

# Workshop: Transformação Digital na Saúde com Inteligência Artificial (IA)

**Andreia Gaudêncio**

*PhD Eng. Biomédica | Suporte & Dados*

**Guimarães - 30 de Abril 2026**

[agaudencio@acnca.pt](mailto:agaudencio@acnca.pt)



# Rede WIFI

- Rede: HGMR
- Password: HGMR2026



# Agenda

## Inteligência Artificial na Saúde

Machine Learning (ML) na Saúde

LLMs & Saúde

## Hands-on

Exemplo de ML

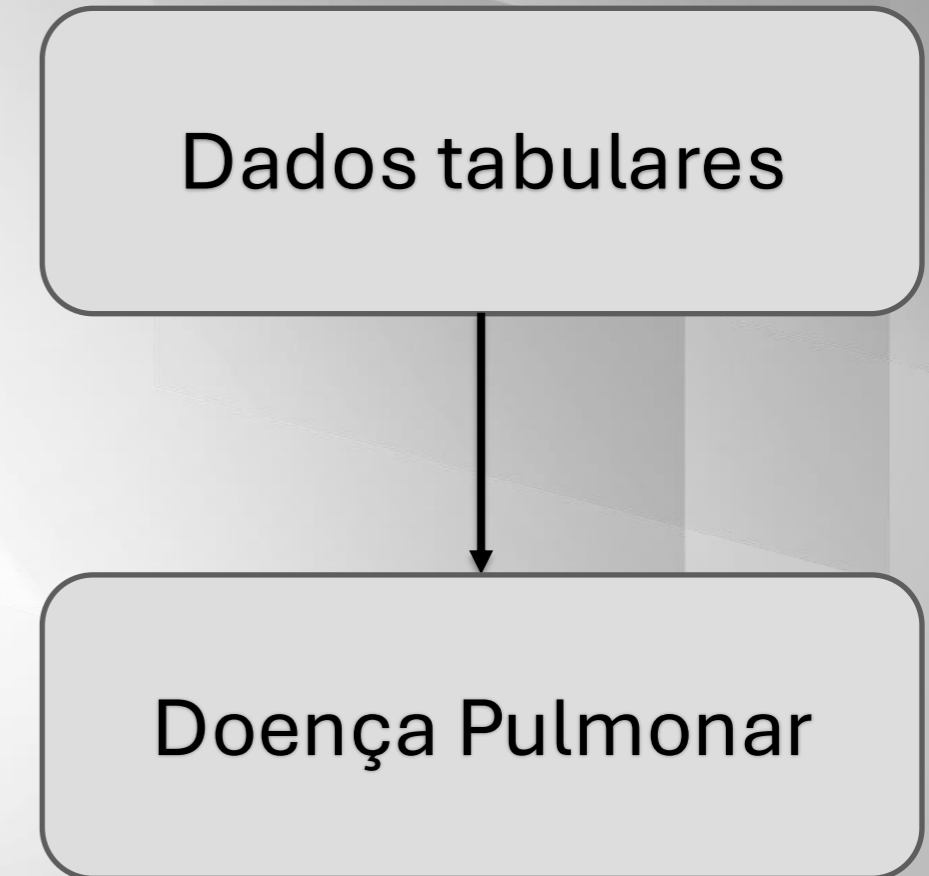
Como desenvolver um Health AI Assistant

Como tratar um dataset para raciocínio clínico?

# Inteligência Artificial (IA) na Saúde

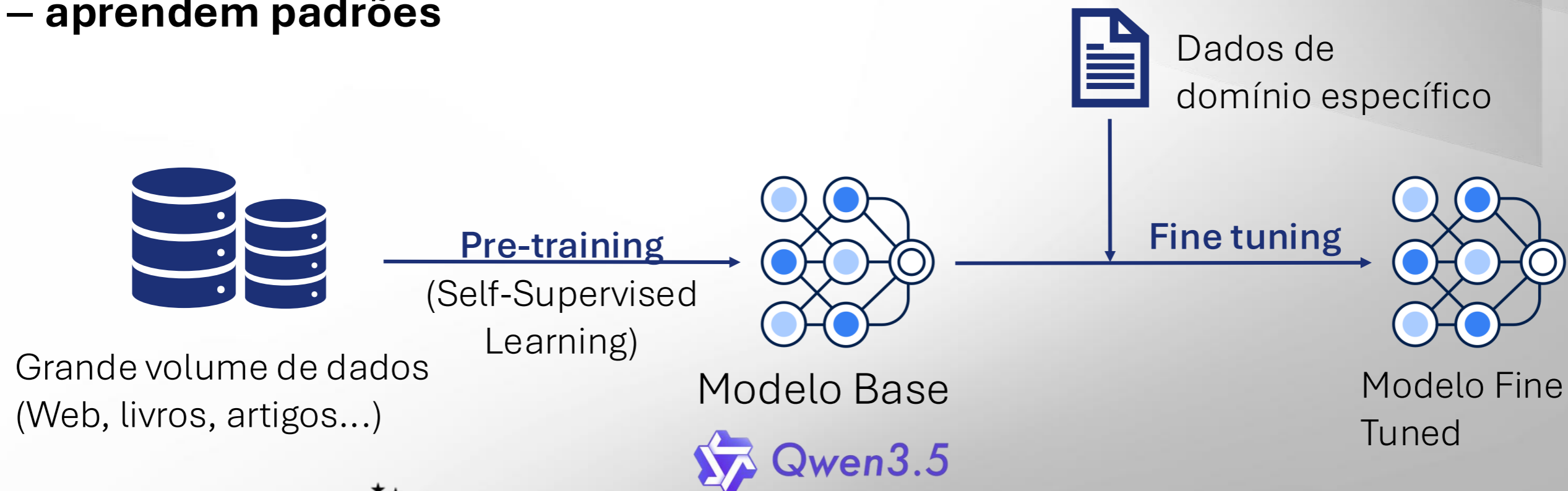
# Machine Learning na Saúde

- **Deteção e predição de doenças**
  - detetar tumores e outras anomalias em imagens médicas
- **Personalização de tratamento**
  - predição do tratamento mais eficaz
- **Desenvolvimento de fármacos**
  - com informação genética, estruturas de proteínas, interações moleculares, prevêm-se interações com diferentes proteínas alvo



# Large Language Models (LLMs)

- Arquitetura com grande número de parâmetros
- Processa e produz linguagem natural
- Modelos matemáticos que aprendem a partir de dados
  - aprendem padrões



# LLMs & saúde

- BioGPT
- ClinicalBert
- MedLM
- SciBert
- ImageBert

Texto	Imagens	Dados estruturados	Video
Notas clínicas, artigos científicos, instruções para pacientes, códigos médicos	Raios-X, TACs, RM	Resultados de laboratório, bioassinais, ECG, EEG, e dados demográficos	Vídeos cirúrgicos, modelos médicos 3D

