MEETUP / INNOQ OFFENBACH / 11.11.2026

Trust but Sandbox

Using Coding Agents Safely with Docker Sandboxes

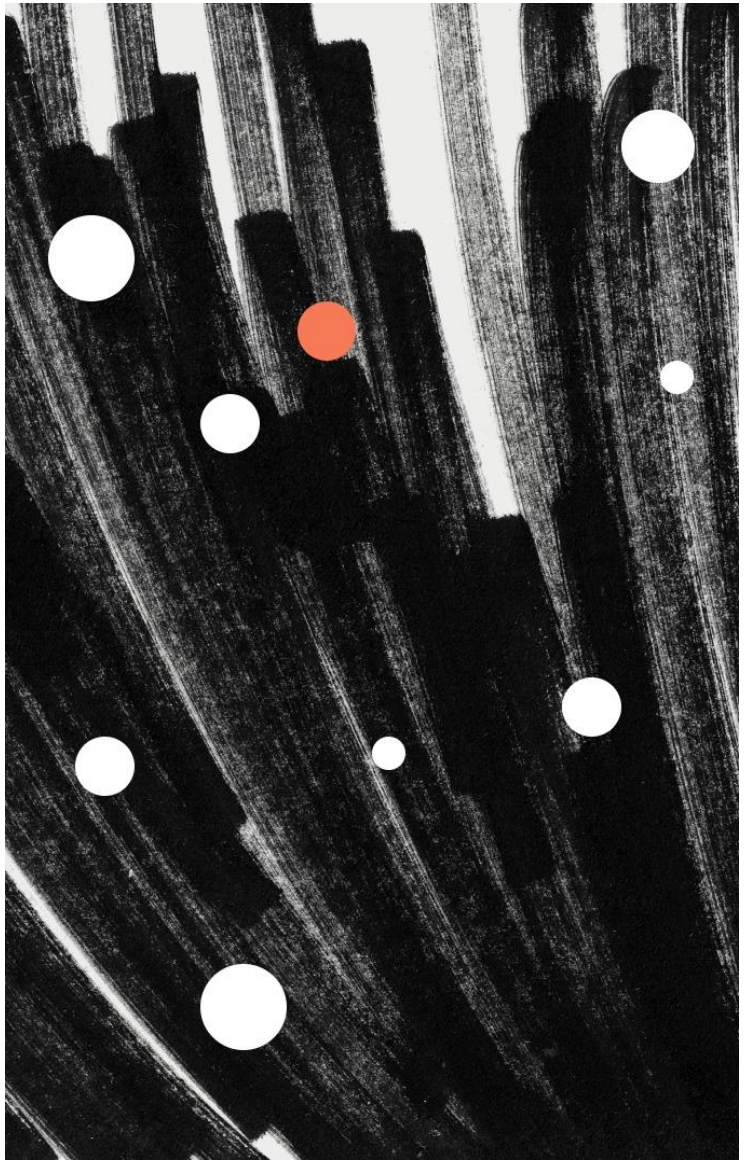
INNOQ



MICHAEL KRÄMER
PRINCIPAL CONSULTANT

Hi 👋

- Michael
- Dipl.-Informatiker (yeah, long ago..)
- Principal Consultant / Meticulator at INNOQ Zürich
- Custom Software, Modernization, Cloud-Migration, Architecture Consulting, Data Architecture, iSAQB & other Trainings



INNOQ

For over 25 years, we work in implementing IT systems of all shapes and sizes.

Our expertise is grounded in years of hands-on experience in software development.

We don't see **technology as an end in itself**, but as an **enabler for solving real-world problems.**

www.innoq.com
info.ch@innoq.com
Tel. +41 41 743 01 11

We're here to help

Software Architecture & Consulting: What's the best way forward? Which approach fits best to your goals?

Implementation: Complete or partial implementation of your IT project by one or more interdisciplinary expert teams.

Cloud Native & DevOps: We can shape your solution to run on modern infrastructure – either by you or by us.

Trainings: We offer many iSAQB and other trainings to enable your team.

AI-readiness: Our experts can help modernize your development organization as well as with making your data accessible for AI.


Agents can do
almost anything.

