



INNOVIUS · SEMANTIC ENCRYPTION

ShinrAI

The AI reads your text — never your identity. ShinrAI replaces every person, place and organization with a statistically faithful stand-in, matched in culture, language, gender and context — and maps the answer back on your side. **You use the AI. The AI does not use you.**

Multi-dimensional replacement

Bias-free by design

14 language tracks

You hold the key

CPU or GPU

Apache 2.0 open weights

2.8×

detection quality vs a tuned rule engine on real German medical letters (F1 71.8 vs 25.6)

94.4

best macro F1 across six languages (range 89.2–94.4, strict span matching)

<1 ms

static screening path — emails, phones, IBANs, IDs — before any model runs

26 ms

p50 neural detection, short text, laptop-class CPU (fp32, measured)

9.4 ms

GPU inference per paragraph — measured, 4.3× CPU; production end-to-end p95 41.6 ms

WHY REPLACEMENT BEATS MASKING

ORIGINAL	DETECTED REALM	MASK	PRESERVE	WHY IT MATTERS
Robert	PERSON · DE · common · masc	[FIRSTNAME]	Paul	Never “Mohammed” — flipping cultural origin changes AI answers.
Bad Neustadt	CITY · DE · small	[CITY]	Biberach	Never “München” — a small town stays a small town.
Mailand (in German text)	CITY · IT · major	[CITY]	Turin	The entity is Italian even in German text — exonym-aware.
Siemens	ORG · DE · international	[COMPANY]	Bosch	Scale is meaning: international stays international.

Three replacement modes per entity type and deployment: **preserve** (default), **mask**, and **neutralize** — deliberate bias-stripping for HR screening or credit decisions. Gender is never flipped; rarity tier is never dropped; six validation gates guarantee lossless restoration.

FIVE DIMENSIONS PER REPLACEMENT

Type & tier

person, city, street, org — rarity preserved

Cultural origin

50 classes, independent of text language

Gender expression

flips forbidden; fallback → neutral

Name part & register

a nickname stays a nickname

Validation

6 gates · injective map · restore guaranteed

LANGUAGES & CULTURES

IN PRODUCTION TODAY

German · English · Japanese protect live traffic; **Italian, French, Spanish and Polish** are in v1.2 — release imminent. Automatic per-language model routing.

V1.3 “WORLDWARD” — IN TRAINING

Russian, Ukrainian, Turkish, Portuguese (PT + BR), Korean, Arabic — trained on a ~13M-record mix at the Jülich Supercomputing Centre; all current languages retrain alongside and improve drastically. Up next: **medical** and **aerospace** specialists, already in training.

RESULTS ON REAL MEDICAL LETTERS



Golden ePA corpus: real German doctor letters, strict span matching. On six-language generated suites, v1.2 scores **89.2–94.4 macro F1** — at or above the ~60–85% reported by cloud PII services and the ~85–95% self-reported by open RoBERTa-class models on their own corpora. **v1.3** (in training) and **v1.4** (in preparation) are pinned at a **95-point release gate** per category. Attribute accuracy (frozen suite): origin 97.7%, gender 98.8%, name part 100%.

MEASURED PERFORMANCE

PATH	MEASURED
Static screening (structured PII)	<1 ms / stage
Replacement draw	~0.4 ms
Neural, short text (CPU, fp32)	26.2 ms p50 · 380 req/s @ c=4
Neural, paragraph (CPU, fp32)	101.1 ms p50 · 104 req/s @ c=4
Neural, paragraph (GPU, P40)	9.4 ms inference · 28.3 ms service p50 (measured)
Neural, paragraph (Mac mini M4, est.)	~105 ms p50
Production end-to-end, paragraph (live)	37.6 ms p50 · 41.6 ms p95
Modern GPU (L40S/A100), paragraph (est.)	~2–4 ms inference · ~13 ms service
Batch API (GPU service)	64 docs × 200k chars / call

Laptop-class CPU, sustained load; any server error fails the run. Overload degrades to the static path — never fail-open. ~9M paragraph requests/day per CPU node, scaling linearly. Zero 5xx across all cells and live failure drills.

DEPLOYMENT OPTIONS

BACKEND · KUBERNETES

ShinrAI Platform

Drop-in OpenAI-compatible endpoint between your apps and any LLM. Encrypt outbound, restore mid-stream inbound. Helm charts, CPU-only or GPU-served, air-gap ready. The replacement map never exists server-side — the client holds the key.

DESKTOP · TRAY APP

ShinrAI Connector

Routes local AI tools — Claude Code, Codex, anything on the vendor SDKs — through on-device protection. Monitor mode first, Protect when ready; per-request transcript, loopback-only, notarized. macOS today, Windows/Linux in preparation.

PRODUCTS

ChtSafe & Secure AI Suite

ChtSafe brings ShinrAI-protected access to 50+ models to individuals. The Secure AI Suite deploys the governed enterprise stack — ShinrAI is its encryption layer.

WHAT DO I NEED? — RAM + FAST CPU, OR A GPU. BOTH WORK.

SETUP	RAM	CPU	GPU	FITS
Static only	256 MB–1 GB	1–2 cores	–	structured PII, absolute minimum footprint
+ Neural compact (INT8)	2–4 GB	2–4 cores	optional	full semantic detection on commodity servers & laptops
+ Neural full (fp32)	4 GB (~1.5 GB/model)	2–4 cores	optional	reference quality, still CPU-friendly
GPU serving	6–10 GB system	2–4 cores	1 GPU · ~4.5 GB VRAM	tightest tail latencies; even 2016-class GPUs qualify

SECURITY POSTURE

- ✓ No server-side mapping store — restoration keys stay with you
- ✓ Content-free logs: counts and types only, no plaintext retention
- ✓ Config secrets wiped after boot into a memory-only store
- ✓ Session-consistent, injective replacement — restore is guaranteed

OPEN SOURCE & RESEARCH

- ✓ Open weights (Apache 2.0) on huggingface.co/innovius
- ✓ Developed with the EECC Research Lab
- ✓ Trained on JURECA at Jülich (WestAI) — scaling onto JUPITER
- ✓ Synthetic training data, license-tracked per record — no proprietary APIs

ENTERPRISE ADD-ON: ONION ROUTING

ShinrAI protects **what** is in a request; onion routing hides **who** sends it. Multi-hop routing on **llmproxy**, our low-level C++ proxy engine — no attribution via provider billing, IP or country of origin. The highest security level our enterprise customers run. Optional; we demo it and set it up as a paid service.

WE ARE LOOKING FOR PARTNERS

Cloud & hosting: run the full Innovius stack — ShinrAI, Secure AI Suite, ChtSafe — in your cloud, region or sovereign environment. **Domain partners:** medical, governance and other regulated fields — build sector solutions and shape the specialist models. Contact: info@innovius.ai