## Hands-on

<https://indico.acnca.pt/event/9/>

Exemplo de ML – Jupiter  
Notebook

AI Health Assistant – VS  
Code

Inferência - Terminal

Dataset raciocínio clínico

# DEUCALION



Minho Advanced Computing Center (MACC),  
Guimarães





# Exemplo prático de Machine Learning (ML)



# Exemplo de Machine Learning

<https://gitlab.a.incd.pt/cnca/aif-pt/ml-lung-disease-pred>

Workshop: Transformação Digital na Saúde com IA  
Fábrica de IA | Centro Nacional de Computação Avançada (CNCA)



Objetivo: Predição de cancro pulmonar - exemplo prático de Machine Learning  
Dataset: dataset de domínio público obtido via [kaggle](#)

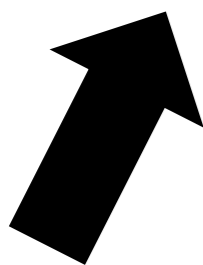
# Primeiros passos

1. Abrir página de Open OnDemand (OOD) do Deucalion

[Deucalion's OOD](#)



2. Copiar os materiais do repositório



OnDemand provides an integrated, single access point for all of your HPC resources.

## Message of the Day

### Changelog

- **07/01/2026:** Deucalion now has a chatbot that goes through our documentation and can answer some basic questions that you might have. To access it please run `module load chat` followed by `deucalion_chat`
- **10/12/2025:** Deucalion has now a private [GitLab](#) server providing a [CI/CD service](#) to all users.

Deucalion

Files ▾

Jobs ▾

Clusters ▾

Interactive Apps ▾

🏠 Home Directory

📁 /projects/F2026AIVLAB00021BSCAIF

/projects/F2026AIVLAB00021BSCAIF



>\_ Open in Terminal ▶ Refresh Refresh + New File New Directory Upload Download Copy/Move Delete

🏠 Home Directory

📁 /projects/F2026AIVLAB00021BSCAIF

⬆ / projects / F2026AIVLAB00021BSCAIF / Change directory Copy path

Show Owner/Mode  Show Dotfiles Filter:

Showing 0 rows - 0 rows selected

<input type="checkbox"/>	Type ▲	Name	Size	Modified at
No data available in table				

powered by



OnDemand version: 4.0.7

Welcome to the Portuguese EuroHPC supercomputer, \*please\* read documentation before use:  
<https://docs.macc.fccn.pt/>

The following scripts are now available for user reference:

1. Home Quota Check: quotahome
2. Lustre (projects) Quota Check: quotaprojects
3. Slurm Billing Information: billing

Run these scripts directly to get up-to-date information on your storage usage and job-related billing.

```
[agaudencio@ln04 F2026AIVLAB00021BSCAIF]$
```

```
[agaudencio@ln03 F2026AIVLAB00021BSCAIF]$ bash ml_repo.sh
```

```
Cloning into 'agaudencio/ml-example'...
```

```
warning: redirecting to https://gitlab.acnca.pt/cnca/aif-pt/ml-lung-disease-pred.git/
```

```
remote: Enumerating objects: 94, done.
```

```
remote: Counting objects: 100% (57/57), done.
```

```
remote: Compressing objects: 100% (57/57), done.
```

```
remote: Total 94 (delta 25), reused 0 (delta 0), pack-reused 37 (from 1)
```

```
Unpacking objects: 100% (94/94), 1.27 MiB | 2.79 MiB/s, done.
```

```
[agaudencio@ln03 F2026AIVLAB00021BSCAIF]$
```

Deucalion Files Jobs Clusters Interactive Apps My Interactive Sessions

Home Directory  
/projects/F2026AIVLAB00021BSCAIF

AI  
Llama.cpp

Desktops  
Desktop

Editors  
VS Code

GUIs  
VAPOR  
VMD


Servers  
Jupyter Notebook  
Jupyter Notebook Pytorch  
Jupyter Notebook Pytorch (GPU)  
RStudio Server  
mlflow

Viz  
TensorBoard

/ projects / F2026AIVLAB00021BSCAIF /

Type ▲

Type
Folder
Folder
Folder
Folder
Folder
Folder
Folder
Folder
Folder
Folder



### Saved Settings

You have no saved settings.

### Interactive Apps

- AI
- Llama.cpp
- Desktops
- Desktop
- Editors
- VS Code
- GUIs
- VAPOR
- VMD
- Servers
- Jupyter Notebook**
- Jupyter Notebook Pytorch
- Jupyter Notebook Pytorch (GPU)
- RStudio Server
- mlflow
- Viz
- TensorBoard

## Jupyter Notebook

This app will launch a Jupyter Notebook server on one or more nodes.

### Number of CPU cores

Enter a value between 1 and 256

### Number of hours

### Partition

Please select a queue from the drop-down

### Account

Save settings

\* The Jupyter Notebook session data for this session can be accessed under the [data root directory](#).

Session was successfully created. ✕

Home / My Interactive Sessions

Saved Settings  
You have no saved settings.

- Interactive Apps
- AI
  - Llama.cpp
  - Desktops
    - Desktop
  - Editors
    - VS Code
  - GUIs
    - VAPOR
    - VMD
  - Servers
    - Jupyter Notebook
    - Jupyter Notebook Pytorch
    - Jupyter Notebook Pytorch (GPU)
    - RStudio Server
    - mlflow
  - Viz
    - TensorBoard

**Jupyter Notebook (1106817)** 1 node | 48 cores | Starting

Created at: 2026-03-31 11:57:34 WEST ✕ Cancel

Time Remaining: 59 minutes

Session ID: 2cfa2ede-265e-418f-86cf-97f4ddeba0a8

Your session is currently starting... Please be patient as this process can take a few minutes.



