Let’s take back the AI… and integrate it properly to our desktops.

A call to action for the Linux community to do it ourselves and do it right

Michal Kohútek, PhD.

Table of contents

How we got here?

- **1950s-1970s**
  - Turing Test
  - Perceptron
  - ELIZA
  - First AI winter
- **1980s**
  - Expert systems
  - Mycin
  - XCON
  - Second AI Winter
- **1990s**
  - Machine learning
  - Decision trees, SVMs, k-NNs, naive Bayes
  - Deep Blue beats Kasparov (1997)
- **2000s**
  - Internet -> massive datasets.
  - Geofrey Hinton combines neural networks and deep learning (2006)
  - ImageNet (2009)
- **2010s**
  - AlexNet wins ImageNet using CNNs on GPUs (2012)
  - Facial recognition, speech recognition in CE
  - Siri, Alexa, Google Assistant, ~~Mycroft~~
  - AlphaGo beats Lee Sedol (2016)
  - reinforcement learning, GANs, etc.
- **2020s**
  - GPT-3 shows up (2020)
  - ChatGPT and StableDiffusion go viral (2022)
  - Multimodal AI, agents, Open Source models
  - Democratization of AI

AI is everywhere

It's being used in wide variety of applications, from generating silly memes to revolutionizing protein folding.

- Common use-cases:
  - Assistants and chatbots
  - Image generation and computational photography
  - Facial recognition
  - Real-time translations
- But also:
  - Healthcare
  - Climate science
  - Accessibility
  - Physics
  - Logistics
  - Education
  - Finance

The current "state of art" on the desktop

Attempts at integrating AI tech into desktop operating systems have so far missed the mark.

- **Microsoft ~~Cortana~~ Copilot**
- **"Apple" Intelligence**

The current "state of art" on the desktop

Attempts at integrating AI tech into desktop operating systems have so far missed the mark.

- **Microsoft ~~Cortana~~ Copilot** - reportedly competent, but problematic in the privacy area
- **"Apple" Intelligence**

The current "state of art" on the desktop

Attempts at integrating AI tech into desktop operating systems have so far missed the mark.

- **Microsoft ~~Cortana~~ Copilot** - reportedly competent, but problematic in the privacy area
- **"Apple" Intelligence** - lacking features, laugheable implementation, Siri still appearing lobotomized

The current "state of art" on the desktop

Attempts at integrating AI tech into desktop operating systems have so far missed the mark.

- **Microsoft ~~Cortana~~ Copilot** - reportedly competent, but problematic in the privacy area
- **"Apple" Intelligence** - lacking features, laugheable implementation, Siri still appearing lobotomized

- Both locked to specific hardware by a specific vendor

The current "state of art" on the desktop

Attempts at integrating AI tech into desktop operating systems have so far missed the mark.

- **Microsoft ~~Cortana~~ Copilot** - reportedly competent, but problematic in the privacy area
- **"Apple" Intelligence** - lacking features, laugheable implementation, Siri still appearing lobotomized

- Both locked to specific hardware by a specific vendor

A Linux desktop assistant done right should be…

My proposal for an AI assistant on the Linux desktop

- **Open Source**
- **Modular and configurable**
- **Local where possible**
- **Integrated to the desktop**
- **Know its user, but keep their secrets**

A Linux desktop assistant done right should be…

My proposal for an AI assistant on the Linux desktop

- **Open Source** - no vendor lock-in, no data collection by a third party
- **Modular and configurable**
- **Local where possible**
- **Integrated to the desktop**
- **Know its user, but keep their secrets**

A Linux desktop assistant done right should be...

My proposal for an AI assistant on the Linux desktop

- **Open Source** - no vendor lock-in, no data collection by a third party
- **Modular and configurable** - no monolithic architecture, no single point of failure
- **Local where possible**
- **Integrated to the desktop**
- **Know its user, but keep their secrets**

A Linux desktop assistant done right should be...

My proposal for an AI assistant on the Linux desktop

- **Open Source** - no vendor lock-in, no data collection by a third party
- **Modular and configurable** - no monolithic architecture, no single point of failure
- **Local where possible** - no need to send data to the cloud, no need for an internet connection
- **Integrated to the desktop**
- **Know its user, but keep their secrets**

A Linux desktop assistant done right should be...

My proposal for an AI assistant on the Linux desktop

- **Open Source** - no vendor lock-in, no data collection by a third party
- **Modular and configurable** - no monolithic architecture, no single point of failure
- **Local where possible** - no need to send data to the cloud, no need for an internet connection
- **Integrated to the desktop** - some use-cases require being able to interact with the desktop
- **Know its user, but keep their secrets**

A Linux desktop assistant done right should be...

My proposal for an AI assistant on the Linux desktop

- **Open Source** - no vendor lock-in, no data collection by a third party
- **Modular and configurable** - no monolithic architecture, no single point of failure
- **Local where possible** - no need to send data to the cloud, no need for an internet connection
- **Integrated to the desktop** - some use-cases require being able to interact with the desktop
- **Know its user, but keep their secrets** - learn from and about their user, in order to be useful, but keep data strictly on device

Features and capabilities

Inspired by others, with a twist.

