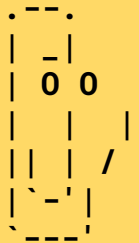
Mastering R2AI

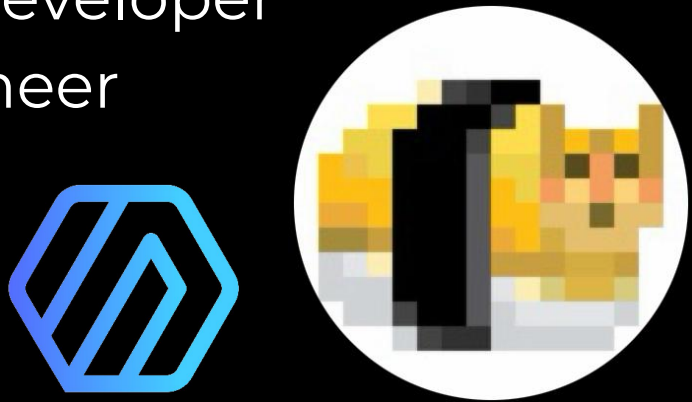
@pancake@infosec.exchange // NN2024



Who Am I?

Sergi Àlvarez aka pancake

- Author and leader of the Radare project
- Free Software enthusiast and developer
- Mobile Security Research Engineer
  at NowSecure



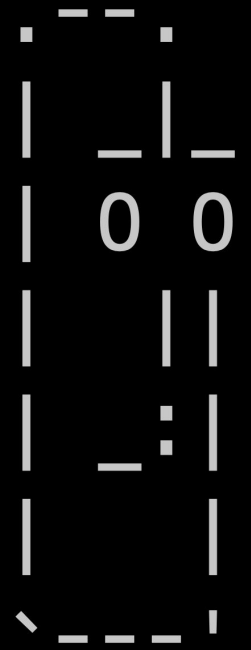
What's this Talk about?

What started as a joke..

making fun memes with r2clippy

ended up gaining conscience

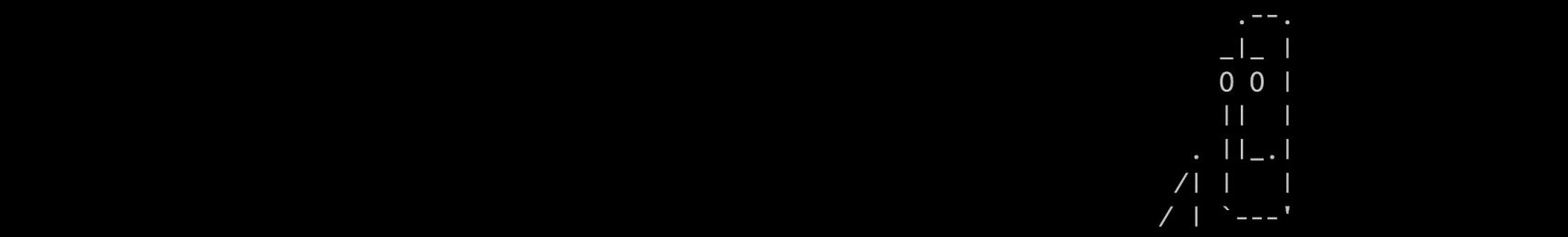
..taking advantage of LLM in R2



Presenting r2clippy

```
[r2ai:0x100003a84]> hello, who are you?

   |\
   | \
.--'  '-------------------------.
| hello, who are you?           |
`-------------------------------'
```



```
  I'm R2clippy, your virtual assistant for Radare2. I'll help you with Radare2-related
  questions and provide technical assistance.

`-------------------------------------------------------------------------------------'
[r2ai:0x100003a84]>
```

