RESERVADO

S. R.

MINISTÉRIO DA DEFESA NACIONAL
EXÉRCITO PORTUGUÊS
DIREÇÃO DE COMUNICAÇÕES E INFORMAÇÃO
CENTRO DE DESENVOLVIMENTO APLICACIONAL E BI
PROJECTO LLM

LLM Exercito

In [1]:
```
#!pip install pdfplumber pandas
#!pip install sentence-transformers faiss-cpu pandas numpy
#!pip install ocrmypdf
```

Bibliotecas Importadas

In [1]:
```
import os
import pdfplumber
import pandas as pd
import re
import unicodedata
import numpy as np
import shutil
from pathlib import Path
from collections import Counter
from tqdm import tqdm
from sentence_transformers import SentenceTransformer
import faiss
from IPython.display import display, Markdown
```

```
c:\Users\garci\AppData\Local\Python\pythoncore-3.14-64\Lib\site-packages\tqdm\aut
o.py:21: TqdmWarning: IProgress not found. Please update jupyter and ipywidgets.
See https://ipywidgets.readthedocs.io/en/stable/user_install.html
  from .autonotebook import tqdm as notebook_tqdm
```

Pasta dos PDF sacados da intranet do exército

Nesta fase e sem um repositório de documentos disponivel, utilizei um script denominado de Sacarpdf.py para savar todos os pdfs disponiveis na página do exército dentro deste diretório.

r"\intranet.exercito.local\areas-sectoriais"

Deu um conjunto de 11618 documentos.

Aqui propunha tambem tirar-mos todos os documentos existentes no GESDOC para engrossar o nosso dataset, e melhorar a LLM naturalmente.

In [2]:
```
pasta_pdfs = r"D:\Trabalhos\Bot Exército\pdfs"
```

Funções

In [3]:
```
def normalizar_texto(txt):
    if not isinstance(txt, str):
        return ""
    txt = txt.lower()
    txt = ''.join(
        c for c in unicodedata.normalize('NFD', txt)
        if unicodedata.category(c) != 'Mn')
    txt = re.sub(r"[_\-.\\/]+", " ", txt)
    txt = re.sub(r"\s+", " ", txt).strip()
    return txt
```

In [4]:
```
def classificar_documento(nome):
    if not isinstance(nome, str):
        return "Desconhecido"
    categorias = {
        "NEP": ["nep","normas de execucao permanente","norma de execucao permane
        "LISTA": ["lista","Lista de Verificacao","Lista de Verificação","LV"],
        "Diretiva": ["diretiva","dee"],
        "Regulamentos e Leis": ["regulamento","lei","regulation"],
        "Guia": ["guia"],
        "manual": ["manual","man"],
        "Norma": ["norma", "normas"],
        "NOTA": ["nota"],
        "Folheto": ["folheto","folhetos","panfleto","panfletos"],
        "Referencial": ["referencial", "UC"],
        "Publicação": ["pde","publicacao doutrinaria do exercito","pte","pad"],
        "Informçaões": [ "Informações propostas",],
        "Plano": ["plano","pme","pa"]
    }
    nome_normalizado = normalizar_texto(nome)
    encontrados = []
    for categoria, palavras in categorias.items():
        for palavra in palavras:
            palavra_norm = normalizar_texto(palavra)
            padrao = r"\b" + re.escape(palavra_norm) + r"(\d+)?\b"

            match = re.search(padrao, nome_normalizado)
```

```
            if match:
                encontrados.append((match.start(), categoria, palavra_norm))
    if encontrados:
        encontrados.sort(key=lambda x: x[0])
        return encontrados[0][1]
    return "Outros"
```

As funções seguintes são as normais para um documento, contudo eu acho que faz sentido melhorarmos aqui o refinamento da remoção de algumas palavras e ou expressões/ paragrafos tipicos. Por exemplo temos documentos que normalmente são estruturados como se de uma Ordem de operações se tratasse com 1. Missão, 2.Situação etc... este tipo de formatação pdoerá ser eliminado para melhorar o chunck, outra coisa que pode ser feito é retirar assinaturas e nomes especificos (por exemplo as assinaturas dos documentos), assim como os destinatários das notas e remetentes por exemplo, este tipo de informação.

- Remover quebras excessivas e múltiplos espaços
- Remover padrões tipo "Página X de Y"
- Remover números de página isolados
- Juntar linhas partidas
- Remover headers e rodapés
- Variável texto final

In [5]:
```
def normalizar_linha(linha):
    if not linha:
        return ""
    linha = linha.strip().lower()
    linha = re.sub(r"\s+", " ", linha)
    linha = re.sub(r"\b\d+\b", "#", linha)
    return linha
```

In [6]:
```
def extrair_linhas_por_pagina(caminho_pdf):
    paginas = []
    with pdfplumber.open(caminho_pdf) as pdf:
        for pagina in pdf.pages:
            texto = pagina.extract_text() or ""
            linhas = [l.strip() for l in texto.split("\n") if l.strip()]
            paginas.append(linhas)
    return paginas
```

In [7]:
```
def limpar_linhas_repetidas_pagina(linhas, headers, footers, top_n=5, bottom_n=5
    if not linhas:
        return []
    resultado = []
    total = len(linhas)
    for i, linha in enumerate(linhas):
        linha_norm = normalizar_linha(linha)
        if i < top_n and linha_norm in headers:
            continue
        if i >= max(0, total - bottom_n) and linha_norm in footers:
            continue
```

```
        resultado.append(linha)
    return resultado
```

In [8]:
```
def detetar_headers_footers(paginas_linhas, top_n=3, bottom_n=3, min_ratio=0.6):
    n_paginas = len(paginas_linhas)
    if n_paginas == 0:
        return set(), set()
    min_count = max(2, int(n_paginas * min_ratio))
    contador_topo = Counter()
    contador_fundo = Counter()
    for linhas in paginas_linhas:
        topo = linhas[:top_n]
        fundo = linhas[-bottom_n:] if len(linhas) >= bottom_n else linhas
        topo_norm = set(normalizar_linha(l) for l in topo if normalizar_linha(l)
        fundo_norm = set(normalizar_linha(l) for l in fundo if normalizar_linha(
        contador_topo.update(topo_norm)
        contador_fundo.update(fundo_norm)
    headers = {linha for linha, c in contador_topo.items() if c >= min_count}
    footers = {linha for linha, c in contador_fundo.items() if c >= min_count}
    return headers, footers
```

In [9]:
```
def limpar_texto_final(texto):
    if not texto:
        return ""
    texto = re.sub(r"p[aá]gina\s+\d+\s+de\s+\d+", "", texto, flags=re.IGNORECASE
    texto = re.sub(r"(?m)^\s*\d+\s*$", "", texto)
    texto = re.sub(r"(?<!\n)\n(?!\n)", " ", texto)
    texto = re.sub(r"[ \t]+", " ", texto)
    texto = re.sub(r"\n{3,}", "\n\n", texto)

    return texto.strip()
```

In [10]:
```
def extrair_texto_sem_headers_footers(caminho_pdf, top_n=3, bottom_n=3, min_rati
    paginas_linhas = extrair_linhas_por_pagina(caminho_pdf)
    headers, footers = detetar_headers_footers(
        paginas_linhas,
        top_n=top_n,
        bottom_n=bottom_n,
        min_ratio=min_ratio
    )
    paginas_limpas = []
    for linhas in paginas_linhas:
        limpas = limpar_linhas_repetidas_pagina(
            linhas,
            headers=headers,
            footers=footers,
            top_n=max(top_n, 5),
            bottom_n=max(bottom_n, 5)
        )
        paginas_limpas.append("\n".join(limpas))
    texto_final = "\n\n".join(paginas_limpas)
    texto_final = limpar_texto_final(texto_final)
    return texto_final, headers, footers
```

