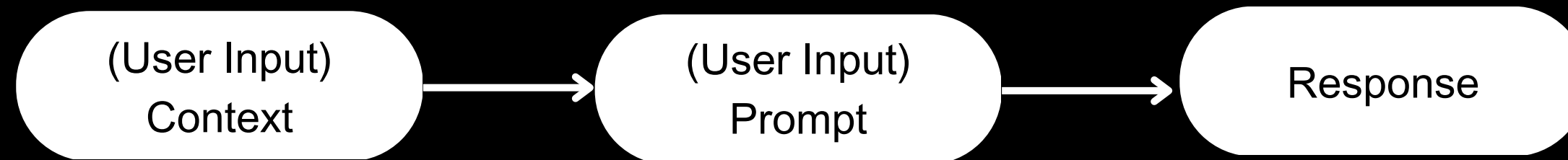# >_*Erica*

OS-Level agent that Listens, Reflects, and Responds to your every action.
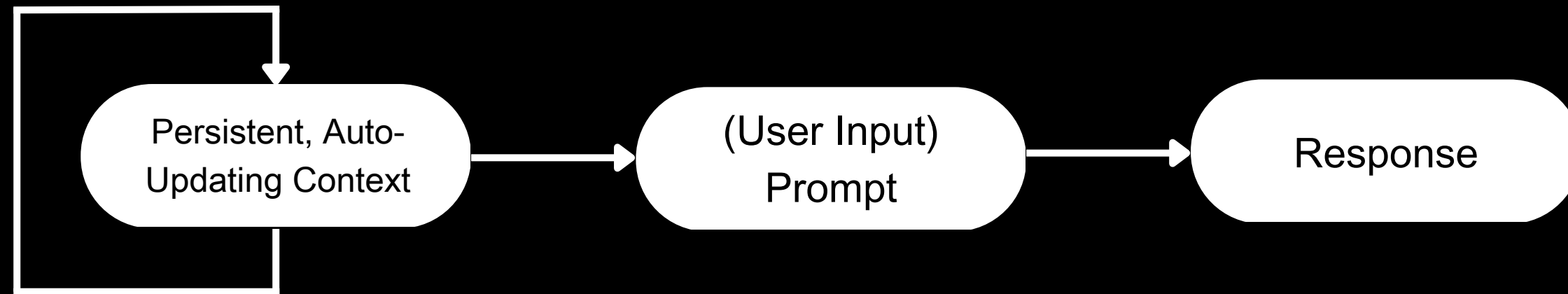
Erica has...

- Eyes
  - Erica captures all screen activity, parsing with Tesseract OCR
- Memory
  - Erica remembers previous conversations and user actions using embeddings
- Mind
  - Erica is capable of complex reflection upon self and user actions using our Reflexion framework
- Voice
  - Erica speaks fluidly using ElevenLabs Synthesis API
- Ears
  - Erica listens intently for commands using Whisper and activation cues
- Hands
  - Erica can autonomously retrieve documents and webpages with ease.

Problem:

(User Input) Context → (User Input) Prompt → Response

- User input of files/text into context window to prompt LLMs is unnatural and frictional.

- AI Assistants operating in walled gardens of SaaS offer poor general use.

- All human sensory input of digital interfaces is hypothetically accessible to AI, but currently radically underutilyzed
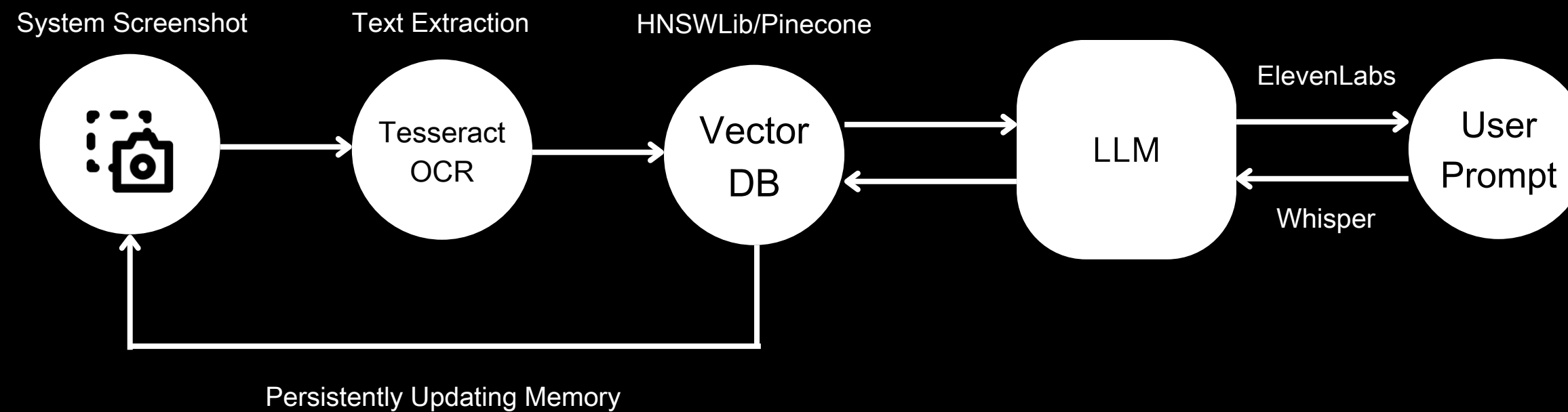
Solution:

- Erika is a persistent OS-Level agent who knows what you're talking about, all the time. No need for feeding in context.

- Erika persistently ingests user inputs and displayed data to enable on-demand QA about anything the user interacts with, ever.

What can Erika do?