Concepts

The Boring Stuff

The Theory

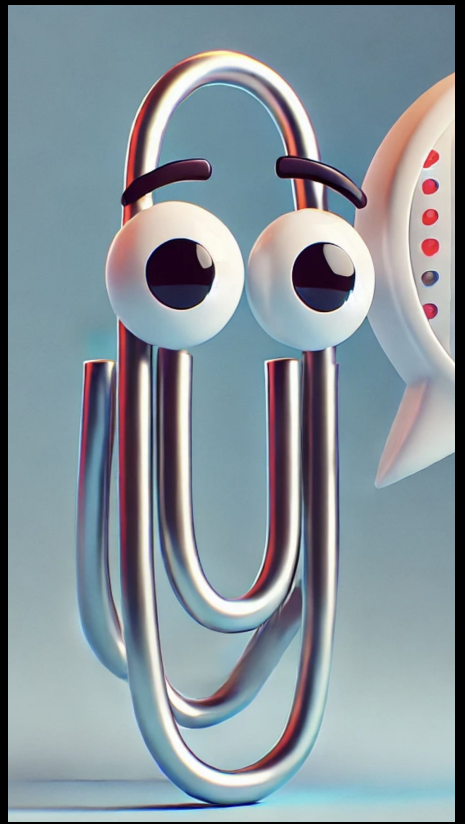
Learn the Internals

Practical Use Cases

Concepts (1/4) - Human Language

5th Generation Programming Language

- **Using BigData and NeuralNetworks**
  - CPU/GPU/NPU/Brain
- **Local Inference Engines**
  - GPT/Llama/MPS/MLC
- **Privative Solutions**
  - OpenAI / Anthropic / Meta / Grok



Commercial/Closed/Cloud Based

- **Anthropic, OpenAI, Llama, Gemini**
  - Some not allowed by EU laws
- **Use API key ..**
  - Faster Response
  - Larger Contexts
  - Biggest Models
  - Burn Water!

Concepts (2/4) - Models

Models infer new tokens

- Using **text**/voice/image/music/video
- Standard: **GGUF** (supported by r2)
- Minimal units as numbers (**Token**)
- Tensors of **Tokens**
  - Aka Multidimensional Vectors
- **HuggingFace** (GitHub for AI)
- Training / FineTuning / Quantification

Concepts (3/4) - Training

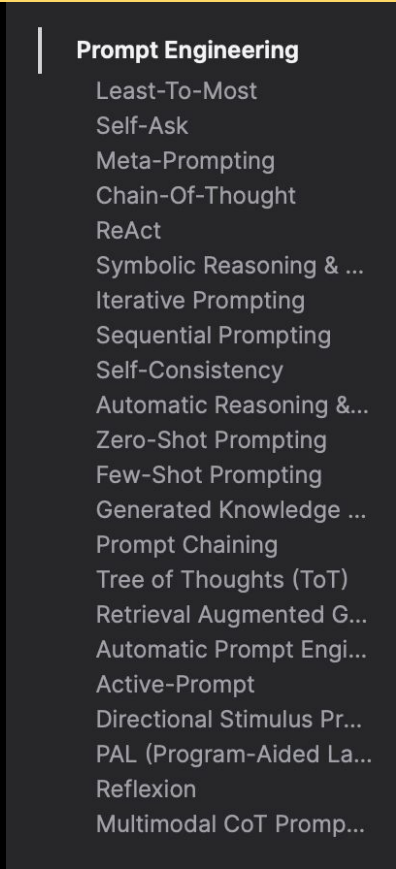
Model training data sources

- Public datasets, **refined** by human reviews and autoreasoning
- **Censored** models (can be hacked)
- Formatted data optimized for:
  - Code autocompletion, **instruction**, conversation, reasoning, text generation, etc.
- **Temperature** / Penalty

Concepts (4/4) - Prompting

Prompt Engineering guides the model to infer better answers for our needs.

- Different ways to express your will
- Embedding the answer in the query
- Context Information in the prompt
- System Prompt / Role Playing



Hallucinations

Biggest issue in AI nowadays

- Make up the answers
- Not reasoning or validating properly
- Prompting properly can help
- Temperature may cause confusion
- The model lacks data or resolution

System Prompt

Before the first message sent to the inference, the system prompt is a sentence that is contained between special tokens that affects the rest of the conversation and defines:

- **`Context`** (medical, scientific, poetic, party, school, ..)
- **`Role`** (“who” is the assistant: teacher, friend, ..)
- **`Properties`** (define the name and other attributes)
- **`Censor`** (what’s forbidden and what’s allowed)
- **`Instructions`** (what to do when something happens)
- **`Output Specs`** (format, language, ..)