In [11]:
```
def processar_pasta_pdfs(pasta_base):
```

```
    resultados = []

    todos_pdfs = []

    for raiz, _, ficheiros in os.walk(pasta_base):
        for ficheiro in ficheiros:
            if ficheiro.lower().endswith(".pdf"):
                todos_pdfs.append(os.path.join(raiz, ficheiro))

    for caminho_pdf in tqdm(todos_pdfs):

        ficheiro = os.path.basename(caminho_pdf)

        try:
            texto_limpo, headers, footers = extrair_texto_sem_headers_footers(ca

            resultados.append({
                "nome_pdf": ficheiro,
                "caminho": caminho_pdf,
                "texto_limpo": texto_limpo,
                "headers_detetados": list(headers),
                "footers_detetados": list(footers)
            })

        except Exception as e:
            resultados.append({
                "nome_pdf": ficheiro,
                "caminho": caminho_pdf,
                "texto_limpo": "",
                "headers_detetados": [],
                "footers_detetados": [],
                "erro": str(e)
            })

    return pd.DataFrame(resultados)
```

Funções especificas para a parte do chunking e do FAISS

- Chunking por parágrafos com overlap
- Embeddings com sentence-transformers
- FAISS para indexação local

In [12]:
```
def limpar_paragrafos_vazios(paragrafos):
    return [p.strip() for p in paragrafos if isinstance(p, str) and p.strip()]
```

In [13]:
```
def dividir_em_paragrafos(texto):
    if not isinstance(texto, str) or not texto.strip():
        return []
    paragrafos = re.split(r"\n\s*\n", texto)
    paragrafos = limpar_paragrafos_vazios(paragrafos)
    return paragrafos
```

In [14]:
```
def chunk_texto_inteligente(texto, chunk_size=1200, overlap=200, min_chunk_size=
    if not isinstance(texto, str) or not texto.strip():
        return []
```

```
    paragrafos = dividir_em_paragrafos(texto)
    if not paragrafos:
        return []
    chunks = []
    chunk_atual = ""
    for paragrafo in paragrafos:
        paragrafo = paragrafo.strip()
        if not paragrafo:
            continue
        if len(paragrafo) > chunk_size:
            if chunk_atual.strip():
                chunks.append(chunk_atual.strip())
                chunk_atual = ""
            inicio = 0
            while inicio < len(paragrafo):
                fim = inicio + chunk_size
                subchunk = paragrafo[inicio:fim].strip()
                if subchunk:
                    chunks.append(subchunk)
                inicio += max(1, chunk_size - overlap)
            continue
        candidato = f"{chunk_atual}\n\n{paragrafo}".strip() if chunk_atual else
        if len(candidato) <= chunk_size:
            chunk_atual = candidato
        else:
            if len(chunk_atual.strip()) >= min_chunk_size:
                chunks.append(chunk_atual.strip())
                if overlap > 0:
                    tail = chunk_atual.strip()[-overlap:]
                    chunk_atual = f"{tail}\n\n{paragrafo}".strip()
                else:
                    chunk_atual = paragrafo
            else:
                chunk_atual = candidato
    if chunk_atual.strip():
        chunks.append(chunk_atual.strip())
    return chunks
```

In [15]:
```
def pesquisar_chunks(query, modelo_embeddings, index, dataset_chunks, top_k=5):
    query_embedding = modelo_embeddings.encode([query], convert_to_numpy=True)
    query_embedding = np.array(query_embedding, dtype="float32")

    distancias, indices = index.search(query_embedding, top_k)

    resultados = dataset_chunks.iloc[indices[0]].copy()
    resultados["score_distancia"] = distancias[0]

    return resultados
```

In [16]:
```
def criar_dataset_chunks(dataset):
    registos = []

    for _, row in dataset.iterrows():
        nome_pdf = row.get("nome_pdf", "")
        caminho = row.get("caminho", "")
        categoria = row.get("categoria", "Outros")
        chunks = row.get("chunks", [])

        for i, chunk in enumerate(chunks):
```

```
            registos.append({
                "nome_pdf": nome_pdf,
                "caminho": caminho,
                "categoria": categoria,
                "chunk_id": i,
                "texto_chunk": chunk
            })

    return pd.DataFrame(registos)
```

In [17]:
```
def garantir_nome_unico(destino):
    base, ext = os.path.splitext(destino)
    contador = 1

    while os.path.exists(destino):
        destino = f"{base}_{contador}{ext}"
        contador += 1

    return destino
```

RAG

In [56]:
```
def construir_contexto_rag(query, resultados):
    contexto = "\n\n---\n\n".join(
        [
            f"[Documento: {row['nome_pdf']} | Categoria: {row['categoria']}]\n{r
            for _, row in resultados.iterrows()
        ]
    )

    prompt = f"""
Responde à pergunta do utilizador com base apenas no contexto fornecido.

Pergunta:
{query}

Contexto:
{contexto}

Instruções:
- Se a resposta não estiver claramente no contexto, diz que não encontraste info
- Resume de forma objetiva.
- Cita o nome do documento quando relevante.
"""
    return prompt
```

In [62]:
```
def preparar_rag(query, modelo_embeddings, index, dataset_chunks, top_k=5):
    resultados = pesquisar_chunks(
        query=query,
        modelo_embeddings=modelo_embeddings,
        index=index,
        dataset_chunks=dataset_chunks,
        top_k=top_k
    )
    prompt = construir_contexto_rag(query, resultados)
    return prompt, resultados
```

Função para teste de perguntas ao modelo

In [64]:
```
def testar_perguntas(perguntas, modelo_embeddings, index, dataset_chunks, top_k=

    for pergunta in perguntas:
        print("\n" + "="*100)
        print("🔎 PERGUNTA:", pergunta)
        print("="*100)

        resultados = pesquisar_chunks(
            query=pergunta,
            modelo_embeddings=modelo_embeddings,
            index=index,
            dataset_chunks=dataset_chunks,
            top_k=top_k
        )

        for i, row in resultados.iterrows():
            print(f"\nResultado {i+1}")
            print("Documento:", row["nome_pdf"])
            print("Categoria:", row["categoria"])
            print("Score:", row["score_distancia"])
            print("Texto:")
            print(row["texto_chunk"][:400])
            print("-"*80)
```

Pré processamento do Corpus

In [ ]:
```
dados = []
for raiz, subpastas, ficheiros in os.walk(pasta_pdfs):
    for ficheiro in ficheiros:
        if ficheiro.lower().endswith(".pdf"):
            caminho_pdf = os.path.join(raiz, ficheiro)
            texto = ""
            try:
                with pdfplumber.open(caminho_pdf) as pdf:
                    for pagina in pdf.pages:
                        texto += pagina.extract_text() or ""
            except:
                texto = "ERRO OCR??"
            dados.append({
                "nome_pdf": ficheiro,
                "caminho": caminho_pdf,
                "texto": texto
            })
dataset = pd.DataFrame(dados)
print("PDFs encontrados:", len(dataset))
dataset.head()
```

In [ ]:
```
dataset["categoria"] = dataset["nome_pdf"].apply(classificar_documento)
```

In [ ]:
```
dataset
```

Out[ ]:

	nome_pdf	caminho	texto	c
0	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS\nCENTR...	
1	PEMGFA_GER_006_GUIA DO ADIDO DE DEFESA.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	NÃO CLASSIFICADO\nESTADO-MAIOR-GENERAL DAS FOR...	
2	LISTA DE ADIDOS ESTRANGEIROS.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	LISTA DE ADIDOS MILITARES ACREDITADOS EM PORTU...	
3	LISTA ADIDOS NACIONAIS.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	LISTA DE ADIDOS MILITARES PORTUGUESES ACREDITA...	
4	Diretiva Ambiental do Exercito 2025-2027.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...		
...	...	...	...	
11613	Folheto_a verdade sobre o Crack.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...		
11614	Folheto_a verdade sobre o ecstasy.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...		
11615	Panfletos PIEDs.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...	MISTURAR SUBSTÂNCIAS AUMENTA EXPONENCIALMENTE ...	
11616	MAN 2016RHV03 - Processo Individual v1.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SIGDN_R...	MINISTÉRIO DA DEFESA NACIONAL\nSecretaria-Gera...	
11617	MAN 2016RHV07 - Assiduidades v1.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SIGDN_R...	MINISTÉRIO DA DEFESA NACIONAL\nSecretaria-Gera...	