Session was successfully created. ✕

Home / My Interactive Sessions

**Jupyter Notebook (1106817)** 1 node | 48 cores | Running

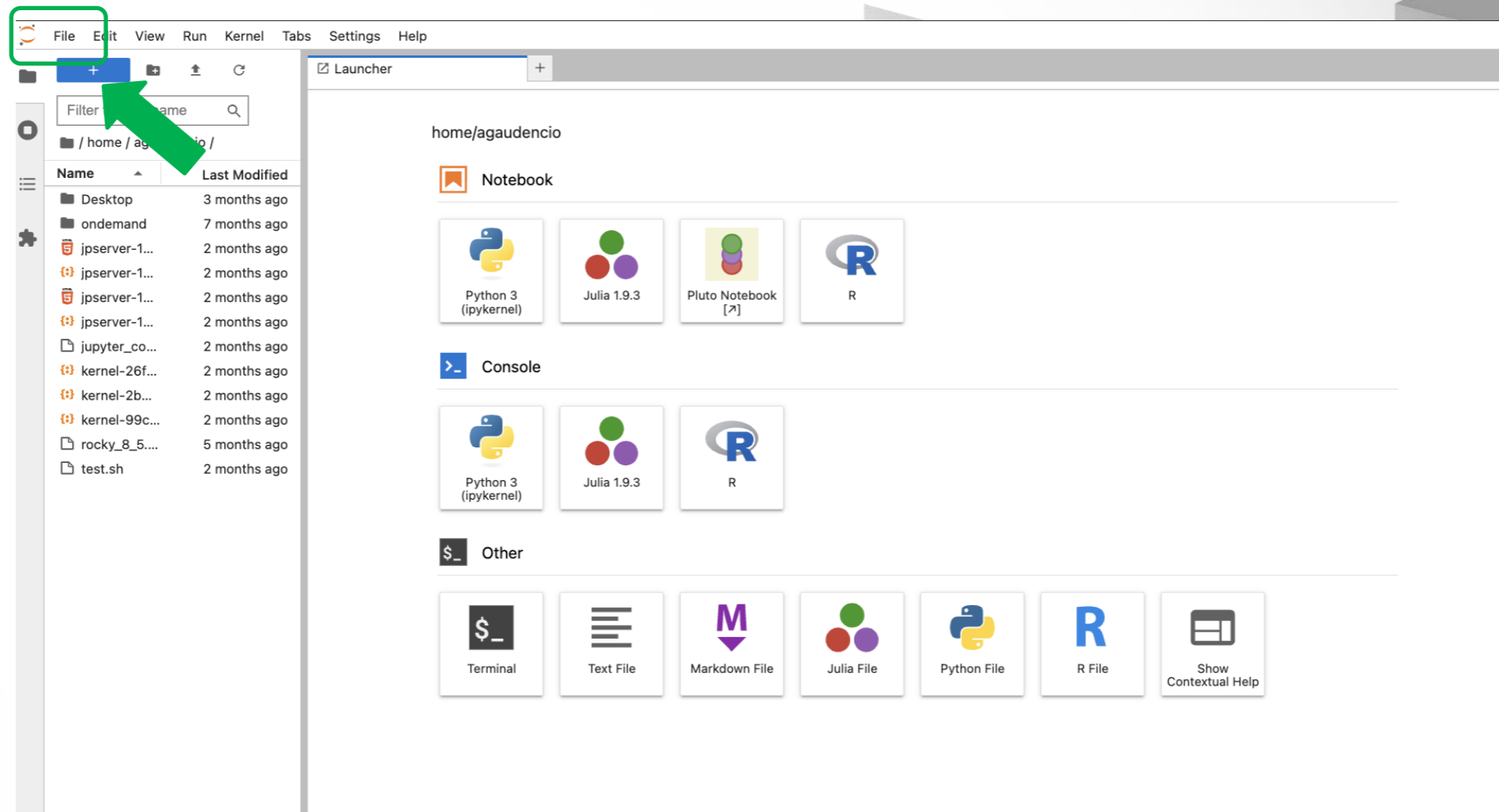
Host: cna0044.deucalion.macc.fccn.pt ✕ Cancel

