

Kani (力二)

Andrew Zhu*, Liam Dugan*,
Alyssa Hwang, Chris Callison-Burch
{andrz[†], ldugan, ahwang16, ccb}@seas.upenn.edu
[†]correspondence ^{*}equal contribution

A Lightweight and Highly Hackable Framework for Building Language Model Applications

The diagram illustrates the Kani framework architecture. On the left, a user asks "What's the weather like in Tokyo?" and receives a response: "It's 22 degrees Celsius and sunny in Tokyo right now." The central part shows the Kani interface with a "Chat History" section containing a message about Tokyo weather, a "Function Context" section with Python code for `get_weather` and `fahren_to_celsius`, and an "Engine" section with icons for OpenAI, Google, and others. On the right, the Python code for the Kani framework is shown, including the `Unit` enum, the `OpenAIEngine` class, the `MyKani` class with an `@ai_function` decorator, and the `ai = MyKani` initialization.

CORE FEATURES

- CHAT HISTORY MANAGEMENT:** Estimate and truncate tokens to ensure prompts never exceed the context window
- ROBUST FUNCTION CALLING:** Control flow and validate LLM-generated calls of developer-written Python functions
- MODEL-AGNOSTIC:** Common Engine and ChatMessage interfaces afford easy interoperability among many LLMs

USE CASES

- RETRIEVAL-AUGMENTED GENERATION:** Write functions to access and use knowledge bases for enhanced responses
- AGENT SPAWNING:** Spawn a new Kani instance within a Kani to collaborate, specialize, or delegate tasks autonomously
- API HOSTING:** Host and productionize Kani instances smoothly through our convenient asynchronous design

AND MORE

- NO PROMPT HACKS:** Interact with LLMs without hidden adjustments that affect the output without your knowledge
- OVERRIDABLE METHODS:** Hook into every part of the library and customize it to your exact specifications
- MULTIMODAL SUPPORT:** Official extensions for vision-language models (LLaVA, GPT-4V), with more on the way
- MINIMALIST IMPLEMENTATION:** Lightweight, unopinionated framework for simpler, shorter, less bloated code
- MAXIMALIST DOCUMENTATION:** Ever-growing, example-driven documentation covering a wide range of applications
- REPRODUCIBILITY:** Conduct research or product experiments with greater control and consistency over model behavior
- OPEN TO CONTRIBUTIONS:** Submit a pull request, issue, or comment on GitHub—we are excited to see what you make!

pip install kani

github.com/zhudotexe/kani

kani.readthedocs.io

discord.gg/eTepTNDxYT

Penn NLP @ University of Pennsylvania

Paper



GitHub