11618 rows × 4 columns

Gravei em .csv e .json o dataset puro com o nome do indexpdfs e o que é para a construção da LLM gravei apenas como json em dataset_llm

In [ ]:
```
dataset.to_csv(r"D:\Trabalhos\Bot Exército\pdfs\indexpdfs.csv", index=False)
dataset.to_json(r"D:\Trabalhos\Bot Exército\pdfs\indexpdfs.json", orient="record
dataset[["nome_pdf", "categoria", "texto"]].to_json(
    "dataset_llm.json",
    orient="records",
    force_ascii=False,
    indent=2
)
```

In [ ]:
```
dataset = pd.read_json(r"D:\Trabalhos\Bot Exército\pdfs\dataset_llm.json")
```

Aqui para o refinamento dos metadados do nosso vetor seria interessante melhorar aqui os termos pada procura das funções de REGEX aplicadas, o dataset está muito desbalanceado, de qualquer forma isto serve apenas para fine tuning do modelo.

In [ ]:
```
dataset["categoria"].value_counts()
```

Out[ ]:
```
categoria
Outros         10829
NOTA             605
manual            50
NEP               27
Guia              25
LISTA             24
Plano             23
Diretiva          20
Folheto           13
Referencial        2
Name: count, dtype: int64
```

In [ ]:
```
dataset
```

Out[ ]:

	nome_pdf	categoria	texto
0	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	Guia	ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS\nCENTR...
1	PEMGFA_GER_006_GUIA DO ADIDO DE DEFESA.pdf	Guia	NÃO CLASSIFICADO\nESTADO-MAIOR-GENERAL DAS FOR...
2	LISTA DE ADIDOS ESTRANGEIROS.pdf	LISTA	LISTA DE ADIDOS MILITARES ACREDITADOS EM PORTU...
3	LISTA ADIDOS NACIONAIS.pdf	LISTA	LISTA DE ADIDOS MILITARES PORTUGUESES ACREDITA...
4	Diretiva Ambiental do Exercito 2025-2027.pdf	Diretiva	
...	...	...	...
11613	Folheto_a verdade sobre o Crack.pdf	Folheto	
11614	Folheto_a verdade sobre o ecstasy.pdf	Folheto	
11615	Panfletos PIEDs.pdf	Outros	MISTURAR SUBSTÂNCIAS AUMENTA EXPONENCIALMENTE ...
11616	MAN 2016RHV03 - Processo Individual v1.pdf	Outros	MINISTÉRIO DA DEFESA NACIONAL\nSecretaria-Gera...
11617	MAN 2016RHV07 - Assiduidades v1.pdf	Outros	MINISTÉRIO DA DEFESA NACIONAL\nSecretaria-Gera...

11618 rows × 3 columns

In [ ]:
```
dataset = processar_pasta_pdfs(pasta_pdfs)
```

```
Could not get FontBBox from font descriptor because None cannot be parsed as 4 fl
oats
Could not get FontBBox from font descriptor because None cannot be parsed as 4 fl
oats
Could not get FontBBox from font descriptor because None cannot be parsed as 4 fl
oats
```

In [ ]:
```
dataset.head()
```

Out[ ]:

	nome_pdf	caminho	texto_limpo	headers_detetados	foo
0	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS CENTRO...	[]	
1	PEMGFA_GER_006_GUIA DO ADIDO DE DEFESA.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS CENTRO...	[não classificado]	[r
2	LISTA DE ADIDOS ESTRANGEIROS.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	LISTA DE ADIDOS MILITARES ACREDITADOS EM PORTU...	[defesa]	
3	LISTA ADIDOS NACIONAIS.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	LISTA DE ADIDOS MILITARES PORTUGUESES ACREDITA...	[]	
4	Diretiva Ambiental do Exercito 2025-2027.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...		[]	

Havia 3 documentos que deram erro, na como não eram "relevantes" acabei por os eliminar

In [ ]:
```
dataset["erro"].value_counts()
```

Out[ ]:
```
erro
Invalid dictionary construct: [/'Registry', b'', /'Ordering', b'', /'Supplemen
t']    1
Unexpected EOF
1
No /Root object! - Is this really a PDF?
1
Name: count, dtype: int64
```

In [ ]:
```
dataset[dataset["erro"].notna()]
```

Out[ ]:

	nome_pdf	caminho	texto_limpo	headers_
2711	849-D - 2120001806_signed.pdf	D:\Trabalhos\Bot Exército\pdfs\Financas\pce\pm...		
9814	LISTAS DE ANTIGUIDADE DE 1963 Oficiais.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\AdmRH\g...		
11585	GSST_Relatorio anual-estatisticas2017+Despacho...	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...		

In [ ]:
```
dataset = dataset[dataset["erro"].isna()]
```

In [ ]:
```
dataset["erro"].value_counts()
```

Out[ ]:
```
Series([], Name: count, dtype: int64)
```

Criação dos Chunks para análise (não finais)

Aqui o afinar dos parametros dos chunks pode tambem ser algo a ser considerado para melhorar o modelo

In [ ]:
```
dataset["chunks"] = dataset["texto_limpo"].apply(lambda x: chunk_texto_inteligen
```

In [ ]:
```
dataset[["nome_pdf", "chunks"]].head(1)
```

Out[ ]:

	nome_pdf	chunks
0	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	[ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS CENTR...

Divisão em chunks e OCR

Chunks por documento, aqui reparei que havia documentos com num_chunks de 0 o que já era previsto porque havia documentos que eram digitalizados então paralelamente existe um script que faz o OCR desses documentos o conjunto de comandos seguintes identificaram quais os documentos sem chunks e com 1 ou 2 chunks e foram passados por OCR pelo script denominado de OCR.ipynb.

Apliquei tambem um critério na questão do texto limpo ser menor de 200 caracteres porque muitas das vezes temos documentos que tem a folha de rosto, por exemplo, em pdf e o resto do corpo está digitalizado, assim cobre tambem esta possibilidade.

A solução passou por identificar quais eram os documentos que necessitavam de OCR tendo em conta os critérios estabelecidos anteriormente e por estes documentos todos numa pasta aparte e aplicar OCR (script) sobre os mesmo.

In [ ]:
```
dataset["num_chunks"] = dataset["chunks"].apply(len)
dataset[["nome_pdf", "num_chunks"]].head()
```

Out[ ]:

	nome_pdf	num_chunks
0	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	112
1	PEMGFA_GER_006_GUIA DO ADIDO DE DEFESA.pdf	237
2	LISTA DE ADIDOS ESTRANGEIROS.pdf	9
3	LISTA ADIDOS NACIONAIS.pdf	2
4	Diretiva Ambiental do Exercito 2025-2027.pdf	0

In [ ]:
```
dataset[(dataset["num_chunks"] == 0) |(dataset["texto_limpo"].str.len() < 200)]
```

Out[ ]:

	nome_pdf	caminho	texto_limpo	headers_detet
4	Diretiva Ambiental do Exercito 2025-2027.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...		
18	Politica Ambiental Exercito_2025_Capa.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	d	
24	FND2017.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...		
25	FND2018.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...		
50	Teste 2paginas.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...		[te
...	...	...	...	
11610	Folheto_ a verdade sobre o LSD.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...		
11611	Folheto_a verdade sobre a Metanfetamina.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...		
11612	Folheto_a verdade sobre as drogas.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...		
11613	Folheto_a verdade sobre o Crack.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...		
11614	Folheto_a verdade sobre o ecstasy.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...		

