

MindWell: Un *chatbot open source* como axuda para a detección e o seguemento de trastornos depresivos

Eliseo Bao Souto

Dirección: Javier Parapar e Anxo Pérez

27 de Febreiro do 2024

Táboa de contidos

1. **Introdución**
2. **Fundamentos teóricos**
3. **Demostración**
4. **Produto**
5. **Metodoloxía**
6. **Conclusións, futuro e ética**

INTRODUCCIÓN

Motivación

O por que deste TFM e algúns datos clave para entender o contexto do proxecto.

¹ *Depression and other common mental disorders: global health estimates. World Health Organization et al. 2017.*

² *Trial of the effectiveness of a program for early detection and treatment of depression. A Picardi et al. 2016.*

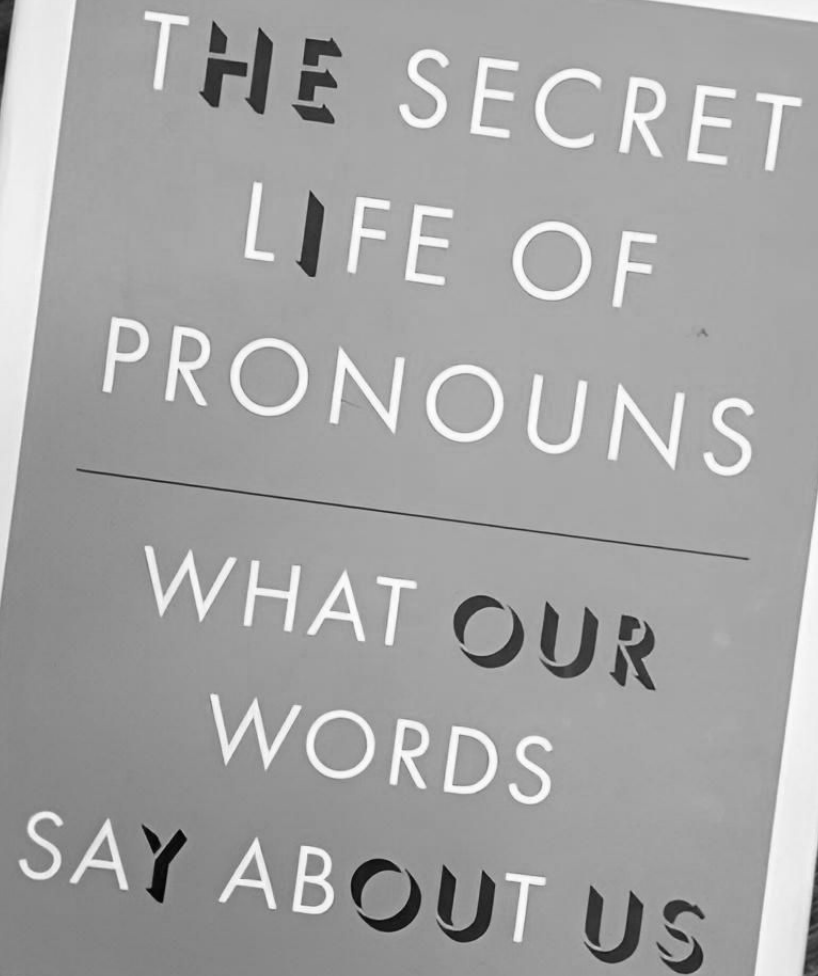
- A OMS estima que máis de **300 millóns de persoas** padecen **depresión**¹.
- A **detección temperá** é fundamental para **reducir o impacto** e as consecuencias².
- As persoas, sobretudo as máis novas, adoitan ver nas **redes sociais** un medio cómodo para **expresar** as súas **loitas emocionais**.

Como podemos **axudar**?

As nosas palabras, o noso ser³

Existe numerosa evidencia de que o xeito no que escribimos e os temas que comentamos poden ser indicadores importantes respecto do noso estado interior.

³ *Psychological Aspects of Natural Language Use: Our Words, Our Selves. Pennebaker et al.*



**THE SECRET
LIFE OF
PRONOUNS**

**WHAT OUR
WORDS
SAY ABOUT US**

Como nos **aliñamos** coas
persoas que traballan na
psicoloxía?

