

Project Ground Truth

A verifiable, cross-indexed encyclopedia of real martial-arts techniques — using fighting games as the discovery layer. **V1 DESIGN DOC · DRAFT**

status DRAFT / RFC — for review, not commitment · **date** 2026-07-24

authors David Kwok Ho Chan · AI Sifu Claude · **thread** Project Ideas · **public name** "Martial Simulacra" · **codename** "Ground Truth"

builds on Fei Kune Do method (verification) · movement-as-data (schema+engine) · the mcp-sport vault (sources)

1 TL;DR

Fighting games are full of real martial-arts techniques wearing flashy names — and full of pure fiction wearing the same names. **Ground Truth is the reference that tells them apart**, technique by technique, backed by primary sources and shown with a human-executable clip. It is organised **by art, not by game**: a Karate practitioner gets every karate technique across every game in one place, each verified and demonstrated. The method is already proven — we ran it three times (Reina/Taido, Xiaoyu/Bagua, Eddy/Capoeira) inside the Practice Book. This doc scopes a **V1: one art (Karate), one filled cell of the matrix**, on the stack we already own.

2 Problem & Motivation

- **The question is real and unanswered well.** Millions play these games; a recurring curiosity is *"is that an actual move?"* The existing answers are clickbait ("Top 10 Moves IRL") with no verification and no primary sources.
- **The knowledge is scattered and locked by language.** As Source Log 1 showed, the real material often lives in the art's native language and in out-of-print manuals — exactly what the vault already holds.
- **Nobody indexes it by art.** Game-by-game listicles exist. A cross-game index — *"here is every karate technique any game has ever used, and here's the real one"* — does not.

3 Goals & Non-Goals (V1)

Goals	Non-Goals (explicitly deferred)
✓ Fill one cell: Karate, ~8–12 techniques, fully verified + demonstrated. ✓ Prove the production loop end-to-end (research → verify → clip). ✓ Formalise the data model (art × game × technique). ✓ Answer the two killer unknowns: <i>is the clip pipeline sustainable</i>, and <i>does anyone care</i>.	✗ Monetisation / pricing / a store. ✗ Multiple arts or the full cross-index UI. ✗ Filmed video <i>at scale</i> (animation-first in V1). ✗ Fantasy-heavy games (Mortal Kombat et al.). ✗ Any use of a game's trademarked branding or footage in a paid artefact.

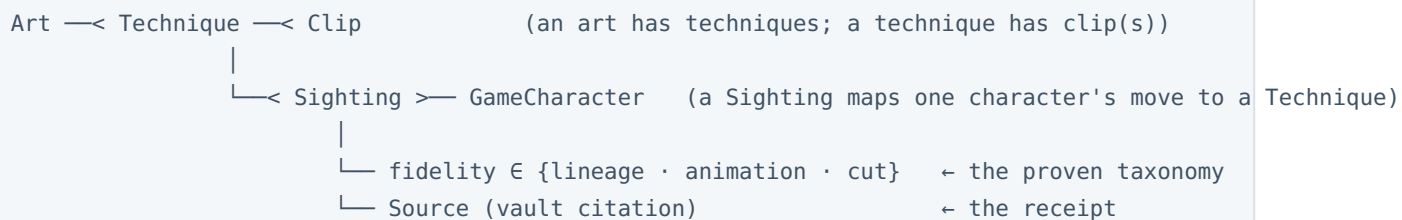
4 Why us — the moat

Anyone can list moves. Four things make this defensible, and they are things we *already have*:

- **A verification method, proven 3×**. Capture movelist → extract names → cross-check the vault → classify → breakdown. It already caught a real error (the *Senjomanji* misfiling) — that self-correction *is* the credibility.
- **A primary-source vault**. The Kyohan, the Gairon, Capoeira 100, the Shotokan syllabus — receipts nobody else brings.
- **A practitioner author** who trains multiple arts and can perform/validate.
- **The convergence lens** — "one law, three languages" — that no fight-game channel has.

5 Core design — the data model

The product *is* a database; the media is a view on it. This is the movement-as-data schema, finally with a reason to exist:



Entity	Key fields
Art	name · language · lineage notes (e.g. 空手道 Karate → Shotokan / Gōjū-ryū)
Technique	real name (native + romanised) · family · description · source refs · difficulty
Sighting	game · character · in-game name · input · fidelity · frame-data snapshot · screenshot
Source	vault collection · page/loc · quote — the primary citation
Clip	type ∈ {animation · self-filmed · collab · licensed-link} · asset path · angle(s)

Why art-centric wins: the same `Technique` row can carry *many* `Sighting`s (Lidia's version, Ryu's version, Makoto's version) and *one* canonical clip. Content compounds instead of repeating; each new game *enriches* existing rows rather than starting a new list.

6 The content pipeline (already proven)

```
[capture movelist] → [extract names] → [cross-check vault] → [classify: lineage/animation/cut]
→ [write breakdown] → [attach source citation] → [produce clip] → [publish view]
```

Every arrow except the last two has been executed three times already. V1's job is to harden the last two — **clip** and **publish** — into a repeatable step.

7 Production — the clip strategy

The historical failure point of this genre is *sustained video*. We de-risk it by **not depending on filmed video for V1**, and treating clips as a spectrum:

Clip type	IP	Scales	Role
animation	 owned	∞	V1 default. The stick-figure engine — zero IP, on-brand, infinite. Poses render clean; rotational moves are the known weak spot (a v-next task).
self-filmed	 owned	low	Hero moves the author can throw. Authenticity, no cost.
collab	 owned	med	Local dojo / students, for moves outside the author's range. Exposure-for-footage.
licensed-link	 link only	high	Point to existing tutorials / Creative-Commons clips. Never reproduce inside a paid artefact.