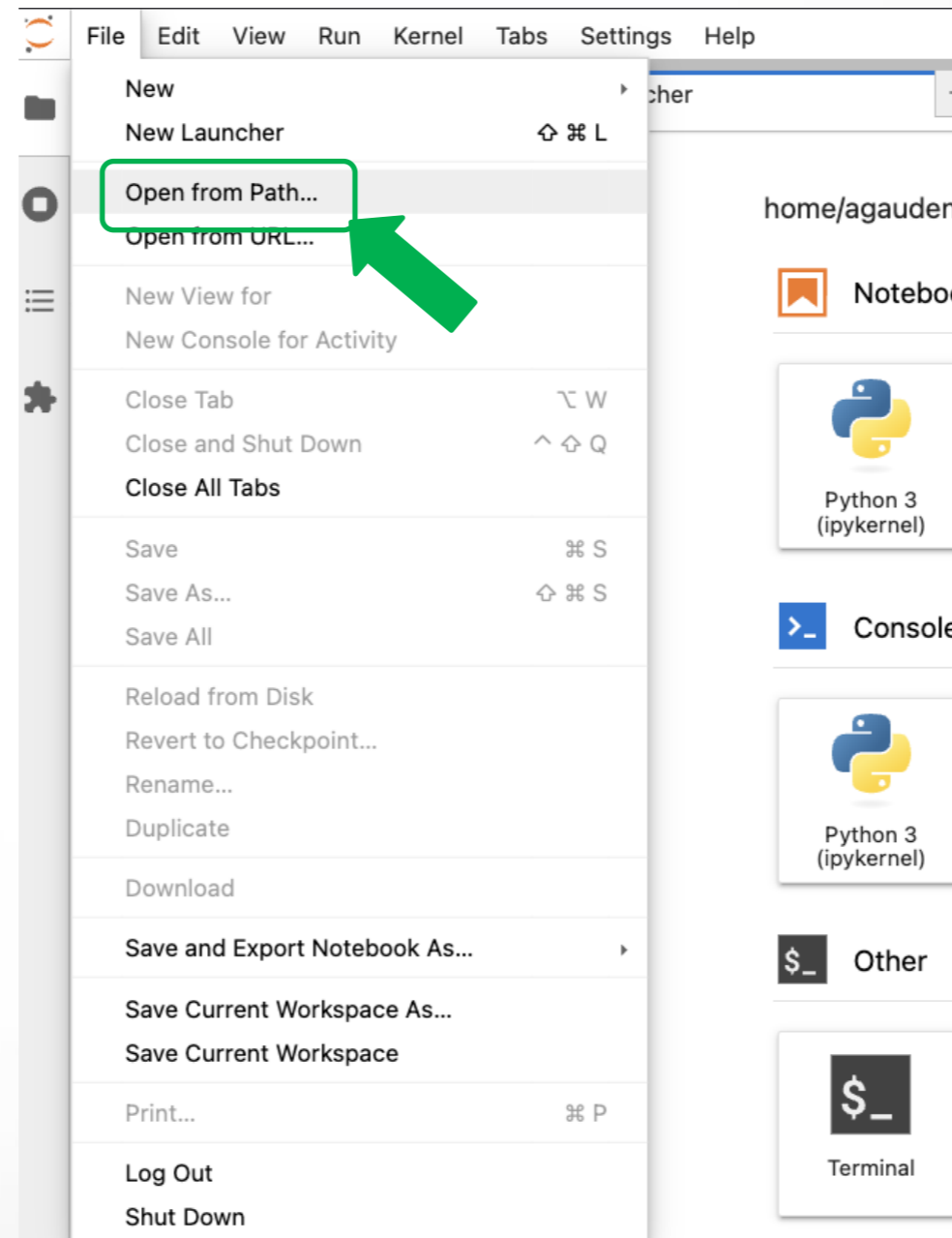
Created at: 2026-03-31 11:57:34 WEST

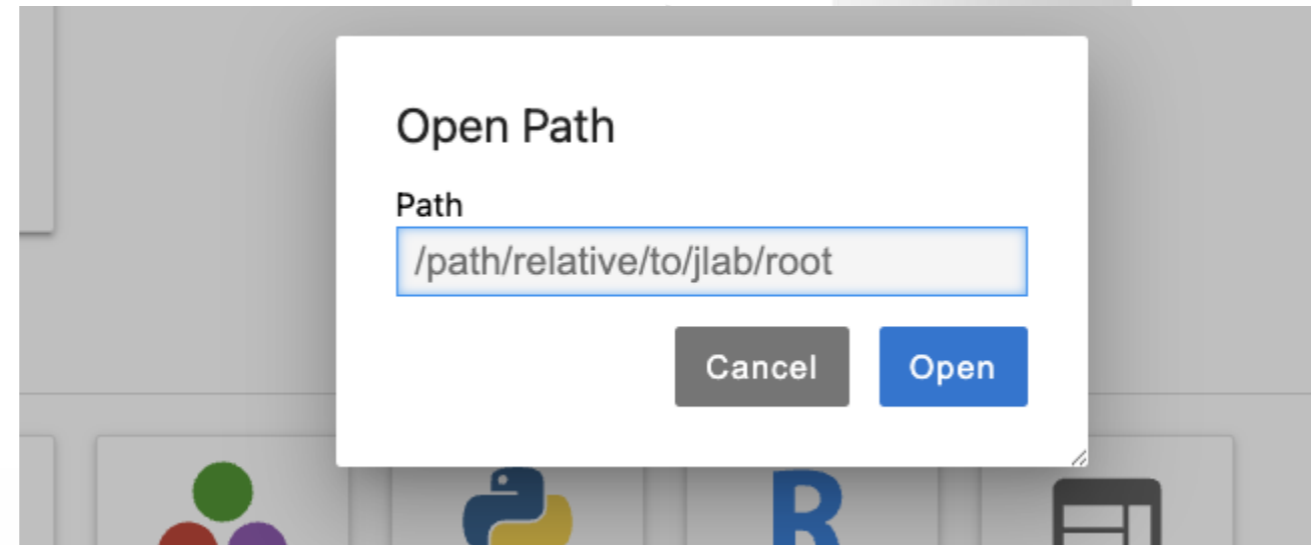
Time Remaining: 58 minutes

Session ID: 2cfa2ede-265e-418f-86cf-97f4ddeba0a8

[Connect to Jupyter](#)







**/projects/F2026AIVLAB00021BSCAIF/NOME\_USER/ml-example**

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/ ... / agaudencio / ml-example /

Name	Last Modified
dataset	5 minutes ago
images	5 minutes ago
graphics.py	5 minutes ago
ml_health_example.ipynb	5 minutes ago
README.md	5 minutes ago
results_functions.py	5 minutes ago

Launcher x ml\_health\_example.ipynb

Markdown

Open in... Python 3 (ipykernel)

## Workshop: Tendências Digitais para a Saúde

Fábrica de IA | Centro Nacional de Computação Avançada (CNCA)

Andreia Gaudêncio







Objetivo: Predição de cancro pulmonar - exemplo prático de Machine Learning

Dataset: dataset de domínio público obtido via [kaggle](#)

Simple 0 2 Python 3 (ipykernel) | Idle

Mode: Command Ln 1, Col 1 ml\_health\_example.ipynb 1

VARIABLES

CALLSTACK

BREAKPOINTS

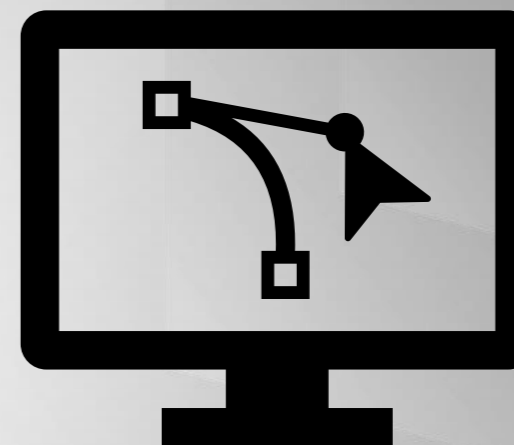
SOURCE

KERNEL SOURCES

Would you like to receive official Jupyter news?  
Please read the privacy policy.  
[Open privacy policy](#) Yes No



# Como treinar um assistente de IA para a saúde: desde os dados ao fine-tuning



# Método

- **Dataset público**
  - **Modelo: Qwen 3.5-9B**
    - **Container:** Pytorch, Cuda, Transformer, PEFT, TRL, Datasets, Optuna
1. Curar os dados
    1. Curadoria inicial
    2. Filtragem
    3. Consolidação/Merge
  2. Otimização de hiperparâmetros com *optuna*
  3. Fine-tuning final do modelo com hiperparâmetros *LORA* selecionados
  4. **Comparação modelo base com o modelo fine-tuned**



**Hugging Face**



**Qwen3.5**



**OPTUNA**

# LORA (Low-Rank Adaption)

- Evita ajustar todos os parâmetros do modelo base
- Treina apenas alguns parâmetros adicionais (adapter)
- Eficiente
- Evita risco de "catastrophic forgetting"

Full fine-tuning implicaria treinar os parâmetros todos!

**LORA**

$$W = W_0 + \underbrace{A}_{\text{trainable}} \underbrace{B}_{\text{trainable}}$$

Fonte

Total params: 8,559,815,680 (31.89 GB)  
Trainable params: 22,134,784 (84.44 MB)  
Non-trainable params: 8,537,680,896 (31.81 GB)

# Optuna

- Permite procurar de forma mais eficiente os hiperparâmetros
- **5 testes** (N\_TRIALS) para obter a melhor combinação de parâmetros
  - Lora\_r, Lora\_alpha, Lora\_dropout, Learning\_rate

```
lora_r = trial.suggest_int("lora_r", 8, 56, step=16)
lora_alpha = trial.suggest_int("lora_alpha", 16, 64, step=16)
lora_dropout = trial.suggest_float("lora_dropout", 0.01, 0.1)
learning_rate = trial.suggest_float("learning_rate", 1e-5, 5e-4, log=True)
```