- **Natural language understanding**
- **Multimodal**
- **Task automation**
- **Integration with existing tools**

Features and capabilities

Inspired by others, with a twist.

- **Natural language understanding** - understand and process natural language commands
- **Multimodal**
- **Task automation**
- **Integration with existing tools**

Features and capabilities

Inspired by others, with a twist.

- **Natural language understanding** - understand and process natural language commands
- **Multimodal** - understand and process text, images, audio, video
- **Task automation**
- **Integration with existing tools**

Features and capabilities

Inspired by others, with a twist.

- **Natural language understanding** - understand and process natural language commands
- **Multimodal** - understand and process text, images, audio, video
- **Task automation** - automate repetitive tasks, like file management, system maintenance, etc.
- **Integration with existing tools**

Features and capabilities

Inspired by others, with a twist.

- **Natural language understanding** - understand and process natural language commands
- **Multimodal** - understand and process text, images, audio, video
- **Task automation** - automate repetitive tasks, like file management, system maintenance, etc.
- **Integration with existing tools** - integrate with existing tools and applications, like email, calendar, text editor, etc.

Low hanging fruit use-cases

What we could implement with current tech without too much effort

- **Calendar management**
- **Web search**
- **IM integration**
- **Context-aware screenshot OCR**
- **Translation**
- **LLM integration**
- **Work context snapshot**

Low hanging fruit use-cases

What we could implement with current tech without too much effort

- **Calendar management** - My proof of concept can already add calendar events to my personal Nextcloud calendar.
- **Web search**

- **IM integration**
- **Context-aware screenshot OCR**

- **Translation**

- **LLM integration**

- **Work context snapshot**

Low hanging fruit use-cases

What we could implement with current tech without too much effort

- **Calendar management** - My proof of concept can already add calendar events to my personal Nextcloud calendar.
- **Web search** - Simple web search is easy enough to implement. A proper deep search equivalent is within our reach too.
- **IM integration**
- **Context-aware screenshot OCR**

- **Translation**

- **LLM integration**

- **Work context snapshot**

Low hanging fruit use-cases

What we could implement with current tech without too much effort

- **Calendar management** - My proof of concept can already add calendar events to my personal Nextcloud calendar.
- **Web search** - Simple web search is easy enough to implement. A proper deep search equivalent is within our reach too.
- **IM integration** - Integrate with existing IM services, like Matrix, Telegram, Signal, etc.
- **Context-aware screenshot OCR**

- **Translation**

- **LLM integration**

- **Work context snapshot**

Low hanging fruit use-cases

What we could implement with current tech without too much effort

- **Calendar management** - My proof of concept can already add calendar events to my personal Nextcloud calendar.
- **Web search** - Simple web search is easy enough to implement. A proper deep search equivalent is within our reach too.
- **IM integration** - Integrate with existing IM services, like Matrix, Telegram, Signal, etc.
- **Context-aware screenshot OCR** - Take a screenshot of the current screen and use OCR to extract text from it and visual LLM model to provide context to it.
- **Translation**

- **LLM integration**

- **Work context snapshot**

Low hanging fruit use-cases

What we could implement with current tech without too much effort

- **Calendar management** - My proof of concept can already add calendar events to my personal Nextcloud calendar.
- **Web search** - Simple web search is easy enough to implement. A proper deep search equivalent is within our reach too.
- **IM integration** - Integrate with existing IM services, like Matrix, Telegram, Signal, etc.
- **Context-aware screenshot OCR** - Take a screenshot of the current screen and use OCR to extract text from it and visual LLM model to provide context to it.
- **Translation** - VLC already got their realtime translation of played media. We could do the same to any played media with a subtitle overlay.
- **LLM integration**