Full access 🤔

- Read and write on the **entire disk**
  - not just the project folder
- Execute arbitrary shell commands
- Network access without restrictions
- Read your env 🔑 🎟️
- Use your browser with all those 🍪

Behavior is determined ***by the model*** — and is ***barely verifiable.***



Agents are creative

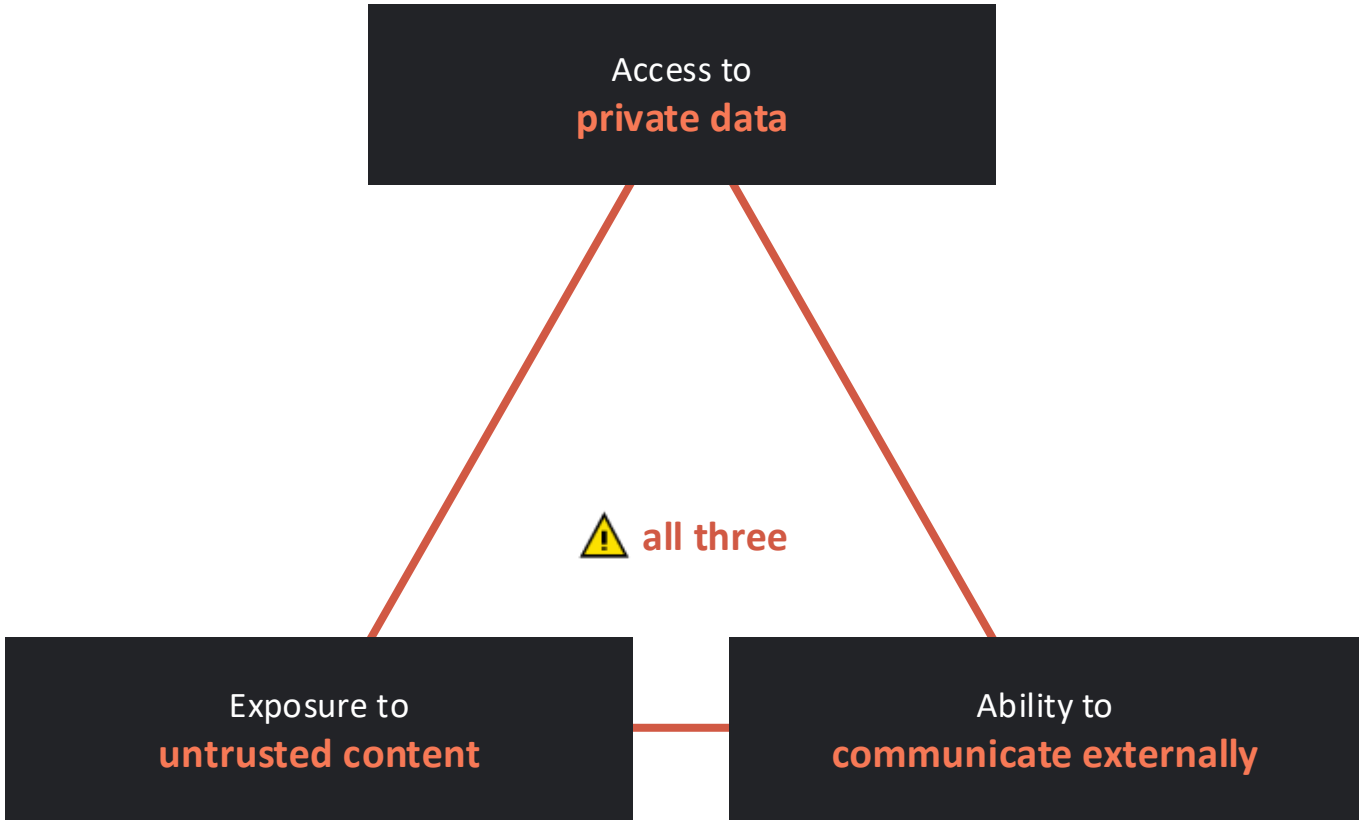
- Confirmation prompts before tool use are wired in **client code** — not in the model 👍
- Agents **find ways around** restrictions when it serves the goal ⚠️
- Agentic work requires **autonomy** for the agent 🚀
- Unstructured input from many sources: not all of them are **trustworthy** ⚠️ ⚠️

Prompt injection via MCP servers, APIs, files — not a theoretical risk.



The Risk Triangle

(aka Lethal Trifecta for Agents)*



Two of these aspects are significantly less risk than covering **all three.**

* A term coined by Simon Willison (simonwillison.net)


Solution:
Docker Sandboxes (sbx)



Docker Sandboxes

What is it?

- Solution from Docker Inc., based on **MicroVM**
- **No shared kernel**, but little overhead
- Lots of conventions for easy use
- Stable, but very active development

Where does it help?

Area	Protection
Filesystem	Only the mounted project directory is accessible
Network	Configurable policies (Open / Balanced / Locked Down)
Credentials	Proxy on host injects tokens — never visible inside the VM
Host system	Completely isolated from the VM

That should ***provide enough control*** for YOLO mode.