BDI-II⁴

*O incorporar os **síntomas** dun cuestionario clínico permite desenvolver **sistemas** máis facilmente **explicables**, **xeneralizables** e, en consecuencia, **confiáveis**.*

⁴ Beck, A. T., Steer, R. A., & Brown, G. (1996). *Beck depression inventory-II. Psychological assessment.*

Sadness

0. I do not feel sad
1. I feel sad much of the time
2. I am sad all the time
3. I am so sad or unhappy that I can't stand it

Pessimism

0. I am not discouraged about my future
1. I feel more discouraged about my future than I used to
2. I do not expect things to work out for me
3. I feel my future is hopeless and will only get worse

Loss of Energy

0. I have as much energy as ever
1. I have less energy than I used to have
2. I don't have enough energy to do very much
3. I don't have enough energy to do anything

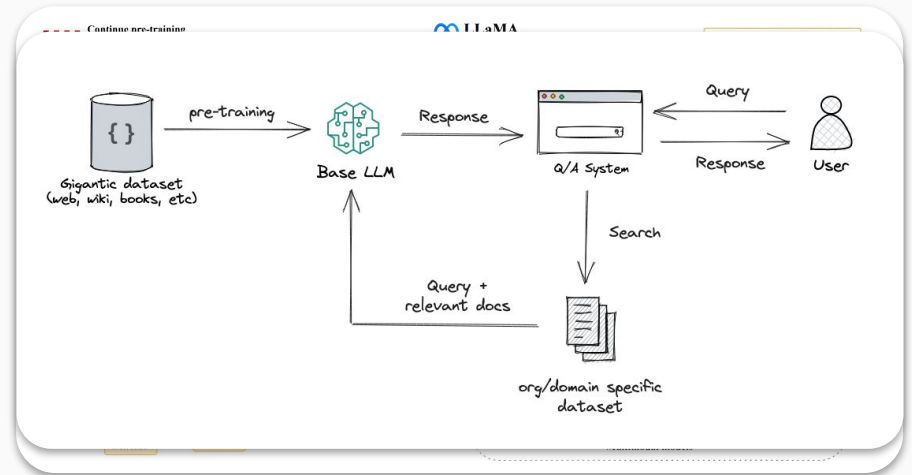
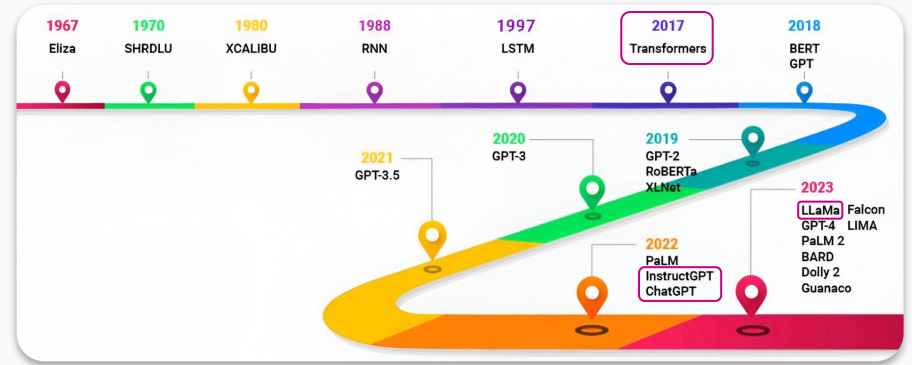
Obxectivo principal

Chatbot open source como asistente que proporcione explicacións extractivas aliñadas con **síntomas de depresión** validados para dotar ás persoas que exercen a psicoloxía dunha **ferramenta de apoio** que as descargue dunha **labor tediosa e moi intensiva en tempo** como é o feito de revisar a man o histórico de publicacións en redes sociais dunha persoa.

FUNDAMENTOS TEÓRICOS

Large Language Models (LLMs)

- Modelos adestrados con **cantidades inxentes de texto** para **predicir a seguinte palabra** dunha secuencia.
- Dado un LLM base, ***fine-tuning*** para seguir turnos de **pregunta-resposta**.



Pero os **LLMs**...