- **Work context snapshot**

Low hanging fruit use-cases

What we could implement with current tech without too much effort

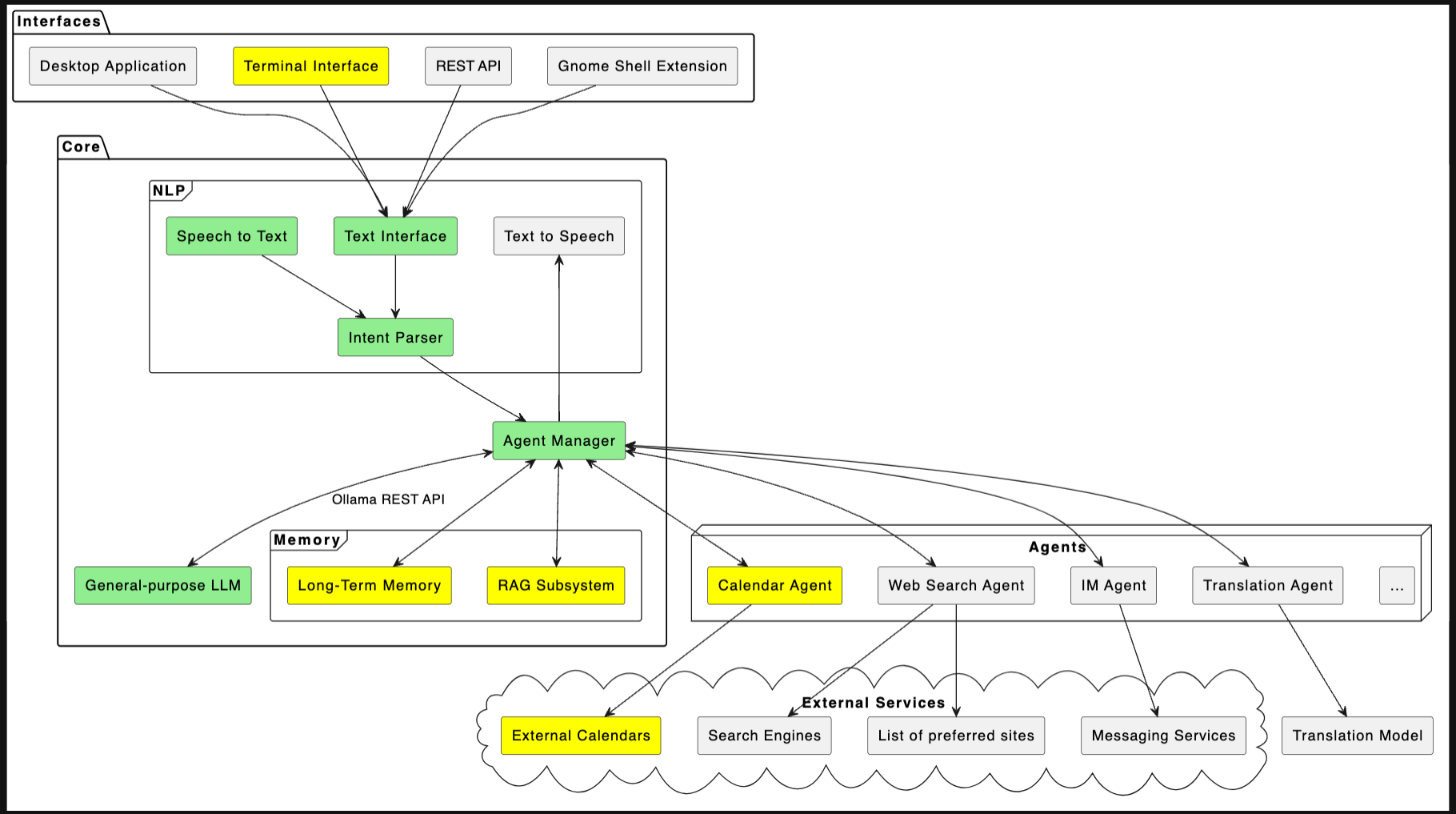
- **Calendar management** - My proof of concept can already add calendar events to my personal Nextcloud calendar.
- **Web search** - Simple web search is easy enough to implement. A proper deep search equivalent is within our reach too.
- **IM integration** - Integrate with existing IM services, like Matrix, Telegram, Signal, etc.
- **Context-aware screenshot OCR** - Take a screenshot of the current screen and use OCR to extract text from it and visual LLM model to provide context to it.
- **Translation** - VLC already got their realtime translation of played media. We could do the same to any played media with a subtitle overlay.
- **LLM integration** - Intent parser can fall back to a general Local LLM model, like Gemma3:27B or a smaller; or as an opt-in could call out to cloud-based LLM .
- **Work context snapshot**

Low hanging fruit use-cases

What we could implement with current tech without too much effort

- **Calendar management** - My proof of concept can already add calendar events to my personal Nextcloud calendar.
- **Web search** - Simple web search is easy enough to implement. A proper deep search equivalent is within our reach too.
- **IM integration** - Integrate with existing IM services, like Matrix, Telegram, Signal, etc.
- **Context-aware screenshot OCR** - Take a screenshot of the current screen and use OCR to extract text from it and visual LLM model to provide context to it.
- **Translation** - VLC already got their realtime translation of played media. We could do the same to any played media with a subtitle overlay.
- **LLM integration** - Intent parser can fall back to a general Local LLM model, like Gemma3:27B or a smaller; or as an opt-in could call out to cloud-based LLM .
- **Work context snapshot** - Take a snapshot of the current work context, including open files, running applications, etc. and use it to provide context-aware assistance.

Assistant architecture



The tech stack

Subject to change at any point

- **Speech to Text**
  - Porcupine
  - Whisper
- **Intent Parsing**
  - Finetuned DistilBert-base-uncased
  - DistilBert-base-uncased as tokenizer
- **General-purpose LLM**
  - Ollama REST API
- **RAG Subsystem**
  - ChromaDB
  - LangChain
  - SentenceTransformers
  - PyMuPDF
  - python-docx
  - ebooklib
  - bs4
- **Long-Term Memory**
  - Embeddings model
  - Generation model
  - Summarization model
  - FAISS or ChromaDB
  - SentenceTransformers
- **Agents**
  - Specific libraries
  - Custom agents
- **Desktop Integration**
  - GJS
  - Python
  - GTK4
  - QT6

Q&A Time

Thank you for your attention!