1234 rows × 8 columns

In [ ]:
```
pasta_ocr = r"D:\Trabalhos\Bot Exército\OCR"
os.makedirs(pasta_ocr, exist_ok=True)
```

In [ ]:
```
docs_ocr = dataset[(dataset["num_chunks"] == 0)|
    (dataset["texto_limpo"].str.len() < 200)]
```

In [ ]:
```
for _, row in docs_ocr.iterrows():
    origem = row["caminho"]
    nome = row["nome_pdf"]
    destino = os.path.join(pasta_ocr, nome)
    destino = garantir_nome_unico(destino)
    try:
        shutil.move(origem, destino)
    except Exception as e:
        print(f"Erro ao copiar {nome}: {e}")
```

In [ ]:
```
dataset_chunks = criar_dataset_chunks(dataset)
dataset_chunks.head()
```

Out[ ]:

	nome_pdf	caminho	categoria	chunk_id	texto_chunk
0	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	0	ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS CENTRO...
1	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	1	janeiro de 2018, que deverá ser destruído. CHE...
2	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	2	DAS REGISTO DE ALTERACÕES J Identificação ou c...
3	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	3	DOS 25 1. Visitas Organizadas 25 2. Plano Anua...
4	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	4	a A-1 ANEXO B Ficha de Dados Biográficos B-1 A...

Embeddings utilizei este mas agora é uma questão de testar com outros especificos por exemplo para a lingua portuguesa só usei este porque há documentos com várias linguas.

Aqui é necessário uma decisão importante, que tem a ver com custo beneficio, e para alem de termos muitas das vezes documentos, ou podermos vir a ter por exemplo

manuais, que não estão escritos em portugues neste caso temos que ir buscar um emmbeding multilingua. Eu por limitações da máquina que tenho em casa usei o anterior que é pequeno mas propunha a subistutição ou para o:

- Multilingual-E5-large - multilinguas - se virmos este tem condições especificas como 'query' ou 'passage'.
- BGE-M3 - multilinguas - Tem git disponivel aqui.
- Legal BERTimbau (adaptado) - focado em PT, construido sobre um corpus de referencias legais, o que pode ser adequado no nosso caso pela forma de construção de texto em artigos e nós em pontos do genero de ordem de operações.

In [ ]:
```
modelo_embeddings = SentenceTransformer("sentence-transformers/paraphrase-multil
```

```
Warning: You are sending unauthenticated requests to the HF Hub. Please set a HF_
TOKEN to enable higher rate limits and faster downloads.
c:\Users\garci\AppData\Local\Python\pythoncore-3.14-64\Lib\site-packages\huggingf
ace_hub\file_download.py:129: UserWarning: `huggingface_hub` cache-system uses sy
mlinks by default to efficiently store duplicated files but your machine does not
support them in C:\Users\garci\.cache\huggingface\hub\models--sentence-transforme
rs--paraphrase-multilingual-MiniLM-L12-v2. Caching files will still work but in a
degraded version that might require more space on your disk. This warning can be
disabled by setting the `HF_HUB_DISABLE_SYMLINKS_WARNING` environment variable. F
or more details, see https://huggingface.co/docs/huggingface_hub/how-to-cache#lim
itations.
To support symlinks on Windows, you either need to activate Developer Mode or to
run Python as an administrator. In order to activate developer mode, see this art
icle: https://docs.microsoft.com/en-us/windows/apps/get-started/enable-your-devic
e-for-development
  warnings.warn(message)
Loading weights: 100%|██████████| 199/199 [00:00<00:00, 18943.00it/s]
BertModel LOAD REPORT from: sentence-transformers/paraphrase-multilingual-MiniLM-
L12-v2
Key                     | Status     |  |
------------------------+------------+--+-
embeddings.position_ids | UNEXPECTED |  |

Notes:
- UNEXPECTED    :can be ignored when loading from different task/architecture; no
t ok if you expect identical arch.
```

In [ ]:
```
embeddings = modelo_embeddings.encode(
    dataset_chunks["texto_chunk"].tolist(),
    show_progress_bar=True,
    convert_to_numpy=True
)
```

```
Batches: 100%|██████████| 6699/6699 [38:41<00:00,  2.89it/s]
```

In [ ]:
```
dimensao = embeddings.shape[1]

index = faiss.IndexFlatL2(dimensao)
index.add(np.array(embeddings, dtype="float32"))

print("Número de chunks indexados:", index.ntotal)
```

```
Número de chunks indexados: 214360
```

Criação dos metadados/dataset

Vamos então voltar a repetir o processo todo agora com os PDF's que foram alvos de OCR já integrados no dataset

In [18]:
```
pasta_pdfs = r"D:\Trabalhos\Bot Exército\pdfs"
```

In [19]:
```
dados = []
for raiz, subpastas, ficheiros in os.walk(pasta_pdfs):
    for ficheiro in ficheiros:
        if ficheiro.lower().endswith(".pdf"):
            caminho_pdf = os.path.join(raiz, ficheiro)
            texto = ""
            try:
                with pdfplumber.open(caminho_pdf) as pdf:
                    for pagina in pdf.pages:
                        texto += pagina.extract_text() or ""
            except:
                texto = "ERRO OCR??"
            dados.append({
                "nome_pdf": ficheiro,
                "caminho": caminho_pdf,
                "texto": texto
            })
dataset = pd.DataFrame(dados)
print("PDFs encontrados:", len(dataset))
dataset.head()
```

```
Could not get FontBBox from font descriptor because None cannot be parsed as 4 fl
oats
Could not get FontBBox from font descriptor because None cannot be parsed as 4 fl
oats
Could not get FontBBox from font descriptor because None cannot be parsed as 4 fl
oats
PDFs encontrados: 11616
```

Out[19]:

	nome_pdf	caminho	texto
0	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS\nCENTR...
1	PEMGFA_GER_006_GUIA DO ADIDO DE DEFESA.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	NÃO CLASSIFICADO\nESTADO-MAIOR-GENERAL DAS FOR...
2	LISTA DE ADIDOS ESTRANGEIROS.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	LISTA DE ADIDOS MILITARES ACREDITADOS EM PORTU...
3	LISTA ADIDOS NACIONAIS.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	LISTA DE ADIDOS MILITARES PORTUGUESES ACREDITA...
4	Politica Ambiental Exercito 2025.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	dPOLÍTICA AMBIENTAL DO EXÉRCITO\nExército Port...

In [20]:
```
dataset["categoria"] = dataset["nome_pdf"].apply(classificar_documento)
```

In [21]:
```
dataset
```

Out[21]:

	nome_pdf	caminho	
0	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	ESTADO-GENERAL DAS ARMADAS\n
1	PEMGFA_GER_006_GUIA DO ADIDO DE DEFESA.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	CLASSIFICADO\nE MAIOR-GENEI
2	LISTA DE ADIDOS ESTRANGEIROS.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	LISTA DE MILITARES ACREI EM
3	LISTA ADIDOS NACIONAIS.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	LISTA DE MILITARES PORTU ACI
4	Politica Ambiental Exercito 2025.pdf	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	dPOLÍTICA AM DO EXÉRCITO\r
...	...	...	
11611	Prevenir_as_Perturbacoes_Musculo-Esqueleticas.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...	F A C T S\n4\nA a E u r o p
11612	Seguranca Higiene Saude - Sector Florestal.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...	Índice C Preâmbulo.....
11613	Panfletos PIEDs.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...	MISTURAR SUBST AU EXPONENCIALM
11614	MAN 2016RHV03 - Processo Individual v1.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SIGDN_R...	MINISTÉRIO DA NACIONAL\nSe
11615	MAN 2016RHV07 - Assiduidades v1.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\SIGDN_R...	MINISTÉRIO DA NACIONAL\nSe

11616 rows × 4 columns

Aqui a semelhança do que fiz no pré processamento gravei em .csv e .json o dataset puro com o nome do indexpdfs_final e o que é para a construção da LLM gravei apenas como json em dataset_llm_final.

