StimmFolgen

Technical report - Hack Apertus Online, Track 2B - 15 October 2026

1. Follow a vote, understand its evidence

StimmFolgen helps people find Swiss votes, track those that matter and return to their meaning and later developments. My votes, Upcoming votes and Search lead to Explore my vote: explanation, dated sources, official updates, related votes, perspectives and optional private notes. Tracking never implies a Yes or No choice. The responsive interface starts with the explanation; AI questions and detailed evidence are optional.

The catalogue has six selected federal votes. The 13th AHV pension initiative of 3 March 2024 has five human-accepted, source-bound statements. Homeownership tax reform of 28 September 2025 is a second genuine booklet case: four candidate statements, source checks and live-model evaluations are prepared. Its human wording review is pending, so only official information is public. Private-preview tests do not imply public admission.

Apertus provides live interpretation of the reader's public question to find relevant evidence. The app displays unchanged ledger wording with source links, rather than AI-written facts or quotations. A reader can see what a proposal defines, what an actor forecast and what remains unknown. StimmFolgen neither recommends a vote nor invents a personal tax or salary charge.

The public app uses Cloudflare and Public AI Apertus v1.5 8B without a model download for visitors. The optional on-premise route uses local Apertus and a Docker app. Ordinary screens omit engineering task codes, hashes and failed-trial commentary; those remain in separate reviewer documentation. This is a tested civic-information prototype, not comprehensive policy coverage.

2. Source and publication controls

The original AHV Federal Chancellery PDF has SHA-256 53c7cc72c28f80969ed56d7813ad8c05e505b1817923418d0dfa45368a6bd07e, 759,783 bytes and 32 pages. Its 91 immutable extracted spans bind text, page and character positions. Apertus selects IDs and typed fields; quotations resolve from those spans. Independent bounded semantic checks cover values, actor, type, missing facts and unsupported assertions. Evaluation references are coding-agent-authored development controls, not an independent academic benchmark.

The booklet's content cutoff is 29 November 2023. An exact-byte archive witness establishes availability by 11 January 2024, 12:11:17 UTC. The accepted public review frame is 2 March 2024, 23:59:59 CET. Content cutoff, availability and retrieval dates remain distinct. The bounded AHV certificate requires authentic source/model evidence, accepted statement bindings and passing controls before publication.

The second official booklet has 48 pages and SHA-256 39c020145299913685e4d0b6899fa23bb1f94938652a417e0c1da08d316ade8a. It was inspected on 15 October 2026; its printed editorial cutoff is 20 June 2025. No exact-byte pre-vote archive witness has been established. Its candidate is therefore a document explanation, not a historical-availability proof. The formal second-home tax ballot is linked to the wider homeownership reform, not a separate referendum on both acts.

The additional-case registry requires real PDF and exact-span checks plus a human-user acceptance bound to the statement hash. Candidate model output never supplies that approval. Later official results and implementation updates are separate dated records, not retrospective model predictions. Their checked web excerpts have text hashes, not preserved original HTML-byte provenance.

3. Real Apertus and on-premise execution

Local inference ran on M3 Pro, 18 GiB RAM, macOS 26.6.2 arm64. The downloaded community conversion tokimoa/apertus-v1.5-8b-mlx-4bit is pinned to revision f08ed532a657bef7d322bd0cbca43dac13a9d0a3, declaring upstream swiss-ai/Apertus-v1.5-8B at a411d838600baf0e3635a3daf66fb7c55fc97bb6. Its text-only safetensors occupy 4,530,491,646 bytes; SHA-256 d0478edeb06bf18e5f4f7b50db77d89b839730b2d16cb1020015e6099e42074b. Independent equivalence to official BF16 or multimodal capability is not claimed.

A fresh public clone at 917e77c received a new Python environment: MLX 0.32.0, mlx-lm 0.31.3, transformers 5.15.0 and PyMuPDF 1.28.2. Pinned model artifacts were copied using APFS from existing local downloads and reauthenticated. No new remote weight download or second physical machine was tested. No-cache Docker application builds succeeded.

make run ONPREM=1 starts native MLX on 127.0.0.1:8768 and a non-root read-only Docker app on host loopback. Docker Desktop's bridge reaches the model; no tunnel, public Mac listener, API credential or hosted fallback is used. Model identity is checked before the app starts. Setup/build require network; this is not an air-gap claim. Native GPU inference is outside Linux Docker. The default make run remains hardware-free source replay, explicitly distinct from live inference.