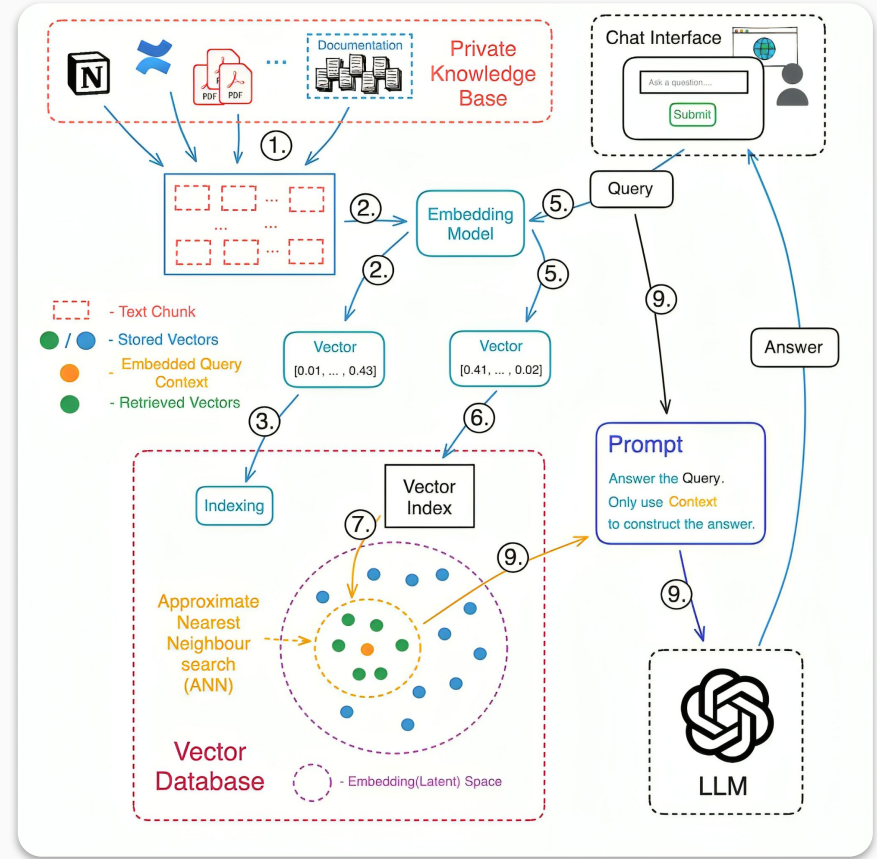
- (i) Sofren ***alucinacións***.
- (ii) Problema de acceso a **datos**.

