapdragon NPU
- `expo-secure-store`
  - biometric lock for OAuth tokens
- Audit log viewer + rate limit dashboard
  - Enterprise features?
- Multi-agent routing
  - work / personal contexts

Call for Contributors

If you're interested in helping out:

MLSG
MACHINE LEARNING
SINGAPORE

UndercoverAgent Interest Form

If you're interested in becoming a contributor for the MobileClaw / UndercoverAgent project (and getting early access to the codebase, for experimentation, new features, testing, etc)

martin.andrews@gmail.com Switch account

Not shared

* Indicates required question

How did you hear about the Project? *

- Machine Learning Singapore MeetUp
- YouTube Video
- X / Twitter
- Linked-In
- Other:

What most interests you about the project *

- Just seems cool
- Seems like a better platform than others
- Actually have a use-case
- Other:

Please give us a way to get hold of you (eg: email, linkedin, ...) *

Your answer

Submit

Clear form

Never submit passwords through Google Forms.

Wrap-Up

- UndercoverAgent: an AI agent on your phone that can actually do things
  - and cannot betray you while doing them.
  - "The LLM can never see your API tokens"
- Agentic Engineering
  - everybody should try latest models
- Self-personalising software FTW!

MLSG
MACHINE LEARNING
SINGAPORE

Machine Learning SG MeetUp Group

- Next Meeting = late-March-2026 @ Google
- Topic(s) : TBA
- Typical Contents :
  - Talk for people starting out
  - Something from the bleeding-edge
  - Lightning Talks
- MeetUp.com / Machine-Learning-Singapore

MLSG
MACHINE LEARNING
SINGAPORE