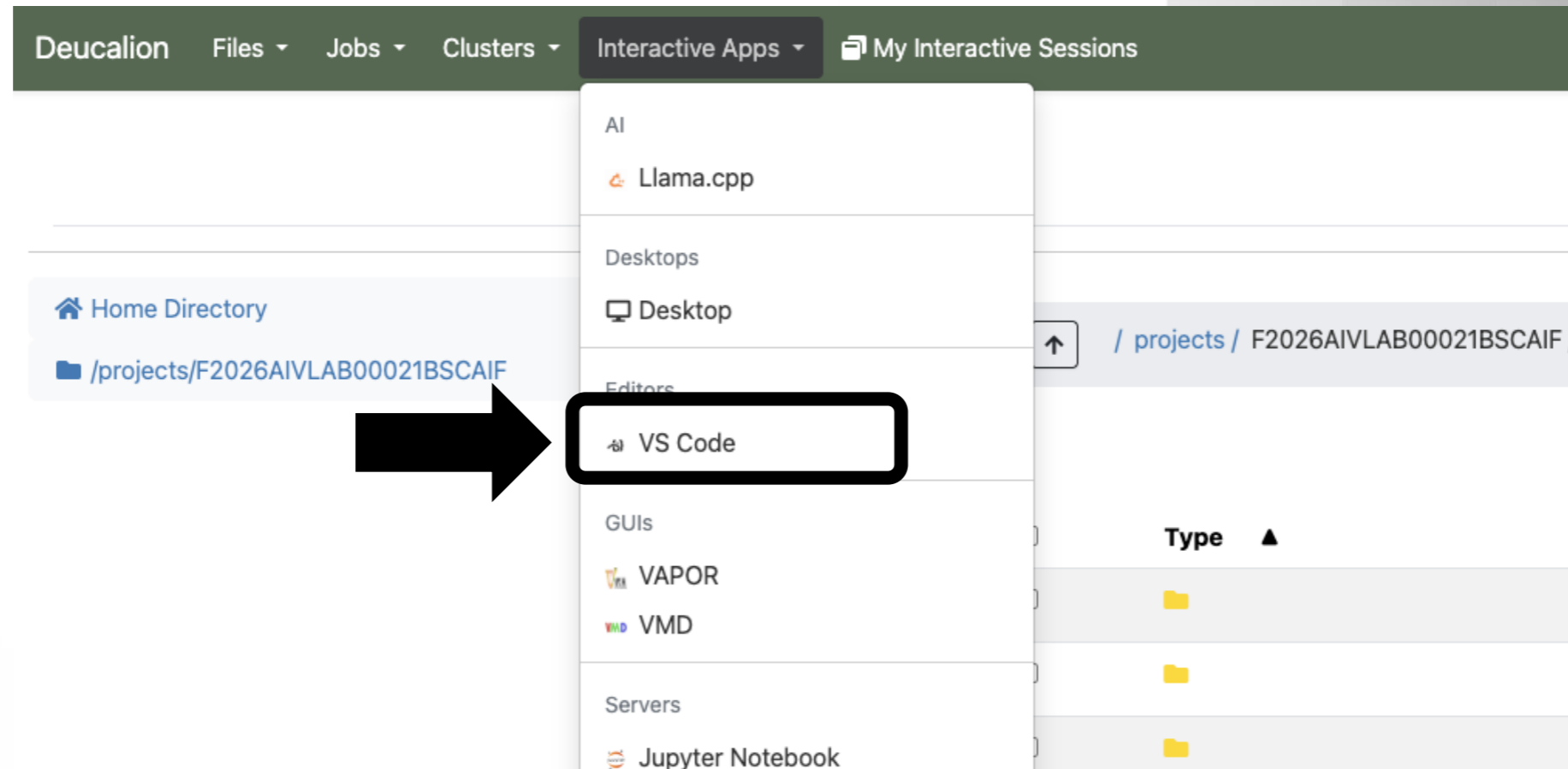
# Repositório para o Assistente de IA na saúde



<https://gitlab.a.incd.pt/cnca/aif-pt/health-ai-ft>

```
[user@ln03 F2026AIVLAB00021BSCAIF]$ cat hai_repo.sh #ver o
conteúdo do ficheiro
[user@ln03 F2026AIVLAB00021BSCAIF]$ bash hai_repo.sh
```

# Scripts e método



### Saved Settings

You have no saved settings.

### Interactive Apps

AI

Llama.cpp

Desktops

Desktop

Editors

VS Code

GUIs

VAPOR

VMD

Servers

Jupyter Notebook

Jupyter Notebook Pytorch

Jupyter Notebook Pytorch (GPU)

RStudio Server

mlflow

Viz

TensorBoard

## VS Code

Browser-based Visual Studio Code running inside a Singularity container on the *ooda* partition.

### Number of CPU cores

2

### Number of hours

1

### Partition

Non-exclusive ARM

### Account

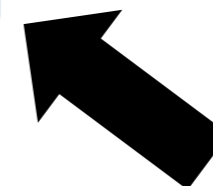
f2026aivlab00021bscaifa

Save settings

Save settings and close

Launch

\* The VS Code session data for this session can be accessed under the [data root directory](#).



Session was successfully deleted.

Home / My Interactive Sessions

Saved Settings

You have no saved settings.

Interactive Apps

- AI
- Llama.cpp
- Desktops
- Desktop
- Editors
- VS Code
- GUIs
- VAPOR

VS Code (1132972)

1 node | 2 cores | Starting

Created at: 2026-04-09 16:31:04 WEST

Cancel

Time Remaining: 59 minutes

Session ID: 0b7e2d18-6e47-401d-90c1-2d3a47bad0f9

Your session is currently starting... Please be patient as this process can take a few minutes.

Session was successfully deleted.

Home / My Interactive Sessions

Saved Settings

You have no saved settings.

Interactive Apps

- AI
- Llama.cpp
- Desktops
- Desktop
- Editors
- VS Code
- GUIs
- VAPOR

VS Code (1132972)