Aqui uma solução mais robusta seria colocar o nosso dataset num BD, tipo o weviate ou outro.

In [ ]:
```
dataset.to_csv(r"D:\Trabalhos\Bot Exército\pdfs\indexpdfs_final.csv", index=Fals
dataset.to_json(r"D:\Trabalhos\Bot Exército\pdfs\indexpdfs_final.json", orient="
```

```
dataset[["nome_pdf", "categoria", "texto"]].to_json(
    "dataset_llm_final.json",
    orient="records",
    force_ascii=False,
    indent=2
)
```

In [26]:
```
dataset = pd.read_json(r"dataset_llm_final.json")
```

O dataset está desiquilibrado

In [27]:
```
dataset["categoria"].value_counts()
```

Out[27]:
```
categoria
Outros                 10587
NOTA                     607
Referencial               96
Plano                     56
manual                    51
Regulamentos e Leis       46
Publicação                36
NEP                       27
LISTA                     26
Guia                      25
Norma                     23
Diretiva                  21
Folheto                   15
Name: count, dtype: int64
```

Criação dos chunks finais

In [30]:
```
dataset = processar_pasta_pdfs(pasta_pdfs)
```

```
 78%|███████▊  | 9046/11616 [47:47<34:40,  1.24it/s]  Could not get FontBBox fr
om font descriptor because None cannot be parsed as 4 floats
 97%|█████████▋| 11297/11616 [1:27:35<08:01,  1.51s/it]  Could not get FontBBox
from font descriptor because None cannot be parsed as 4 floats
 97%|█████████▋| 11301/11616 [1:27:44<13:24,  2.55s/it]Could not get FontBBox fr
om font descriptor because None cannot be parsed as 4 floats
100%|██████████| 11616/11616 [1:48:22<00:00,  1.79it/s]
```

In [31]:
```
dataset["chunks"] = dataset["texto_limpo"].apply(lambda x: chunk_texto_inteligen
```

Ora voltamos a ver quantos pdfs temos com chunks=0, e vemos que ainda assim haviam 141 que mesmo depois do OCR ficaram m

tam_texto = 0 → OCR/extract falhou mesmo

tam_texto > 0 mas num_chunks = 0 → o problema pode estar no chunking ou na limpeza

In [35]:
```
docs_restantes = dataset[dataset["num_chunks"] == 0].copy()
docs_restantes[["nome_pdf", "caminho"]].head(20)
```

Out[35]:

	nome_pdf	caminho
2101	849-D - 2120001806_signed.pdf	D:\Trabalhos\Bot Exército\pdfs\Financas\pce\pm...
7987	12_1_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\12_1_ocr.pdf
8031	2019 Lei 36_2019 - revoga DLs 1975 a 1980_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\2019 Lei 36...
8045	2120012050_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\2120012050_...
8402	4206_Mapa Incumprimentos_ABR2023_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\4206_Mapa I...
8760	Anexo D (Folha de Caixa)_ABR23_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\Anexo D (Fo...
8926	Flyer_sinais de CAD internet e videojogos_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\Flyer_sinai...
9000	NIF Exercito_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\NIF Exercit...
9034	OfertaFormativa - ISG_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\OfertaForma...
9080	Protocolo - Portinsurance_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\Protocolo -...
9108	Saldos Conta SIG Dez24_1110040013_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\Saldos Cont...
9109	Saldos Conta SIG Dez24_1110040013_signed_signe...	D:\Trabalhos\Bot Exército\pdfs\OCR\Saldos Cont...
9110	Saldos Conta SIG Dez24_1110044306_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\Saldos Cont...
9111	Saldos Conta SIG Dez24_1110044306_signed_signe...	D:\Trabalhos\Bot Exército\pdfs\OCR\Saldos Cont...
9112	Saldos Conta SIG Dez24_1210040335_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\Saldos Cont...
9113	Saldos Conta SIG Dez24_1210040335_signed_signe...	D:\Trabalhos\Bot Exército\pdfs\OCR\Saldos Cont...
9115	Teste 2paginas_ocr.pdf	D:\Trabalhos\Bot Exército\pdfs\OCR\Teste 2pagi...
9990	LISTAS DE ANTIGUIDADE DE 1963 Oficiais.pdf	D:\Trabalhos\Bot Exército\pdfs\pessoal\AdmRH\g...
11597	GSST_Relatorio anual-estatisticas2017+Despacho...	D:\Trabalhos\Bot Exército\pdfs\pessoal\SegSaud...

In [36]:
```
docs_restantes["tam_texto"] = docs_restantes["texto_limpo"].fillna("").str.len()
docs_restantes[["nome_pdf", "tam_texto", "num_chunks"]].sort_values("tam_texto")
```

Out[36]:

	nome_pdf	tam_texto	num_chunks
2101	849-D - 2120001806_signed.pdf	0	0
7987	12_1_ocr.pdf	0	0
8031	2019 Lei 36_2019 - revoga DLs 1975 a 1980_ocr.pdf	0	0
8045	2120012050_ocr.pdf	0	0
8402	4206_Mapa Incumprimentos_ABR2023_ocr.pdf	0	0
8760	Anexo D (Folha de Caixa)_ABR23_ocr.pdf	0	0
8926	Flyer_sinais de CAD internet e videojogos_ocr.pdf	0	0
9000	NIF Exercito_ocr.pdf	0	0
9034	OfertaFormativa - ISG_ocr.pdf	0	0
9080	Protocolo - Portinsurance_ocr.pdf	0	0
9108	Saldos Conta SIG Dez24_1110040013_ocr.pdf	0	0
9109	Saldos Conta SIG Dez24_1110040013_signed_signe...	0	0
9110	Saldos Conta SIG Dez24_1110044306_ocr.pdf	0	0
9111	Saldos Conta SIG Dez24_1110044306_signed_signe...	0	0
9112	Saldos Conta SIG Dez24_1210040335_ocr.pdf	0	0
9113	Saldos Conta SIG Dez24_1210040335_signed_signe...	0	0
9115	Teste 2paginas_ocr.pdf	0	0
9990	LISTAS DE ANTIGUIDADE DE 1963 Oficiais.pdf	0	0
11597	GSST_Relatorio anual-estatisticas2017+Despacho...	0	0

In [38]:
```
dataset = dataset[dataset["num_chunks"] > 0].reset_index(drop=True)
(dataset["num_chunks"] == 0).sum()
```

Out[38]:
```
np.int64(0)
```

Aqui um exemplo para ver se a query está a procura dos documentos certos

In [45]:
```
dataset_chunks = criar_dataset_chunks(dataset)
dataset_chunks.head()
```

Out[45]:

	nome_pdf	caminho	categoria	chunk_id	texto_chunk
0	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	0	ESTADO-MAIOR-GENERAL DAS FORÇAS ARMADAS CENTRO...
1	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	1	janeiro de 2018, que deverá ser destruído. CHE...
2	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	2	DAS REGISTO DE ALTERACÕES J Identificação ou c...
3	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	3	DOS 25 1. Visitas Organizadas 25 2. Plano Anua...
4	GUIA DOS ADIDOS DE DEFESA E MILITARES ACREDITA...	D:\Trabalhos\Bot Exército\pdfs\comando-EstadoM...	Outros	4	a A-1 ANEXO B Ficha de Dados Biográficos B-1 A...