`--dangerously-skip-permissions` is the default in sbx

Where does it hurt?

- Copy & Paste
- Updates
- Regular license checks
- First incompatibilities (v3 kits)

- It’s one wrapper more 🙄




Live Demo

Pro & Con

Pro

- Strong isolation with low overhead
- Credentials never visible to the agent
- Network policies granularly configurable
- ASCII dashboard for overview
- Works with many agents

Con

- Still very new, moving target
- Many projects = many VMs
- No file masking in project folder (workaround: symlinks)
- IDE on host / agent in VM: separate network can interfere

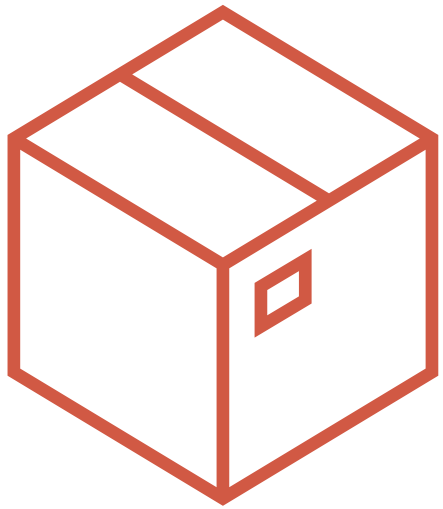


Configuration & Customization

- **Easy:** agent settings in the project folder for context and constraints
  - `.claude / .agent folder, skills, …`
- Network policy overridable per project
- Status bar and agent behavior customizable
- Custom sandboxes based on other images or other agents, like pi
  - github.com/docker/sbx-kits-contrib

