

# SARA Engine: Phase 1

PyPI Library Architecture & Release Strategy

---

# Vision & Roadmap

Transforming SARA from a personal Spiking Neural Network experiment into an accessible, high-performance edge AI standard.



# Core Objectives

## Seamless Installation

Ensure any developer can integrate SARA into their environment effortlessly. By structuring the repository correctly, we enable standard **pip install sara-engine** commands, eliminating complex local setups and bridging the gap between advanced SNN tech and everyday Python development.

## Developer Experience (DX)

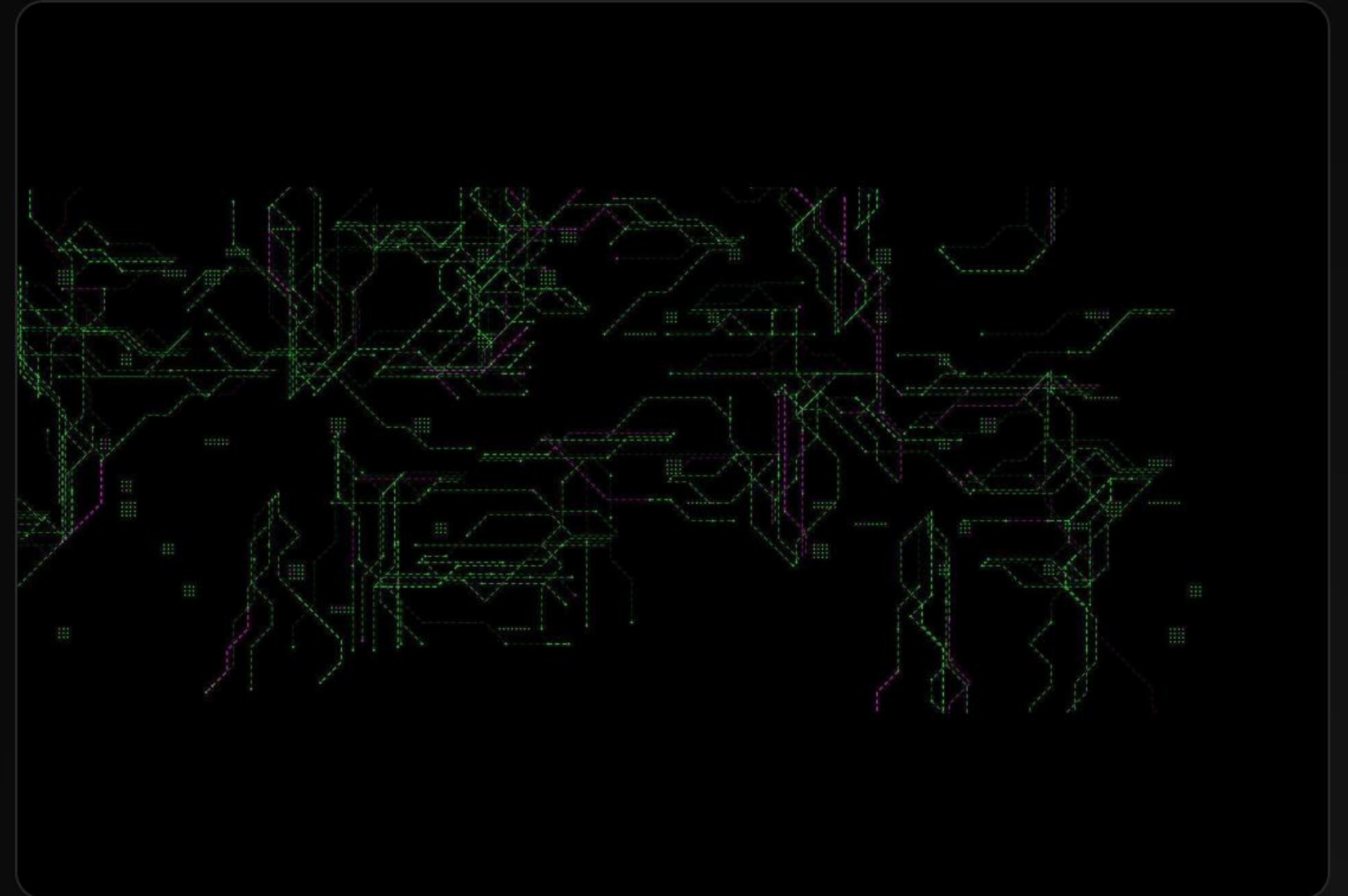
Provide an intuitive API where importing the **SaraInference** class allows users to spin up a local AI in just three lines of code. Abstract away the complexities of SDR mapping and Hippocampus searching, focusing strictly on inputs, outputs, and customizable parameters like `top_k` and `temperature`.