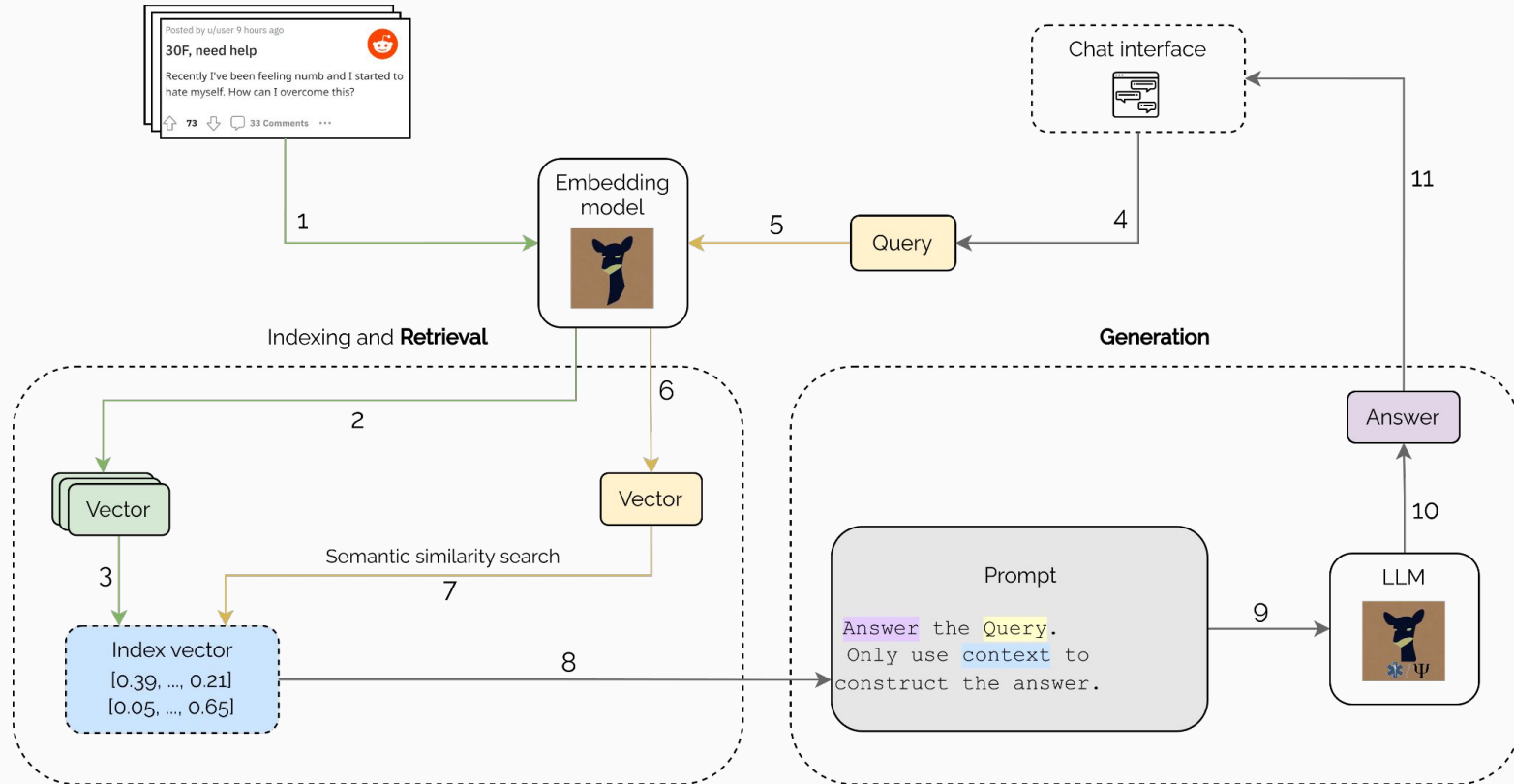
RAG!

Retrieval Augmented Generation (RAG)⁵

- Busca de información nun **corpus extenso** (*retrieval*) para logo empregar ese **contexto** na parte de xeración.
- Moi útil para **QA**, por exemplo.

⁵ Retrieval-augmented generation for knowledge-intensive NLP tasks. P Lewis et al. 2020.





DEMOSTRACIÓN

Demo

Vídeo demostrativo do funcionamento do sistema.

Amósase o máis destacado entre as funcionalidades.



The image shows the cover of a video demonstration. At the top left, there is a logo for MUEI (Módulo de Unidades de Emerxencias e Intervencións) associated with the University of Coruña. To the right of this logo is a small globe icon. The main title of the video is "MindWell: Un chatbot open source como axuda para a detección e o seguemento de trastornos depresivos". Below the title, the presenter's name "Eliseo Bao Souto" is listed, followed by the directors "Dirección: Javier Paropari e Anxo Perez". The date "27 de Febreiro do 2024" is shown at the bottom left. At the bottom right, there are logos for the University of Coruña and QRLAB.

MUEI UNIDADE DE EMERXENCIAS E INTERVENCIÓNS
UNIVERSIDADE DA CORUÑA

MindWell: Un chatbot open source como axuda para a detección e o seguemento de trastornos depresivos

Eliseo Bao Souto
Dirección: Javier Paropari e Anxo Perez

27 de Febreiro do 2024

UNIVERSIDADE DA CORUÑA QRLAB

Capacidades

(i) *Análise de **síntomas** recollidos no BDI-II.*

(ii) *Identificación de **comportamentos** concretos.*

(iii) *Ubicación **temporal e espacial.***

(iv) *Habilidade básica de **profiling.***

PRODUTO

Os datos

“Os datos son algo sumamente valioso e durarán máis que os propios sistemas.”

Tim Berners-Lee.

- **BDI-Sen**⁶ e **PsySym**⁷ como fontes de datos.