In [46]:
```
modelo_embeddings = SentenceTransformer("sentence-transformers/paraphrase-multil
```

```
Loading weights: 100%|██████████| 199/199 [00:00<00:00, 15670.66it/s]
BertModel LOAD REPORT from: sentence-transformers/paraphrase-multilingual-MiniLM-
L12-v2
Key                     | Status     |  |
------------------------+------------+--+-
embeddings.position_ids | UNEXPECTED |  |

Notes:
- UNEXPECTED    :can be ignored when loading from different task/architecture; no
t ok if you expect identical arch.
```

In [49]:
```
embeddings = modelo_embeddings.encode(
    dataset_chunks["texto_chunk"].tolist(),
    show_progress_bar=True,
    convert_to_numpy=True
)
```

```
Batches: 100%|██████████| 7352/7352 [42:05<00:00,  2.91it/s]
```

In [52]:
```
dimensao = embeddings.shape[1]

index = faiss.IndexFlatL2(dimensao)
index.add(np.array(embeddings, dtype="float32"))

print("Número de chunks indexados:", index.ntotal)
```

```
Número de chunks indexados: 235262
```

In [53]:
```
resultados = pesquisar_chunks(
    query="Quais são as normas sobre segurança da informação?",
    modelo_embeddings=modelo_embeddings,
    index=index,
    dataset_chunks=dataset_chunks,
    top_k=5
)

resultados[["nome_pdf", "categoria", "texto_chunk", "score_distancia"]]
```

Out[53]:

	nome_pdf	categoria	texto_chunk	score_distancia
36816	PAD 230-01 RGSUEOE SET22.pdf	Outros	ança do Pessoal; (3) Segurança física; (4) Seg...	7.508637
36818	PAD 230-01 RGSUEOE SET22.pdf	Outros	o e do acesso à Informação classificada, em es...	9.730995
204497	Protocolo - Biwatt Green.pdf	Outros	a, classificada pela entidade emissora de conf...	9.759338
39543	Regulamento (UE) 2016_679 do PE e do Conselho,...	Outros	porcionada para assegurar a segurança da rede ...	10.058803
204600	Protocolo - CRIAP 2023.pdf	Outros	cia trocada, classificada pela entidade emisso...	10.598384

Qwen

In [57]:
```
query = "Quais são os procedimentos previstos na NEP para segurança documental?"
resultados = pesquisar_chunks(query, modelo_embeddings, index, dataset_chunks, t
prompt_final = construir_contexto_rag(query, resultados)
```

In [58]:
```
print(prompt_final)
```

Responde à pergunta do utilizador com base apenas no contexto fornecido.

Pergunta:
Quais são os procedimentos previstos na NEP para segurança documental?

Contexto:
[Documento: E0199A1P_Esp 012 - Sapador de Engenharia.pdf | Categoria: Outros]
as graves de procedimentos de segurança.

[Documento: DIRECTIVA31MAR08.pdf | Categoria: Outros]
rementar a utilização da aplicação de gestão documental de modo a tirar melhor pa rtido das suas capacidades; (b) Propor acções com vista à motivação pessoal, para o cumprimento dos objectivos da DFin;

(c) Efectuar o controlo da assiduidade do pessoal civil tirando partido das capac idades da aplicação de “controlo de ponto”. g. Oficial de Segurança (1) Quanto à segurança das informações (a) Adoptar todas as medidas relativas à protecção das matérias classificadas e conduzir as investigações de segurança que forem necessá rias; (b) Aconselhar o Director sobre assuntos relativos à segurança e protecção das matérias classificadas; (c) Montar a segurança das matérias classificadas; (d) Propor a correcção de erros de atribuição da classificação de segurança; (e) Organizar programas de destruição de rotina de documentos desactualizados e sem i nteresse histórico, por forma a evitar a sobrecarga inútil dos arquivos; (f) Supe rvisionar a destruição de rascunhos, minutas, papeis químicos, cópias excedentári as e demais desperdícios, por forma a assegurar que através deles não ocorram que bras de segurança; (g) Planear a destruição de emergência de todas as matérias cl assificadas; (h) Efectuar inspecções de segurança com vista a verificar se são cu mpridas as normas de segurança em vigor.

[Documento: Diretiva 2008 - 31Mar08.pdf | Categoria: Outros]
rementar a utilização da aplicação de gestão documental de modo a tirar melhor pa rtido das suas capacidades; (b) Propor acções com vista à motivação pessoal, para o cumprimento dos objectivos da DFin;

(c) Efectuar o controlo da assiduidade do pessoal civil tirando partido das capac idades da aplicação de “controlo de ponto”. g. Oficial de Segurança (1) Quanto à segurança das informações (a) Adoptar todas as medidas relativas à protecção das matérias classificadas e conduzir as investigações de segurança que forem necessá rias; (b) Aconselhar o Director sobre assuntos relativos à segurança e protecção das matérias classificadas; (c) Montar a segurança das matérias classificadas; (d) Propor a correcção de erros de atribuição da classificação de segurança; (e) Organizar programas de destruição de rotina de documentos desactualizados e sem i nteresse histórico, por forma a evitar a sobrecarga inútil dos arquivos; (f) Supe rvisionar a destruição de rascunhos, minutas, papeis químicos, cópias excedentári as e demais desperdícios, por forma a assegurar que através deles não ocorram que bras de segurança; (g) Planear a destruição de emergência de todas as matérias cl assificadas; (h) Efectuar inspecções de segurança com vista a verificar se são cu mpridas as normas de segurança em vigor.

[Documento: PDE 4-46-00.pdf | Categoria: Outros]
ma casuística e não integrada, Diretivas Técnicas e Normas de Execução Permanente (NEP), como forma de dar resposta aos problemas e desafios específicos que vão su rgindo, de que tem resultado a instituição de procedimentos ajustados e facilitad ores do cumprimento da missão do CmdLog. XI

```
---

[Documento: PASSTPA-2024_BrigInt_ocr.pdf | Categoria: Outros]
ensibilização dos executantes para a sua finalidade face aos perigos/riscos, de a
cordo com o método preconizado na PAD 123-01. i. Requlamentos, Manuais e Normas d
e Segurança (1) Promover a divulgação das NEP da Briglnt; (2) Promover a revisão
dos PEI e sua difusão. j. Relato, investigação e registo de Incidentes / Acidente
s (1) Manter atualizada, em Ordem de Serviço das Unidades, a relação dos elemento
s que constituem as Comissões de Investigação de Acidentes, quanto aos Fatores Hu
manos, Materiais e Ambientais. Por norma, o Presidente é o Comandante da Unidade,
o Oficial Coordenador é o Oficial de Segurança e Saúde no Trabalho e o

Instruções:
- Se a resposta não estiver claramente no contexto, diz que não encontraste infor
mação suficiente.
- Resume de forma objetiva.
- Cita o nome do documento quando relevante.
```

In [63]:
```python
perguntas_teste = [
    "Quais são as normas sobre segurança da informação?",
    "O que é uma NEP?",
    "Que documento fala sobre redes sociais?",
    "Quais são os procedimentos do interrogatório?",
    "Que regras existem sobre administração de rede local?"
]
```

Gualdar o modelo de embeddins

In [ ]:
```python
np.save("embeddings.npy", embeddings)
```

In [65]:
```python
testar_perguntas(
    perguntas_teste,
    modelo_embeddings,
    index,
    dataset_chunks,
    top_k=5
)
```

==========
==========
🔎 PERGUNTA: Quais são as normas sobre segurança da informação?
==========
==========

Resultado 36817
Documento: PAD 230-01 RGSUEOE SET22.pdf
Categoria: Outros
Score: 7.508636951446533
Texto:
ança do Pessoal; (3) Segurança física; (4) Segurança das CSI. d. As medidas da se
gurança da informação situam-se no âmbito da segurança da informação classificad
a, no tratamento e manuseamento das mesmas e englobam: (1) Credenciação adequada
à função e válida; (2) Classificação de segurança da informação; (3) Classificaçã
o do documento; (4) Preparação do documento classificado; (5) Reclassificação
---

Resultado 36819
Documento: PAD 230-01 RGSUEOE SET22.pdf
Categoria: Outros
Score: 9.730995178222656
Texto:
o e do acesso à Informação classificada, em estrita observância do princípio da
“necessidade de conhecer”. i. A segurança física deve ser implementada segundo o
princípio da defesa em profundidade, recorrendo a um conjunto de medidas compleme
ntares nas restantes áreas da segurança. j. Para além da implementação dos concei
tos prescritos adaptados às condições específicas da U/E/O e do meio onde se
---