- **"What was the summary of the paper I was just looking at a few minutes ago?"**

- **"How does the code I'm looking at now work?"**

- **"Based on my action history over the past day, what are some optimizations I can make to boost my productivity?"**

- Querying, anywhere. Desktop ready, web-intelligent, and optimized for generalization
- Long-Term (cross-device) Recall + QA on tasks via vectorDB
- Reflective ability on actions + User guiding.
- Seamless Cross-Device Memory

Tech Stack:

System Screenshot → Text Extraction → HNSWLib/Pinecone

Tesseract OCR → Vector DB ⇄ LLM → User Prompt

ElevenLabs / Whisper

Persistently Updating Memory

- User Action Logging: Periodic Screenshots
- OCR (Tesseract) + Feature Extraction
- Long-term memory via embeddings + vectorDB (HNSWLib/Pinecone)
- Querying via LLM 3.5-turbo
- User Interface:
  - Query: CLI or Voice (Whisper)
  - Response: CLI or Speech (ElevenLabs)
- Highly deployable dockerized environment, cross-device compatible via cloud vectorDB

Future Development:
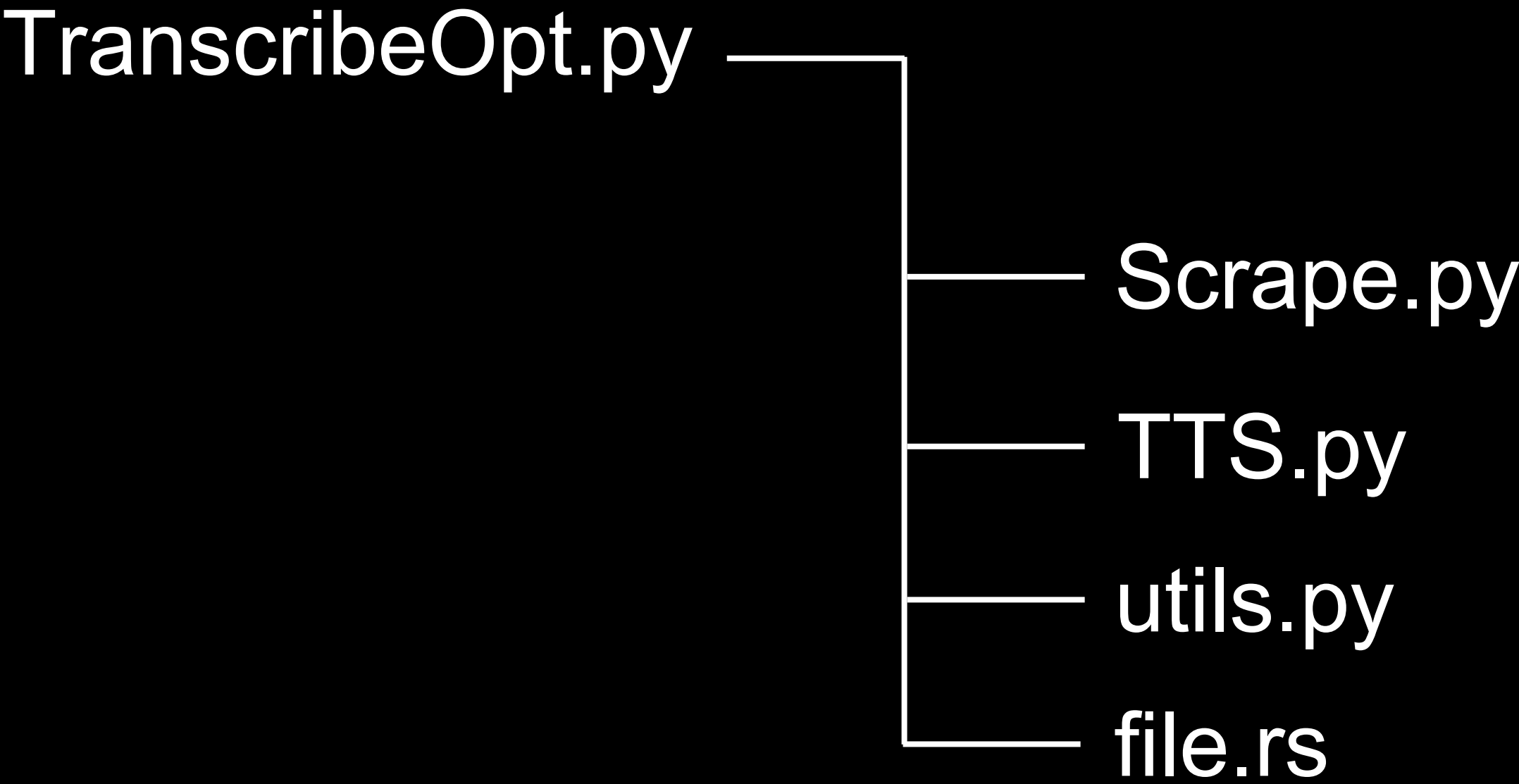
Now: OS-Level Agent

↓

Future: OS-~~Level Agent~~

Future Development:

AI-First Linux Kernel.
Fully User Owned + Open Source Stack:

- Whisper
- LLaMa/GPT4All/StableLM
- Chroma
- OpenTTS

Codebase:

TranscribeOpt.py

Scrape.py

TTS.py

utils.py

file.rs

Team:

Peter Bowman-Davis

Noah Shinn

- Yale Physics '25, Researcher @ MIT
- Prev:
  - ML Engineer, Halborn
  - Technical Cofounder, Politeia Inc. (Acq. '22)

- Northeastern CS '25, Researcher @ MIT + Princeton NLP Group
- First author, "Reflexion: an autonomous agent with dynamic memory and self-reflection"
- Prev:
  - ML Engineer, Marvell Technology