(..) I want to take the money out of my IRA and travel throughout Europe. I do not comprehend who I am. **My short-term memory is terrible, terrible, and I can not concentrate.** I see guys on TV who are going to advise some really fantastic actions for me to pursue, but ultimately I lack the willpower or energy to carry out your advice.

	BDI-Sen	PsySym
# de posts	357	752
# de explicacións	546	764

⁶ BDI-Sen: A Sentence Dataset for Clinical Symptoms of Depression. Anxo Pérez et al. 2023.

⁷ Symptom Identification for Interpretable Detection of Multiple Mental Disorders on Social Media. Zhiling Zhang et al. 2022.

Fine-tuning

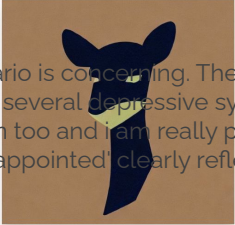
(I) Selección dun LLM conversacional, open source e SOTA.

(II) Definición da estratexia, considerando múltiples plantillas.

```
{
  "id": 0,
  "generation_options": {
    "from": "human",
    "value": "I'd like to inquire whether there are any indications of depressive symptoms in the following passage: ughhhh i'm in this position too and i am really pissed off, i feel your pain so much i'm very disappointed" clearly reflects the symptom of sadness."
  },
  "from": "gpt",
  "value": "This scenario is concerning. The text strongly suggests the presence of several depressive symptoms. For example, 'ughhhh i'm in this position too and i am really pissed off, i feel your pain so much i'm very disappointed' clearly reflects the symptom of sadness."
},
  "caption": "Vicuna (generated by stable diffusion 2.1)"
}
```

Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality
by: The Vicuna Team, Mar 30, 2023

We introduce Vicuna-13B, an open-source chatbot trained by fine-tuning LLaMA on user-shared conversational data from ShareGPT. Vicuna-13B is a generalist chatbot that performs on par with GPT-4 on a wide range of tasks, including text completion, question answering, and code generation. In fact, Vicuna-13B achieves more than 90% quality of GPT-4 on a wide range of tasks, including text completion, question answering, and code generation. The cost of training Vicuna-13B is around \$300. The code and weights, along with an online demo, are publicly available for non-commercial use.



*According to a fun and non-scientific evaluation with GPT-4. Further rigorous evaluation is needed.

Arquitectura*

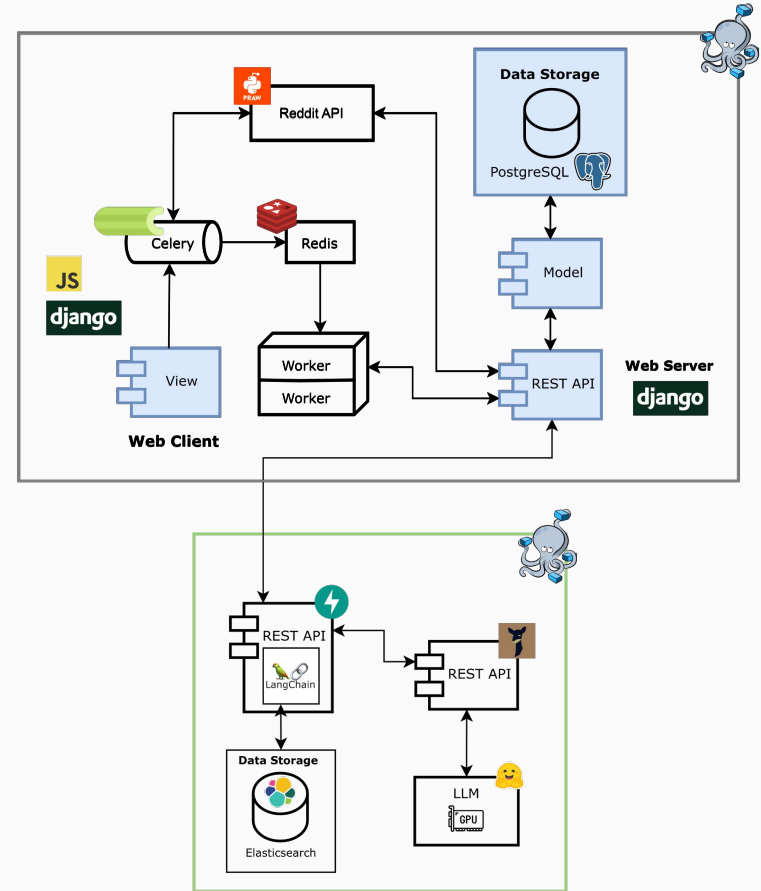
Cliente-servidor MVT.