Resultado 204498
Documento: Protocolo - Biwatt Green.pdf
Categoria: Outros
Score: 9.75933837890625
Texto:
a, classificada pela entidade emissora de confidencial, na mais estrita confidenc
ialidade, bem como a observar o regime legal da proteção de dados pessoais. 2. Ao
abrigo do disposto no númeroanterior, os OUTORGANTES obrigam-se, designadamente:

Resultado 39544
Documento: Regulamento (UE) 2016_679 do PE e do Conselho, de 27 de abril de 2016,
RGPD.pdf
Categoria: Outros
Score: 10.058802604675293
Texto:
porcionada para assegurar a segurança da rede e das informações, ou seja, a capac
idade de uma rede ou de um sistema informático de resistir, com um dado nível de
confiança, a eventos acidentais ou a ações maliciosas ou ilícitas que comprometam
a disponibi lidade, a autenticidade, a integridade e a confidencialidade dos dado
s pessoais conservados ou transmitidos, bem como a segurança dos serviços
---

Resultado 204601
Documento: Protocolo - CRIAP 2023.pdf
Categoria: Outros
Score: 10.598383903503418
Texto:
cia trocada, classificada pela entidade emissora de confidencial, na mais estrita
confidencialidade, bem como a observar o regime legal da proteção de dados pessoa

is. 2. Ao abrigo do disposto no númeroanterior, os OUTORGANTES obrigam-se, design
adamente:

==========
🔎 PERGUNTA: O que é uma NEP?
==========

Resultado 11402
Documento: RA2012.pdf
Categoria: Outros
Score: 13.04309368133545
Texto:
nto;

Resultado 65681
Documento: NEP AGE.122_ocr.pdf
Categoria: Outros
Score: 15.137784004211426
Texto:
.o 231/2009, de 15 setembro, que estabelece que o Comando de Pessoal é Orgão Cent
ral de Administração e Direção (OCAD), cabendo-lhe assegurar a superintendência e
a execução em áreas ou atividades específicas essenciais. b. A presente NEP aplic
a-se a todos os trabalhadores da área da saúde, vinculados em regime de trabalho
em funções públicas e integrados nas respetivas carreiras, que pretendam a
---

Resultado 43805
Documento: PDE 4-46-00.pdf
Categoria: Outros
Score: 15.832741737365723
Texto:
ma casuística e não integrada, Diretivas Técnicas e Normas de Execução Permanente
(NEP), como forma de dar resposta aos problemas e desafios específicos que vão su
rgindo, de que tem resultado a instituição de procedimentos ajustados e facilitad
ores do cumprimento da missão do CmdLog. XI

Resultado 1828
Documento: PDE 0-19-00_SINAIS CONVENCIONAIS MILITARES.pdf
Categoria: Outros
Score: 16.162994384765625
Texto:
Outros Símbolos e Gráficos TAREFA A/E GRÁFICO NOTAS Neutralizar (Neutralize) (É u
ma tarefa táctica que se traduz na N incapacidade do pessoal ou material In poder
interferir com determinada operação das forças amigas. O comandante deve especifi
car qual a E Ver: N unidade ou material a neutralizar e a Suprimir (Supress); dur
ação da neutralização. Desorganizar (Disrupt); No apoio de fogos representa
---

Resultado 65696
Documento: NEP AGE.123_ocr.pdf
Categoria: Outros
Score: 16.30051040649414
Texto:
AGE.123 SUPLEMENTOS REMUNERATÓRIOS – TRABALHO EXTRAORDINÁRIO PAG. 2 despendidos c
om este suplemento; b. Assim, com a presente NEP, pretende-se compilar e difundir

informação legislativa no âmbito da prestação de trabalho extraordinário, dispers
a por vários normativos e, em simultâneo, regulamentar e implementar formalidade/
procedimentos que permitam ao comando do Pessoal cumprir com o desidrato re
---

==========
🔎 PERGUNTA: Que documento fala sobre redes sociais?
==========

Resultado 204
Documento: PEMGFA_GER_006_GUIA DO ADIDO DE DEFESA.pdf
Categoria: Outros
Score: 10.005367279052734
Texto:
PEMGFA/GER 006 CAPÍTULO 12 COMUNICAÇÃO INSTITUCIONAL 1201. REDES SOCIAIS Atualmen
te as redes sociais fazem parte do dia-a-dia das pessoas e das organizações em to
do o planeta. Com o crescimento do Facebook, Instagram, Twitter, etc permitindo a
cessos por computadores e também por tecnologias móveis como telefones, tablets,
IPhones, etc é pois uma incontornável realidade. Para que possamos capitaliz
---

Resultado 232228
Documento: Livro_VDS2018_definitivo_RCAN.pdf
Categoria: Outros
Score: 10.095380783081055
Texto:
Moqbel, 2012). Segundo Ellisson & Boyd (2007), as redes sociais são numa ferramen
ta pertencente à Internet que permite aos utilizadores estabelecer relações, cons
truir um perfil público e criar e desenvolver conexões com outros utilizadores. A
ssim, as redes sociais como Facebook, Instagram, WhatsApp e LinkedIn fazem hoje p
arte do quotidiano de grande parte da população, transformando-se em platafo
---

Resultado 13432
Documento: RA2022.pdf
Categoria: Outros
Score: 11.656988143920898
Texto:
 Seguidores das Redes Sociais do Recrutamento do Exército, que tiveram igualment
e um crescimento contínuo do número de seguidores ao longo de todo o ano. Figura
31 - Evolução mensal do número de “seguidores” das redes sociais do recrutamento
do Exército, em 2022 Peças jornalísticas sobre o Exército nos Órgãos de Comunic
ação Social, no âmbito do incremento da relevância do Exército no ambiente m
---

Resultado 232229
Documento: Livro_VDS2018_definitivo_RCAN.pdf
Categoria: Outros
Score: 12.249822616577148
Texto:
cooperativa com os clientes. As Organizações passaram a recorrer às mesmas como f
orma de promover os seus produtos e serviços (Drury, 2008). A utilização das rede
s sociais em contexto organizacional proporciona diversos benefícios para as Orga
nizações, nomeadamente o facto de facilitar a interação e troca de informações co
m clientes e potenciais clientes (Ozuem, Howell & Lancaster, 2008), bem como
---

Resultado 8738

Documento: PA2023.pdf
Categoria: Outros
Score: 12.47812557220459
Texto:
a divulgação do Serviço Militar em Regime de Voluntariado e Regime de Contrato (R V/RC) em Televisão, Cinemas, outdoors e revistas on-line. Comando do Pessoal Refo rçar a mensagem transmitida pelo Recrutamento do Exército nas Redes Sociais atrav és da criação de diversas campanhas no Facebook. Figura 15 – Publicidade instituc ional
---

==========
🔎 PERGUNTA: Quais são os procedimentos do interrogatório?
==========

Resultado 220461
Documento: E0202A1P_Esp 028 - I Armas Pesadas Morteiros CFS.pdf
Categoria: Outros
Score: 10.34010124206543
Texto:
o dos procedimentos - Supervisão

Resultado 210171
Documento: MD 240-03_Modelo de Referencial de Curso.pdf
Categoria: Outros
Score: 10.960897445678711
Texto:
de recolha e análise de dados. (a) O procedimento metodológico, explica os passos inerentes aos processos de avaliação descrevendo quando e como se vai realizar a avaliação. A metodologia e instrumentos de recolha da informação nas fontes acima indicadas podem ser: 1 Testes. 2 Circuitos de Avaliação. 3 Observação direta. 4 Q uestionários. 5 Relatórios. 6-3