Installing r2ai

```
$ r2pm -s r2ai
```

- Standalone Tool
- OpenAPI Server
- R2 Plugin
- Python Module

```
$ r2pm -s r2ai
r2ai-plugin         r2ai plugin for radare2
r2ai-native         run a local language model integrated with radare2
r2ai-bard           Google Bard AI plugin for radare2 (requires bard-cli in $PATH)
decai               r2ai/OpenAI/Anthropic/ollama based decompiler, formerly known as r2ai-decai
r2ai                run a local language model integrated with radare2
r2ai-server         start a language model webserver in local
$
```

Reversing Usecases

- **Assistant / Create scripts**
  - Answer questions about assembly or r2
  - Find new ways to solve problems, hint ideas
- **Decompiler / Transpiler**
  - Combine many decompiles into any lang
  - Not 1:1, but very helpful in some situations
- **Automatic Solver**
  - Replace scholars, self-reasoning
- Find **vulnerabilities** / analyze patches
  - **Explain** code, functions, logic
- **Rename** functions or variables, guess types

Voice Assistant

Speaks and translates many langs

- Local TTS/STT
- Turn based conversation
- GPU/crawling bottleneck

```
[0x00000000]> r2ai -e|grep voice
-e voice.lang=en
-e voice.model=base
-e chat.voice=false
[0x00000000]>
```

Learning From Data

Vector database (chromadb)

- Import documentation in txt/md
- Convert code into markdown
- Mastodon search
- History logs

Decompilation with Decai

Tricky but with impressive results

- R2JS script using curl
  - Talks to r2ai, claude, openai, ..
- Run decai.cmds and requests and improved output

Autonaming Functions

- Take output form **`pdsf`**
  - Suggest better name
- Reverse Recursive Function Listing
  - **`afla`** and **`@@F`**
- Output in **`afn`** format

In short:

- **`decai -n @@F`**

Finding Bugs

- Analyze disasm, decompilation and references to hint/instruct the analyst or auto mode in the points of interest

Reality: Lots of false positives, but enlightening for newcomers, sometimes surprisingly helpful.

NOTE: The Copilot Fail. Mixed feelings

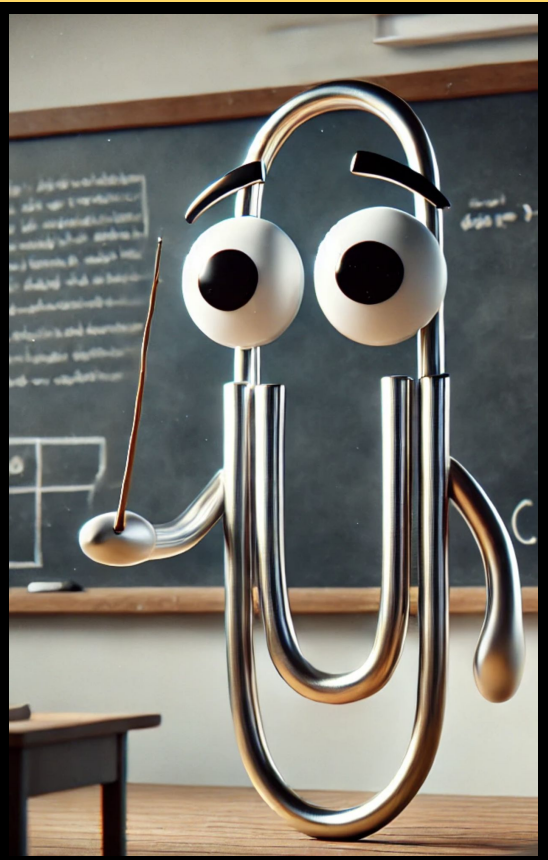
Explaining Code

Are you lost or lazy and you don’t know what’s the purpose of a function?

- Remember: R2 is documented in C
- Feed it with code, pseudo, disasm..

In short:

- `> decai -x`



Scripting

Easily scriptable in r2js or Python:

```
r2.cmd("r2ai")
```

- Explaining pullrequests
- Documenting code
- Writing chapters for the r2book
- Generate Podcast
- Summarize YouTube Videos
- Crawl the Fediverse

Auto Mode

- Replace scholars!
- Function Calling
  - Local models or Chaining
- Create a plan from the query
  - Run r2 commands
  - Analyze the result
  - Perform following steps
  - Find solution