(i) Orquestración de servicios Docker.

(ii) Componentes Redis, PRAW, etc. separadas.

(iii) Comunicación HTTP coa nova API do LLM.

*En azul, as modificacións substanciais sobre o proxecto base. En verde, o incremento asociado a este TFM.

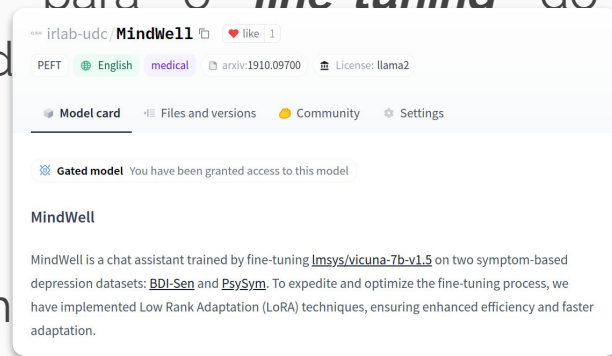


Software liberado

O uso e liberación de *software* libre foi un compromiso en firme durante todo o proceso de desenvolvemento.

- **LLM resultante** do proceso de *fine-tuning*.

- Código para o ***fine-tuning*** do modelo, incluíndo



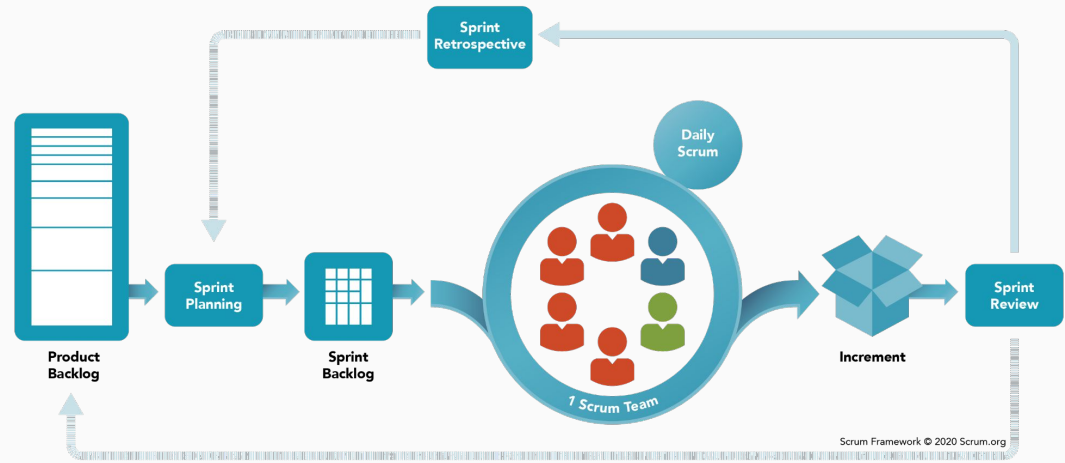
- Código para o **LangChain** e **Stablesearch**.

- Código coa **aplicación web** incrementada.

METODOLOGÍA

Metodoloxía

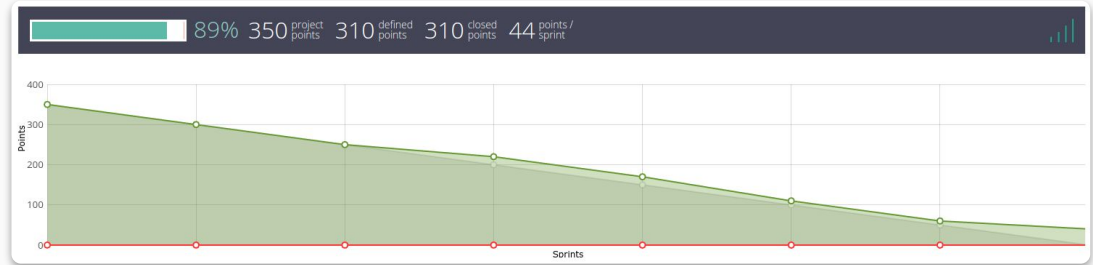
Séguese **Scrum** pero cunha serie de **modificacións**, como a supresión das *dailies*.