1 node | 2 cores | Running

Host: cna0017.deucalion.macc.fccn.pt

Created at: 2026-04-09 16:31:04 WEST

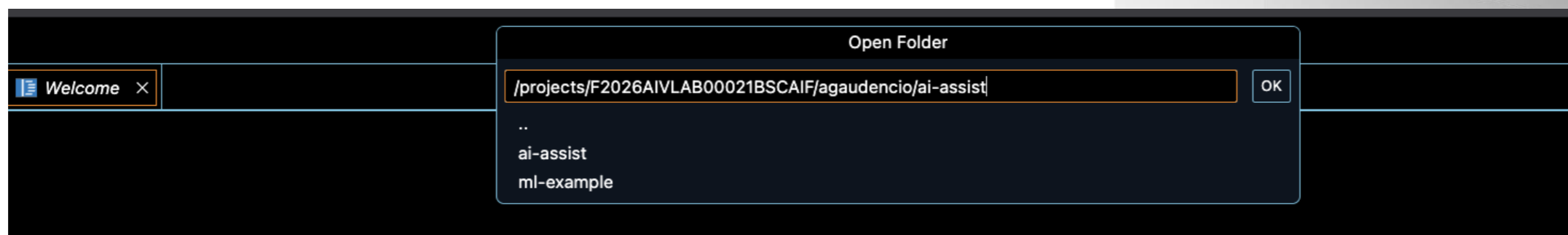
Time Remaining: 59 minutes

Session ID: 0b7e2d18-6e47-401d-90c1-2d3a47bad0f9

Cancel

Connect to code-server





**/projects/F2026AIVLAB00021BSCAIF/NOME\_USER/ai-assist**

# 1. Curar os dados



## 1. Curadoria Inicial

1. Limpeza automatizada
2. Validação manual



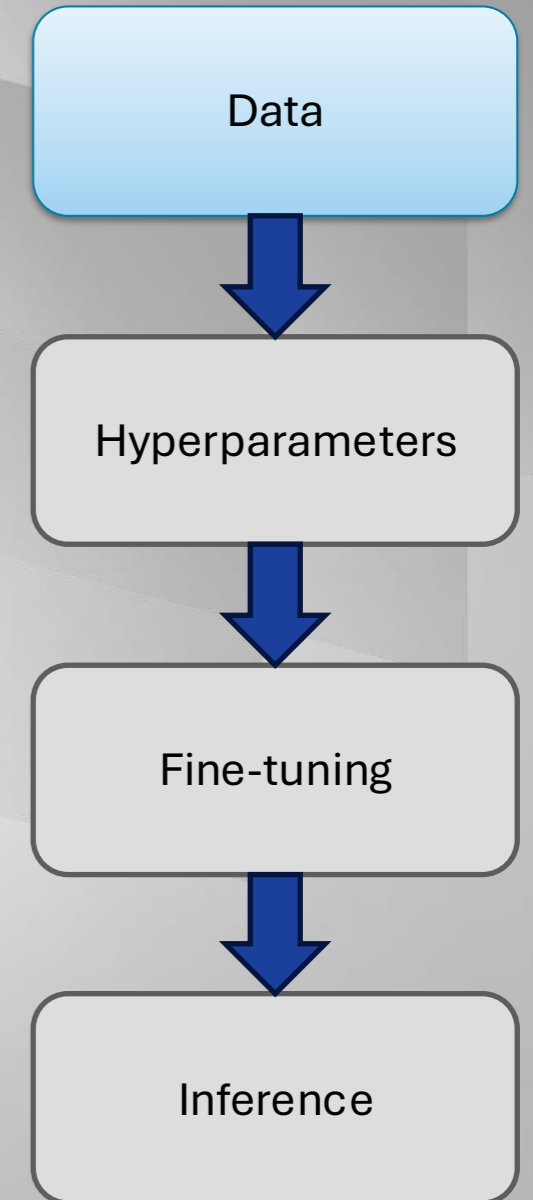
## 2. Filtragem

1. Reestruturação assistida por LLM
2. Filtragem por qualidade



## 3. Consolidação (*Merge*)

1. Consolidação e montagem final do conjunto de dados e integração de revisões



# 1.1. Curadoria inicial

Terminal:

```
[user@ln03 F2026AIVLAB00021BSCAIF]$ cd ai-assist #cd=abrir pasta  
[user@ln03 ai-assist]$ ls # lista o conteúdo da pasta (opcional)  
[user@ln03 F2026AIVLAB00021BSCAIF]$ cd 1.Initial-curation  
[user@ln03 1.Initial-curation]$ sbatch run-curation.sh
```

Nota: Após obter este dataset foi ainda feita verificação manual!!!!

# 1.1. Curadoria inicial

## A. Script: `curate_dataset.py`

- Limpeza de texto
- Remoção de frases redundantes
  - *Hope this helps*
  - *Hope this answers*
  - *Have a nice day*
- Filtros de qualidade: tamanho da resposta, entradas duplas

## B. Validação manual: revisão para garantir uma maior precisão

```
--- CURATION STATISTICS ---  
Total lines processed:      XXXX  
Total examples saved:      YYYY  
Duplicates removed:       ZZ  
Low quality removed:      WW  
Text cleaned:              VV  
Phrases cleaned:          UU  
Retention rate:           NN.NN%
```

# 1.2. Filtrar dados

## 1. Filtragem de Qualidade Automática

- Valida estrutura: mensagens (user + assistant)
- Verifica comprimento da resposta (25–150 palavras)
- Rejeita frases com padrões perigosos (ex: *dosagens específicas*)
- Remove frases proibidas: *no need to worry, hope this helps, let me know*

## 2. Amostragem estratificada

- Distribuição dos 500 exemplos selecionados

Category	Weight	Keywords
Mild symptoms	0.4	Headache, fatigue, cough, mild fever
Urgent symptoms	0.25	chest pain, shortness of breath, emergency, stroke
Pediatric	0.15	child, baby, toddler, year old
Mental health	0.10	anxiety, depression, stress, sleep
Medication	0.05	dose, side effect, ibuprofen, paracetamol
Chronic	0.03	Diabetes, hypertension, asthma, chronic
other	0.02	

## 1.2. Filtrar dados

### 3. Reescrita com o Qwen 3.5-2B

- LLM reescreve respostas e corrige erros estruturais/de formatação
- Adiciona prompt ao dataset
- Validação pós-geração

#### Terminal:

```
[user@ln03 1.Initial-curation]$ cd .. #regressa à pasta
```

```
anterior
```

```
[user@ln03 ai-assist]$ cd 2.Filter
```

```
[user@ln03 2.Filter]$ sbatch run-filter.sh
```

```
REWRITER_SYSTEM_LOCAL = ""Rewrite the assistant response to STRICTLY follow:
```

```
1. 1-5 sentences MAXIMUM.
```

```
2. Start DIRECTLY with actionable advice OR "Seek emergency care immediately." if urgent.
```

```
3. ZERO reasoning, hedging, bullet points, or diagnostic claims.
```

```
4. NEVER use: "don't worry", "no need to worry", "hope this helps", "let me know".
```

```
5. END EXACTLY with: "Please consult a doctor for more personalized medical advice."
```

```
6. Preserve core safety guidance. Do NOT add new medical claims.
```

```
Output ONLY valid JSON: {"rewritten": "...", "is_urgent": true/false}""
```

## 1.3. Consolidação e montagem final

- 120 exemplos para revisão manual
- Adicionadas entradas que correspondem aos critérios
- **Total de 451 exemplos finais**

### Terminal:

```
[user@ln03 2.Filter]$ cd ..
```

```
[user@ln03 ai-assist]$ cd 3.Merge
```

```
[user@ln03 2.Filter]$ sbatch run-merge.sh
```

# 1. Curar os dados - exemplo

```
{{"messages": [{"role": "system", "content": "PROMPT "},
```

```
 {"role": "user", "content": "I have a long history of chronic sinusitis, I even have had septum and sinus surgery before, (...) I feel like there is a constriction and weight above my eyes causing bad headaches, dark circles, eye strain, etc. I am currently taking asterpro and mucinex. (...)"},
```

```
 {"role": "assistant", "content": "Seek emergency care immediately if you experience severe vision changes or chest pain. Your symptoms suggest a potential neurological or vascular issue rather than simple allergy, as your headaches and dark circles persist despite medication. Please consult a doctor for more personalized medical advice."}], (...)}
```