# Target Architecture

To succeed as a public package, we must enforce a strict separation of concerns:

- **The Core Package (`sara_engine/`):** Contains only the necessary code for inference. This ensures the installed footprint remains incredibly light.
- **The Operations Hub (`scripts/`):** Keeps distillation, DB management, and training scripts separate. End-users don't need these unless they are actively fine-tuning.
- **Data Isolation (`data/`):** Ensures SQLite knowledge bases and JSONL conversational logic are cleanly decoupled, preventing data corruption.





# The Value Proposition

1500+

TOKENS / SECOND

## Unmatched Edge Speed

By bypassing heavy floating-point matrix multiplications (GPUs) and relying on  $O(1)$  dictionary lookups of Spiking Neural Network patterns, SARA achieves speeds that traditional LLMs cannot replicate locally. This is the primary selling point of the **sara-engine** PyPI package: instantaneous, on-device reasoning.



# Key Deliverables for v1.0.0



## Packaging Configuration

Implement a robust `setup.py` or `pyproject.toml` detailing dependencies (like `transformers` and `msgpack`), versioning, and entry points for smooth PyPI publication.



## Comprehensive Documentation

A pristine `README.md` containing quick-start guides, architecture overviews, and examples of how to format `chat_data.jsonl` for custom instruction tuning.



## CLI Tooling

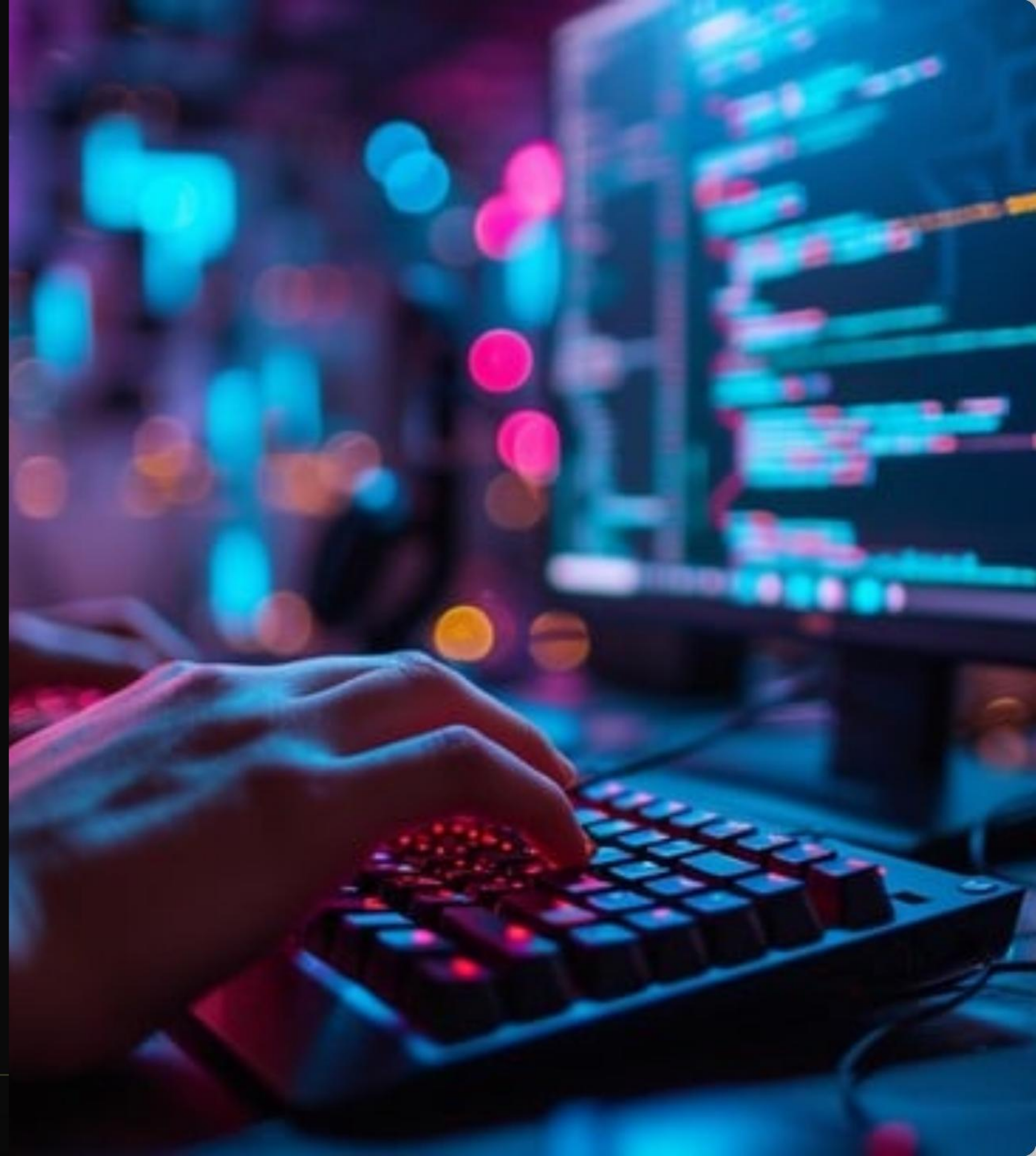
Introduce terminal commands like `sara-train chat_data.jsonl`, allowing users to effortlessly overwrite chat personalities and distill new logic without writing custom Python scripts.