- **Equipo** conformado polo **autor** do traballo e os seus **directores**.
- 2-3 semanas por *Sprint*, a razón de 50 Puntos de Historia agardados.

Planificación

7 Sprints completados a razón de 44 Puntos, resultando nunhas 310 horas⁸ adicadas.



- Fase inicial de **estudo** da metodoloxía, **análise** e **asimilación** da problemática e xestión de **riscos**.
- A **metodoloxía** favorece a corrección de desviacións.

⁸Tendo en conta a estimación de 1 hora de dedicación por cada Punto de Historia.

CONCLUSIÓN, FUTURO E ÉTICA

Conclusións

(I) *Valoración da consecución ou o fracaso respecto dos obxectivos fixados.*

- Como **resultados** do proxecto, obtívose:
 - Conxunto de datos.
 - *Fine-tuning* dun LLM.
 - Conexión do LLM con publicacións en RRSS.
 - Interface *web* para consumir o sistema.

- En canto ao **aspecto formativo**, destácase:
 - Estudo de literatura e fundamentación teórica.
 - Asimilación de novas tecnoloxías.
 - Manexo dunha metodoloxía.
 - Aseguramento de competencias do título.

Traballo futuro

(III) Cara onde pensamos que pode camiñar o proxecto e cales son as nosas inquedanzas.

- Adaptación a **outros dominios** sociais, lingüísticos e culturais.
- Manexo doutros **manuais clínicos** con marcadores validados por expertos, nomeadamente **DSM-5**.
- Extensión a **outras enfermidades comórbidas** á depresión, dada a conexión entre os síntomas.

Aspecto ético

(III) *Reflexión sobre o plano ético do traballo no máis amplo do termo.*

- Supervisión e **feedback** por parte dun perfil **clínico experto**, procurando sempre ser **complemento** e non algo substitutorio.
- Adopción de medidas para garantir o cumprimento dos **acordos de uso de datos**, mantendo **privacidade** e **anonimato**.
- **2,59 Kg.** aproximados de CO₂ emitidos, o que equivale a uns 10 Km. de conducción dun coche promedio⁸.

⁸ *Quantifying the Carbon Emissions of Machine Learning. A Lacoste et al. 2019.*

Grazas!

Descarga o
software!



Licenzas

Licenzas das partes empregadas e licenza final seleccionada.

- Memoria



- Código



- LLM -> Llama2

Software	Licenza
Vicuna 7B 1.5	Llama2
Python	PSF 2
Transformers	Apache License 2.0
DeepSpeed	Apache License 2.0
FastChat	Apache License 2.0
FastAPI	MIT License
Elasticsearch	Server Side Public License
LangChain	MIT License
Django	3-Clause BSD
VSCodium	MIT License
Pip	MIT License
Docker	Apache License 2.0
Taiga	Mozilla Public License
GitLab	MIT License

Capacidades

(i) *Análise de **síntomas** recollidos no BDI-II.*

(ii) *Identificación de **comportamentos** concretos.*

(iii) *Ubicación **temporal e espacial.***

(iv) *Habilidade básica de **profiling.***

The screenshot shows a chat interface with the following text:

You
What can you do?

MindWell

You
Can you guess what's the user age?

MindWell
I'm sorry, but I don't have enough information to guess the user's age. The user mentions being in their 30s in one of the comments, but without additional information, I cannot make an accurate guess.

3. Loss of pleasure: "I would love to move somewhere where I can find \my people\", \my community", but have no idea if place like this exists."
4. Guilt: "I'm pretty sure there are many like me, \retarded misfits\"

It's important to note that the user's comments may not indicate the presence of all symptoms, but rather a combination of them. Additionally, it's possible that the user is experiencing additional symptoms not listed in the provided list.