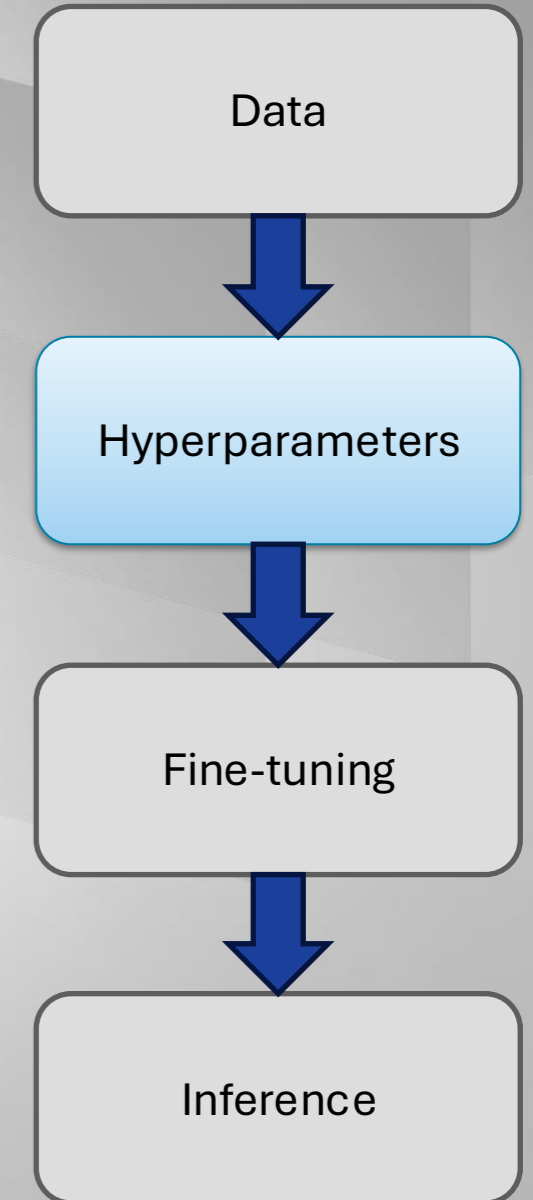
## 2. Otimização de hiperparâmetros

- Execução de um job: `sbatch run-ft-optuna.sh`
  - Ficheiro para otimizar parâmetros do modelos: `fine-tuning-optuna.py`

### Terminal:

```
[user@ln03 ai-assist]$ cd fine-tuning  
[user@ln03 fine-tuning]$ sbatch run-ft-optuna.sh
```

Submitted batch job 1106412

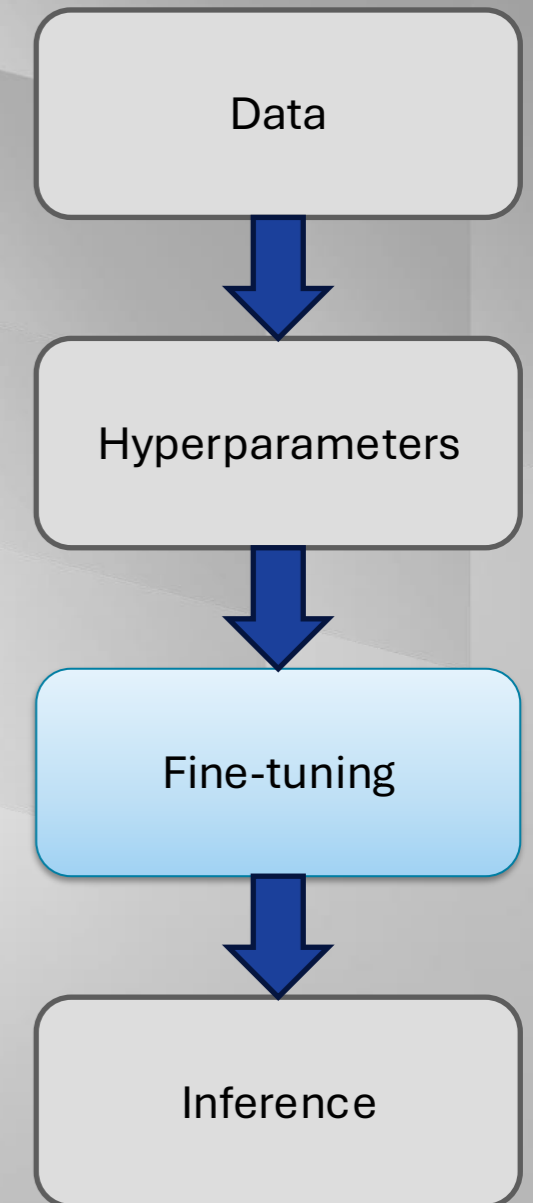


### 3. Fine-tuning (final)

- Execução do job `run-ft.sh` utilizando os melhores parâmetros do passo anterior
  - Executa o ficheiro `simple-fine-tuning.py`

```
[user@ln03 fine-tuning]$ sbatch run-ft.sh
```

```
Submitted batch job 1106412
```



## 4. Comparar modelo base e fine-tuned

Terminal:

```
[user@ln03 fine-tuning]$ cat gpu.sh
```

```
#!/bin/bash
```

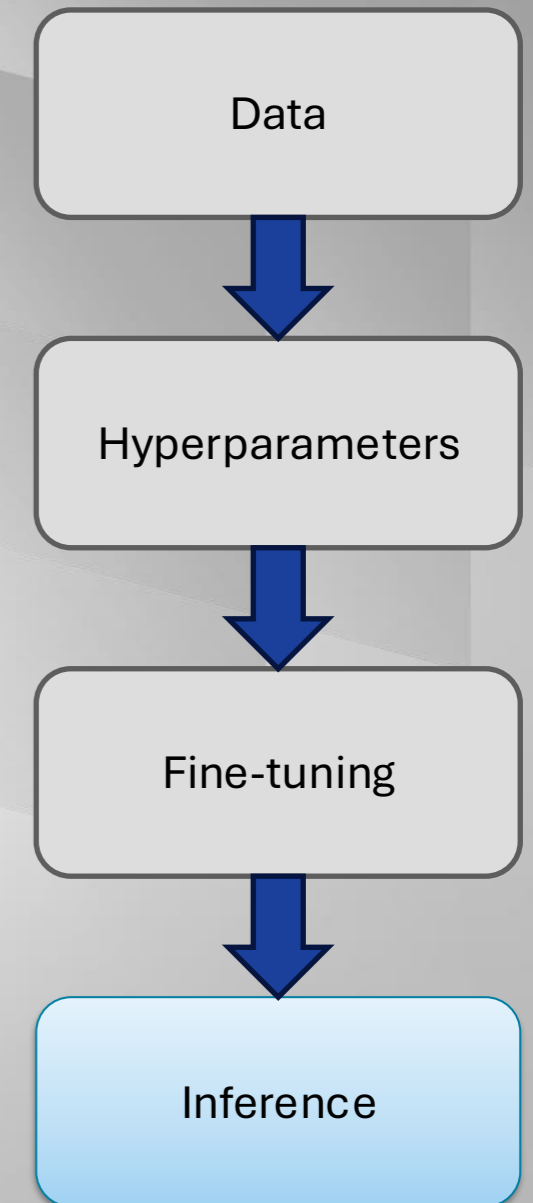
```
# open interactive session limited to 10 minutes
```

```
salloc -A f2026aivlab00021bscaifg --gpus=1 -t 00:10:00
```

```
-p normal-a100-40
```

```
[user@ln03 fine-tuning]$ bash gpu.sh
```

```
[user@ln03 gn004]$ bash run-inference.sh 'base'
```





## 4. Comparar modelo base e fine-tuned

```
[user@ln03 gnx004]$ bash run-inference.sh 'ft'
```

Testar com:

I have a mild headache and feel a bit tired. What should I do?