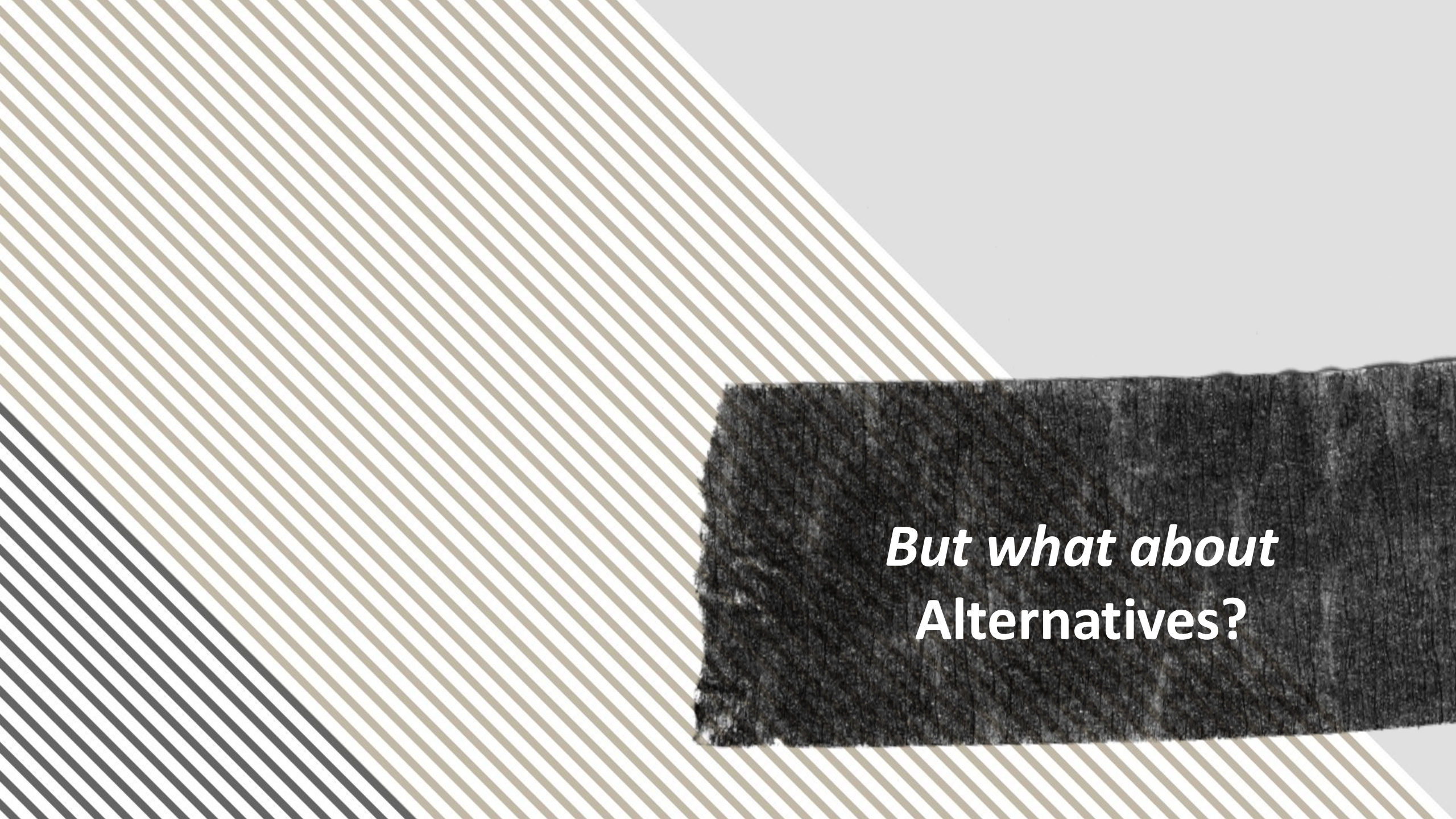
Declarative Packaging with Kits

- Describe the agent **in a file**, not by hand
  - image, tools, skills, network, credentials
- **Versioned in git**, reviewed like code
- **Reproducible**: same setup for team and CI



Side effect of sbx: **Kits** already deliver something like that: a pre-configured sandbox environment you can share and reuse

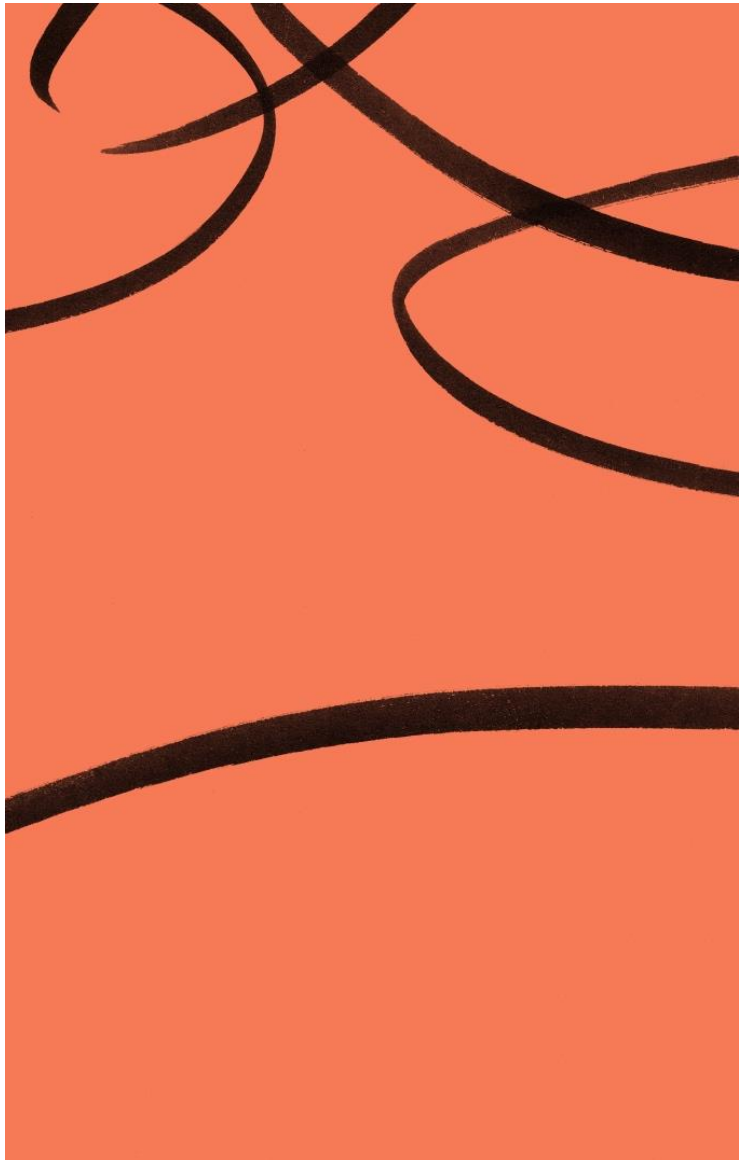
Shift with v3: from **declaration only** (v1/v2, evaluated at sandbox start) to a **build step**, like docker build


But what about
Alternatives?