8 IP & legal — the gating risk

△ BLOCKER — must be resolved before any paid release

Selling anything branded as "[Game] DLC", using trademarked character names as the product, or embedding in-game footage in a paid artefact, invites a takedown and is likely infringement. **This is a hard constraint, not a preference.**

The reframe that clears it: the product is *real martial arts* — techniques and their real names are **public domain, not copyrightable**. Games are referenced **nominatively** ("the Shotokan behind Tekken's Lidia") and cited as *sightings*, not reproduced. All clips are **owned or linked**. The art-centric framing isn't just better product design — it's the legal posture. *Get the final form reviewed by an IP lawyer before charging money. Not legal advice.*

9 Tech stack — reuse everything

Need	Reuse
verification	ChromaDB vault (<code>mcp-sport.db</code> , <code>mcp_ma_theology.db</code>) + the MCP query tools
clips (V1)	the stick-figure engine (<code>~/ .claude/skills/stick-figure</code>)
documents	weasyprint house pipeline (this doc is proof it works)
distribution (V0)	the Hugo blog — free, already a funnel
index (new)	a structured file — start as JSON/SQLite; grows into the queryable front-end later

Net new build for V1 is small: one schema file + a render step from schema → page. Everything else exists.

10 V1 scope — the Karate cell

- **Art:** Karate (best-documented; vault has `Shotokan_Karate_Syllabus` , `Karate_for_Life`).
- **Primary sighting:** Lidia Sobieska (Tekken, Shotokan). **Secondary:** the Mishima karate-family (Kazuya/Reina) and, if reached, Makoto or Ryu (Street Fighter).
- **Deliverable:** ~8–12 techniques, each with real name + source citation + fidelity rating + a stick-figure clip; 1 stretch *self-filmed* hero move.
- **Format:** published free to the blog as the V0 view — the funnel, not the product.

11 Phasing

Phase	Output	Gate to next
P0 now	This design doc + the styles×games map (appendix)	David sleeps on it
P1	The Karate cell, animated, blog-published	Is the clip loop sustainable? Any traction?
P2	Second art + formalise the index (movement-as-data DB)	Does content compound as designed?
P3	Cross-index front-end + IP/legal review + monetisation decision	Only here does money enter

12 Success metrics

- **P1 primary:** can one person produce a full animated cell in a bounded time (target: a weekend)? If no → the model needs collab/animation-tooling before scaling.
- **P1 signal:** any organic engagement on the free cell (views, shares, a single "finally someone verified this" comment).
- **Leading health metric:** content *compounding* — does adding Street Fighter *enrich existing karate rows* rather than spawn a parallel list? If yes, the art-centric thesis holds.

13 Open questions — to sleep on

1. **Brand relationship:** is this a Fei Kune Do sub-project, or its own standalone brand? (Credibility borrows from FKD; reach may want independence.)
2. **Codename:** "Ground Truth" (ML/verification) vs "Movepool" vs "Kata-log" vs something native-language.
3. **The product object:** web database · pack · card deck · course · YouTube series — which is *the* product vs a view?
4. **Animation ceiling:** is stick-figure enough to be compelling, or is a richer animation format (the long-standing GIF idea) a P1 prerequisite?
5. **Entry art:** Karate is safest — but is **Virtua Fighter** (every character a real documented style) the higher-leverage *second* move, given credibility is the moat?

A Appendix — the styles × games map (raw material)

Game	Realism	Richest real-style veins
Virtua Fighter	● highest	<i>Designed</i> around real styles: Akira=Bajiquan · Pai=praying-mantis · Lau=tourou-ken · Jacky=JKD · Wolf=pro-wrestling · Jeffry=Pancrase · Aoi=Aiki-jūjutsu · Lei Fei=Shaolin · Shun=drunken boxing. Best verification mine.
Tekken 8	● semi-real	

Karate-family (Lidia=Shotokan, Mishima) · Reina=Taido · Eddy=Capoeira · Xiaoyu=Bagua · Leroy=Wing Chun · Paul=Judo · Hwoarang=TKD. *(Already prototyped.)*

Street Fighter	● grounded-ish	Goldmine: Ryu/Ken=ansatsuken · Makoto=Rindōkan karate · Fei Long=JKD (Bruce Lee) · Dhalsim=yoga · Sagat=Muay Thai · Zangief=sambo · Chun-Li=Chinese MA.
Mortal Kombat	● fantasy-heavy	Mostly "cut." Selective keeps (Liu Kang, Kung Lao, Kenshi=kenjutsu). High reach, low credibility — the honey, not the meat.
Sifu · others	● researched	Sifu=Bak Mei (obsessively researched) · Samurai Shodown=weapon arts · KOF · Absolver.

"**Martial Simulacra**" (working codename: **Ground Truth**) — **V1 Design Doc (DRAFT)**. Living planning artefact, shared building-in-public alongside the *Martial Simulacra* post; a draft for review, not a commitment. Verification method and the fighting-game→real-art mappings are demonstrated in Practice Book Ch. 12–14. Game titles, character names and style attributions are used descriptively for planning; nothing here reproduces game assets. The IP framing (§8) is a working position, not legal advice — a qualified review is a hard prerequisite before any paid release. Archived at [~/Sync/ProjectIdeas/](#).