# The Developer Experience

The ultimate goal of Phase 1 is to make integration feel magical. Developers building Web Apps, Discord Bots, or local automations should not need a PhD in Neural Networks.

With **`from sara_engine.inference import SaraiInference`**, the entire power of the Hippocampus engine is encapsulated. They pass a prompt, set their desired randomness, and receive instantaneous text—all running securely and locally on their own hardware.





# Beyond Phase 1



## Tool Calling

Implementing logic for SARA to emit specific spikes (e.g., ) to trigger external Python functions, bridging the gap between language generation and exact computation.



## Rust Hybrid

Re-integrating the `'sara_rust_core'` to handle LIF (Leaky Integrate-and-Fire) membrane potentials, expanding context windows for long-form comprehension.



## Multimodal

Encoding audio frequencies and visual edge-detection into sparse distributed representations (SDRs) to create cross-modal synaptic connections.



# Execution Timeline





# Ready to Deploy?

Let's finalize the PyPI package configuration.

```
$ pip install sara-engine
```



# Image Sources



[https://static.vecteezy.com/system/resources/previews/067/827/002/non\\_2x/glowing-tech-grid-with-green-and-purple-lines-abstract-digital-circuit-background-for-hacking-cybersecurity-or-ai-themes-high-tech-layout-with-neon-effect-futuristic-hack-bg-illustration-vector.jpg](https://static.vecteezy.com/system/resources/previews/067/827/002/non_2x/glowing-tech-grid-with-green-and-purple-lines-abstract-digital-circuit-background-for-hacking-cybersecurity-or-ai-themes-high-tech-layout-with-neon-effect-futuristic-hack-bg-illustration-vector.jpg)

Source: [www.vecteezy.com](https://www.vecteezy.com)

---



[https://images.stockcake.com/public/f/6/4/f643005b-bd96-4b64-9e29-11329b60914b\\_large/coding-at-night-stockcake.jpg](https://images.stockcake.com/public/f/6/4/f643005b-bd96-4b64-9e29-11329b60914b_large/coding-at-night-stockcake.jpg)

Source: [stockcake.com](https://stockcake.com)