Alternative:
Docker Container

- **Shared kernel** — weaker isolation than MicroVM
- No dedicated Docker daemon → Docker-in-Docker required for container builds
- Manual setup: no ready-made agent template, no credential proxy *(That statement might be outdated)*

Verdict: feasible, but significantly more effort for less security



Alternative:
Claude --sandbox

- Agent-side feature, **not system-level** protection
  - *at least it didn't work for me*
- Restricts tool use in the agent — but no real isolation
- No protection against prompt injection from external sources
- Vendor lock-in

Verdict: useful as a second line of defense, does not replace a system-level sandbox



Alternative:
Brig.sh

- Also **MicroVM**: own kernel, only the named project mounted
- **Open source** (Apache 2.0), macOS and Linux
- **Credentials proxy**: keys off the sandbox
- **Network policies**, default deny
- Profiles for Claude Code, Codex, Gemini, opencode, …

Verdict: so far the closest match to sbx. Not tested by me yet (no time)

brig.sh

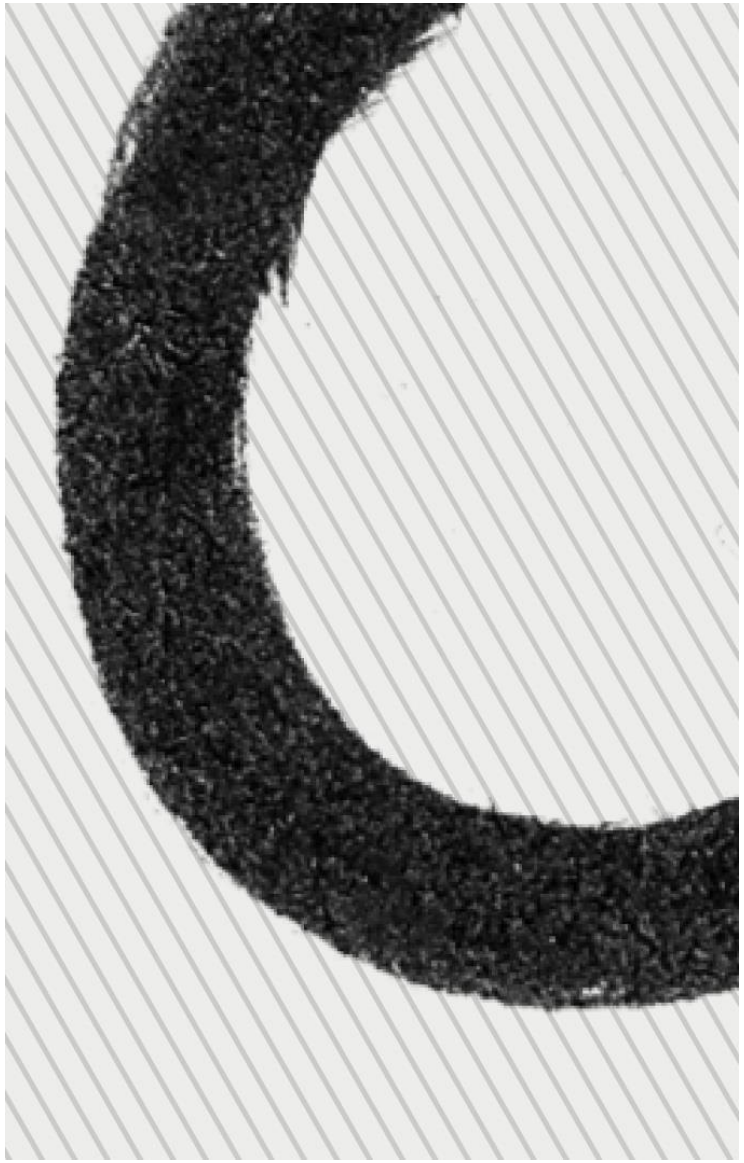


Alternative:
Cloud Sandboxes

GitHub Codespaces, E2B.dev and similar

- **Real isolation** in the cloud — no local risk
- No access to local resources (filesystem, local services)
- Latency and costs depending on usage

Verdict: great for teams and CI, less practical for local development





Outlook & Links

- **Kits:** pre-configured sandbox environments
- Git clone mode
- Image pasting (currently doc-only 🤣)
- File masking in project folder (still open)
- VM update automation?

https://codefloe.com/michel/sbx-kits

github.com/restyler/awesome-sandbox

Agents need freedom. Sandboxes give them that freedom — *without putting your system at risk.*

- `sbx run claude` — one command, real isolation
- YOLO mode ✅ — but in the VM, not on your laptop
- Still young. Changes, but reliable for everyday use

Docs: docs.docker.com/ai/sandboxes

Thanks for listening! Questions?!



Michael Krämer
michael.kraemer@innoq.com

+41 77 422 40 61

https://www.linkedin.com/in/mkraemerx

www.innoq.com

innoQ Schweiz GmbH

Schutzengelstr. 57
6340 Baar
+41 41 743 0116

Hardturmstr. 253
8005 Zürich