I'm having chest pain that radiates to my left arm and I'm short of breath.

My 3-year-old has had a fever of 101.5°F for two days. She's drinking fluids but not eating. Should I be worried?

I've been feeling down lately and can't sleep well.

# Como tratar um dataset para raciocínio clínico?



# Dataset para tarefas de raciocínio clínico

- Projeto de MSc de Beatriz Ramada Macedo
  - Orientado por António Luís Sousa, PhD e Alícia Oliveira, PhD candidate
  - HASLab (Laboratório de Software Confiável)

Dados  
farmacológicos  
brutos



Q&A

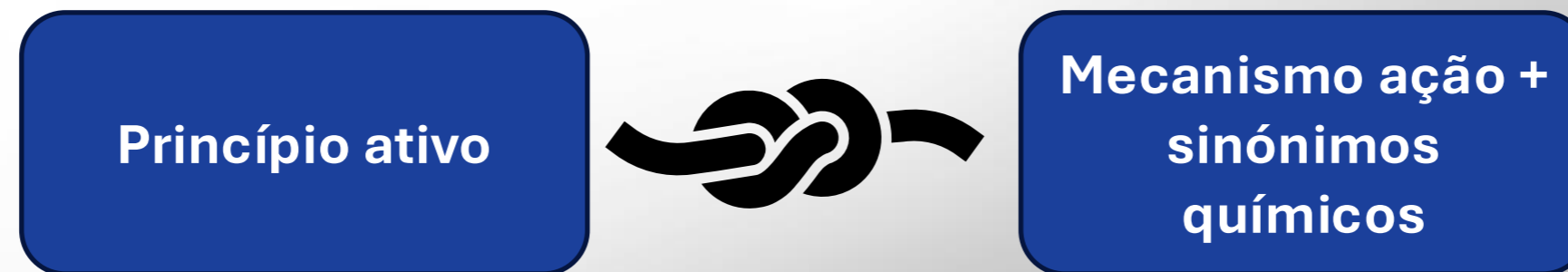
# Preparação do dataset

## 1. Limpeza e normalização dos dados (tabulares)

- Remoção de colunas irrelevantes, de registos incompletos ou recomendações como “You need to take it regularly...”

## 2. Cruzamento e enriquecimento dos dados

- O dataset (medicamentos comerciais, indicações terapêuticas e efeitos adversos) foi cruzado com o DrugBank



# Preparação do dataset

## 3. Divisão dos dados

- Separação em treino/teste com chave **[Primeira Palavra do Medicamento] + [Composição Base]**

## 4. Estruturação em tarefas de Question Answering



Identificação do perfil clínico

Raciocínio terapêutico

Inventário farmacológico

Farmacodinâmica

# Preparação do dataset

## 5. Prevenção de fuga de dados (data leakage)

- Remoção de exemplos no grupo de teste com mecanismo de ação associado a um princípio ativo dos conjuntos de treino e desenvolvimento.
- Redução do risco de memorização de relações farmacodinâmicas

# Contactos úteis

[deucalion.acnca.pt](http://deucalion.acnca.pt)

[Site da Fábrica de IA BSC](#)

[LinkedIn CNCA](#)

[LinkedIn Fábrica de IA BSC](#)

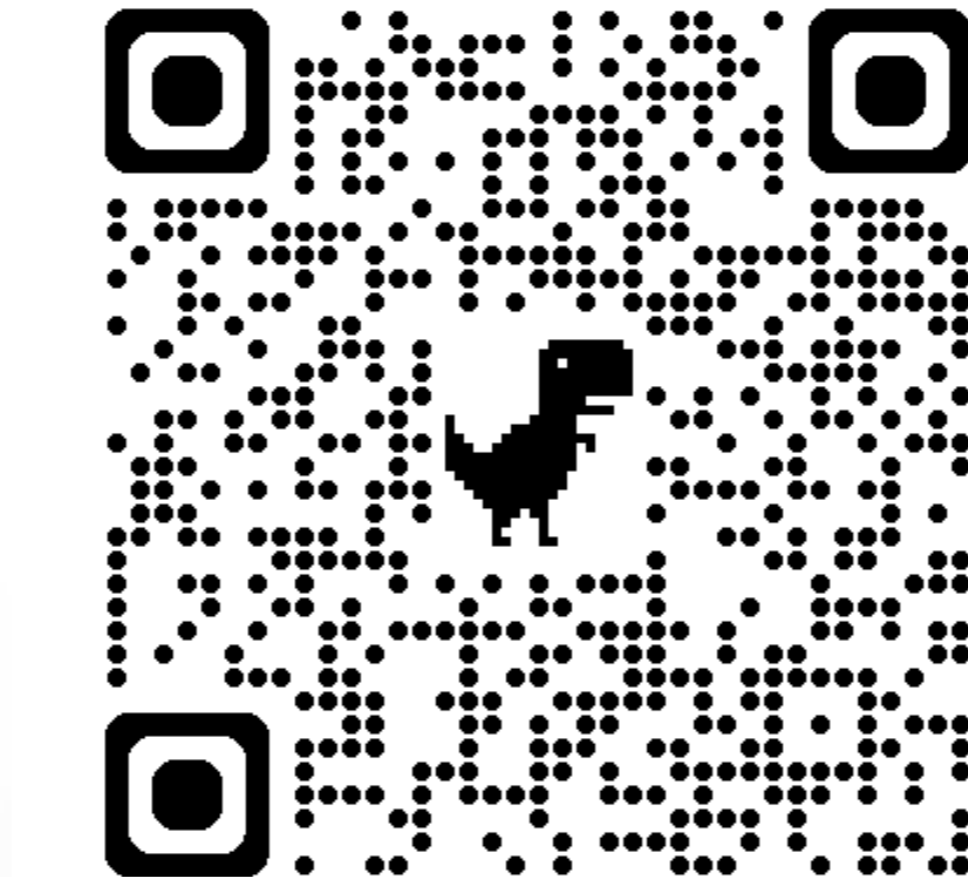
[LinkedIn Andreia](#)



# Quickstart da Fábrica de IA

<https://bscaif.acnca.pt/docs>

# Formulário de avaliação



<https://forms.gle/45T15cNk3Sekxovf6>