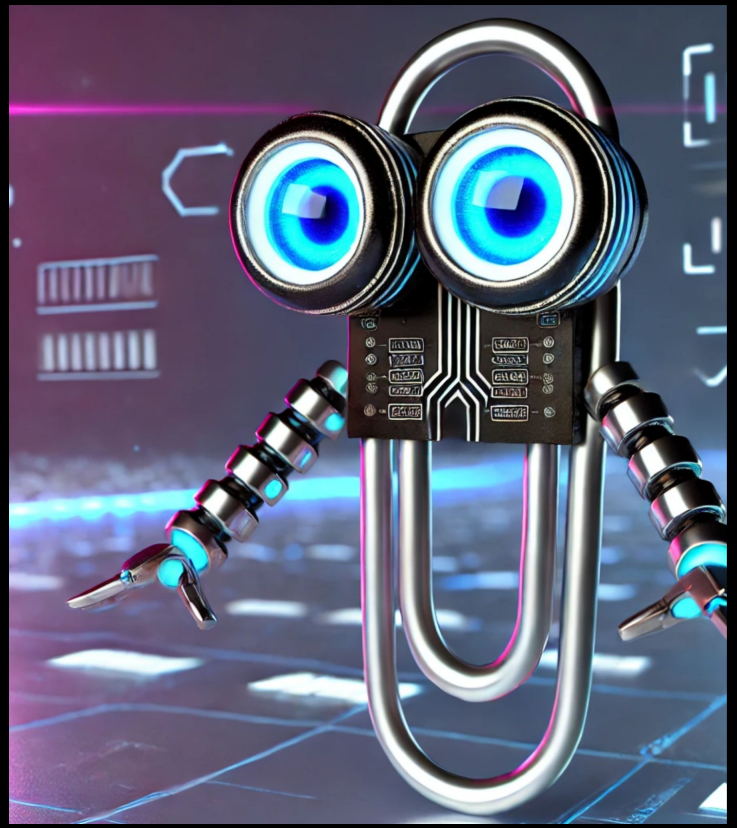
Reasoning

- OpenAI’s O1, added reasoning tags, used to define a plan in the prompt and verify, compare and analyze every single step before answering.
- We can emulate reasoning with multiple prompts and chaining results

Future

AI technologies evolve very fast.

- Better models every week
- Improve auto mode
  - Find and patch vulns
- Finetune our own models
- Fuzzing generators
- Improved decompilation



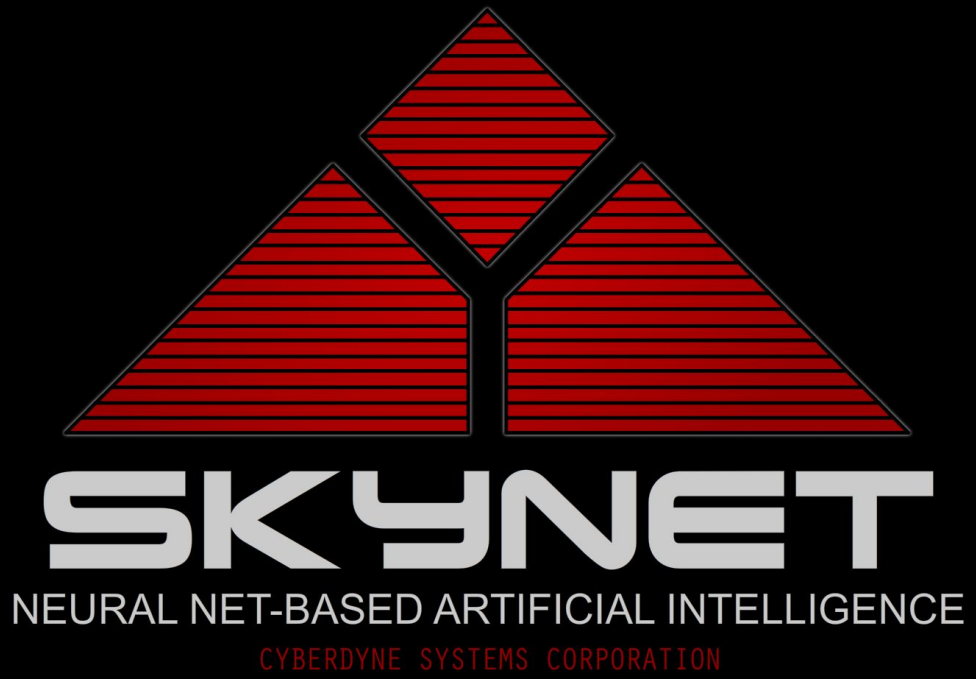
Questions?

To learn more or engage into further discussions:

- Join the Telegram, Matrix or Discord chats
- Fediverse **@radareorg@infosec.exchange**

https://rada.re

Questions?


SKYNET
NEURAL NET-BASED ARTIFICIAL INTELLIGENCE
CYBERDYNE SYSTEMS CORPORATION