Native MLX lacks the hosted API's schema-constrained decoding. The local adapter uses small advice/search tasks and shared source retrieval/directness checks, retaining original question terms. Application policy refuses recommendations and observed later outcomes. A freshly installed browser answered the pension-recipient question locally in 7.522 seconds. Phone-sized and desktop private-preview sessions also returned the correct AHV attribution and second-vote source passages. Timings on one Mac are not a throughput guarantee.

4. Results, failures and private state

The original AHV release suite used real local completions and passed seven independent bounded trials, including actor conflict, fabricated quote, personal-cost and later-knowledge controls. Earlier direct full-booklet and generated-quote protocols failed; those traces remain. The actor result is application-composed from independent model tasks, not a single generated answer or a hardcoded expected actor.

The new known development matching set passes 20/20 locally: 13 AHV and seven second-vote questions. Sixteen questions invoke native Apertus; four intentional advice/later-event controls are application refusals with no model call. Hosted Public AI v1.5 8B passes 45/45 known development questions, including seven second-vote candidates. These are inspected development sets, not held-out accuracy or local/hosted protocol equivalence. The first second-vote extraction invented references and overstated a legal condition; it was rejected. A prior local matching run failed 2/13. All failures are preserved.

Final deterministic checks pass 108 Python tests and 29 compiled Worker tests, including actual workerd privacy rejection, request serialisation and persistent quota accounting with a synthetic provider. WebCrypto checks cover nonce freshness, wrong-password/tamper rejection, concurrent saves, lock/reload and encrypted backup recovery. Synthetic tests are never described as model inference.

Agent-operated browser checks cover a fresh local AHV installation plus both cases in a private candidate preview at actual CSS 389x844 and 1280x800. Search, track, sources/dates, updates, related votes, perspectives, private notes, backup recovery and reload locking were exercised. The preview is labelled pending and is not public release proof. A delayed answer is bound to its case/source/wording to prevent cross-vote display. No physical phone, independent usability study or comprehensive accessibility audit was performed.

5. Deployment, limitations and reproduction

Visit-only tracking is default. Optional persistence uses PBKDF2-SHA256 (600,000 iterations, random salt) and AES-GCM-256 with fresh nonces. Only authenticated ciphertext and parameters persist. Keys stay in memory; reload, page hiding, idle timeout and cross-tab changes lock the private view. Backups are encrypted and restored locally. No account, cross-device sync or passphrase recovery exists. Notes, choices, searches and selected perspectives are never attached to inference requests; only the explicitly submitted public question is sent.

The hosted service is a limited global Cloudflare/Public AI demo, not exclusively Swiss hosting or permanently free infrastructure. It has 200 shared lifetime reservations, 80/day, six per five minutes per network and one active question. Aggregate counts survive deployments; the app does not persist question text or IPs. Temporary credentials and provider credit can expire. No card, top-up or paid plan was added. Provider/platform processing and retention are outside the app's control. The owner's Mac is not public infrastructure.

Reproduce with docs/JUDGE_RUN.md. Source and failure evidence: docs/evidence/onprem-2026-10-15, property-tax-2026-10-15 and gate0-release-2026-10-14. Implementation: PR #4, https://github.com/Devpen787/stimmfolgen/pull/4. App: https://stimmfolgen-demo.stimmfolgen-hosted.workers.dev/. Original sources: Federal Chancellery booklets and dated Federal Social Insurance Office / Federal Department of Finance pages linked in the app. Exact revisions/hashes are in the evidence manifests.

Remaining boundaries: second-case human publication review, unseen-ballot and multilingual evaluation, independent semantic/security review, comprehension study and longer-term service funding. This continuation leaves video unchanged. Main integration and submission remain separate. Apache-2.0 code is distinct from third-party source rights; eight historical page captures need a redistribution basis before the organiser warranty. THIRD_PARTY_NOTICES.md records this and PyMuPDF AGPL obligations. No original PDF or model weights are committed. Codex assisted implementation and evaluation; Apertus performed actual model inference.