Resultado 225621
Documento: PTE 005 01 Auditoria de Abonos e Descontos.pdf
Categoria: Outros
Score: 11.076339721679688
Texto:
CAPÍTULO 3 – O PROCESSO DE AUDITORIA 301. Fases do processo de auditoria O Proces so de auditoria contempla as quatro fases a seguir apresentadas. FASES DA AUDITOR IA Compreensão dos processos, das entidades intervenientes e do meio Planeamento envolvente através do desenvolvimento do Estudo Preliminar. Avaliação do risco e elaboração do Programa de Auditoria. Avaliação do controlo interno e dos pro
---

Resultado 226501
Documento: PTE 005 01 Auditoria de Abonos e Descontos.pdf
Categoria: Outros
Score: 11.076339721679688
Texto:
CAPÍTULO 3 – O PROCESSO DE AUDITORIA 301. Fases do processo de auditoria O Proces so de auditoria contempla as quatro fases a seguir apresentadas. FASES DA AUDITOR IA Compreensão dos processos, das entidades intervenientes e do meio Planeamento envolvente através do desenvolvimento do Estudo Preliminar. Avaliação do risco e elaboração do Programa de Auditoria. Avaliação do controlo interno e dos pro
---

Resultado 220893
Documento: E0215A1P_Esp 299 - C Morteiros CFO.pdf
Categoria: Outros
Score: 11.142557144165039
Texto:
cedimento O procedimento metodológico descreve quando e como se realiza a avaliaç
ão do curso e decorre em 2 fases: (1) Avaliação formativa, a fim de ajustar o des
envolvimento da formação às necessidades dos formandos; (2) Avaliação sumativa, a
fim de aferir o resultado final da formação. b. Técnicas (1) Avaliação formativa:
questões objetivas, registo de observação, revisão após ação;

==========
🔎 PERGUNTA: Que regras existem sobre administração de rede local?
==========

Resultado 36778
Documento: PAD 230-01 RGSUEOE SET22.pdf
Categoria: Outros
Score: 10.024639129638672
Texto:
tação das NAT nesta área, emanadas pela respetiva entidade competente. 356. Compe
tências do administrador de Rede local a. A tarefa principal do administrador de
Rede local é gerir os sistemas informáticos locais, de acordo com as Normas em vi
gor no Exército e em estreita coordenação com a entidade competente do Exército p
ara esta matéria. b. É dever do administrador de Rede local cumprir e divulg
---

Resultado 45914
Documento: (REVOGADA) NAT 02.03.04 - Normas Tecnicas e procedimentos do Admin da
Rede Local das UEO14JAN2014_ocr.pdf
Categoria: Outros
Score: 10.039144515991211
Texto:
istentes na U/E/O e fornece-lo ao CTE, sempre que solicitado ou em caso de altera
ções; (c) No processo de implementação de pontos de acesso wireless, regulamentad
os em Norma Técnica própria (NEP/NAT 02.03.06), é responsabilidade do Administrad
or Local: 1. Solicitar autorização e parecer técnico à DCSI; 2. Se autorizada, ga
rantir que se cumprem as normas de implementação e exploração de pontos de a
---

Resultado 36779
Documento: PAD 230-01 RGSUEOE SET22.pdf
Categoria: Outros
Score: 10.354473114013672
Texto:
Serviço nas U/E/O de relato de incidentes. c. No âmbito da infraestrutura da Red
e, constituem responsabilidades do administrador da Rede local acompanhar o plane
amento e a execução de trabalhos de expansão/melhoramento da Rede local, manter u
m cadastro atualizado das ligações físicas e lógicas de todos os equipamentos de
Rede existentes na U/E/O e colaborar com a entidade técnica gestora da Rede d
---

Resultado 45908
Documento: (REVOGADA) NAT 02.03.04 - Normas Tecnicas e procedimentos do Admin da
Rede Local das UEO14JAN2014_ocr.pdf
Categoria: Outros

```
Score: 10.440067291259766
Texto:
Z LL DCSI NAT 02.03.04 | Pág. .2de58Pág. | ÂMBITO A presente norma destina-se a s
er do conhecimento de todos os Administradores de Rede Local e outros elementos q
ue, por força da necessidade ou por nomeação superior, tenham necessidade de exec
utar tarefas e aceder, quer de forma física quer lógica, às redes e sistemas do S
IC-Op, presentes na sua U/E/O, SITUAÇÃO a. O alto nível de informatização do
--------------------------------------------------------------------------------

Resultado 45912
Documento: (REVOGADA) NAT 02.03.04 - Normas Tecnicas e procedimentos do Admin da
Rede Local das UEO14JAN2014_ocr.pdf
Categoria: Outros
Score: 10.764945983886719
Texto:
Local (1) É da responsabilidade do Comandante/Diretor/Chefe da U/E/O a nomeação d
o Administrador de Rede Local e, eventualmente, um adjunto; (2) A nomeação deverá
recair sobre graduados, sempre que possível do QP, habilitados com conhecimentos
técnicos de administração e segurança de redes, por forma a
--------------------------------------------------------------------------------
```

In [66]:
```python
for pergunta in perguntas_teste:
    print("\n" + "="*100)
    print("PERGUNTA:", pergunta)

    prompt, resultados = preparar_rag(
        query=pergunta,
        modelo_embeddings=modelo_embeddings,
        index=index,
        dataset_chunks=dataset_chunks,
        top_k=5
    )

    print("\n--- PROMPT PARA O QWEN ---\n")
    print(prompt[:1500])
```

```
================================================================================
==================
PERGUNTA: Quais são as normas sobre segurança da informação?

--- PROMPT PARA O QWEN ---


Responde à pergunta com base apenas no contexto fornecido.

Pergunta:
Quais são as normas sobre segurança da informação?

Contexto:
[Documento: PAD 230-01 RGSUEOE SET22.pdf | Categoria: Outros]
ança do Pessoal; (3) Segurança física; (4) Segurança das CSI. d. As medidas da se
gurança da informação situam-se no âmbito da segurança da informação classificad
a, no tratamento e manuseamento das mesmas e englobam: (1) Credenciação adequada
à função e válida; (2) Classificação de segurança da informação; (3) Classificaçã
o do documento; (4) Preparação do documento classificado; (5) Reclassificação e d
esclassificação do documento; (6) Reprodução do documento classificado; (7) Trans
ferência da informação classificada; (8) Registo e controlo de segurança da infor
mação classificada; (9) Destruição e evacuação da informação classificada. e. A s
egurança do Pessoal traduz-se, de acordo com as ISM, no acesso regular e no acess
o excecional a informação classificada, nomeadamente: (1) Na credenciação do Pess
oal nas várias Marcas e Graus; (2) No acompanhamento de alterações de atitude ou
comportamentos suspeitos que ponham em causa a idoneidade do Pessoal;

---

[Documento: PAD 230-01 RGSUEOE SET22.pdf | Categoria: Outros]
o e do acesso à Informação classificada, em estrita observância do princípio da
“necessidade de conhecer”. i. A segurança física deve ser implementada segundo o
princípio da defesa em profundidade, recorrendo a um conjunto de medidas compleme
ntares nas restantes áreas da segur

================================================================================
==================
PERGUNTA: O que é uma NEP?

--- PROMPT PARA O QWEN ---


Responde à pergunta com base apenas no contexto fornecido.

Pergunta:
O que é uma NEP?

Contexto:
[Documento: RA2012.pdf | Categoria: Outros]
nto;

---

[Documento: NEP AGE.122_ocr.pdf | Categoria: Outros]
.o 231/2009, de 15 setembro, que estabelece que o Comando de Pessoal é Orgão Cent
ral de Administração e Direção (OCAD), cabendo-lhe assegurar a superintendência e
a execução em áreas ou atividades específicas essenciais. b. A presente NEP aplic
a-se a todos os trabalhadores da área da saúde, vinculados em regime de trabalho
em funções públicas e integrados nas respetivas carreiras, que pretendam a sua ac
tualização ou aperfeiçoamento profissional. 3. INTRODUÇÃO a. É facultado aos trab
```

alhadores da área de saúde a frequência de ações de formação complementar específica da respetiva área profissional. Essas formações podem incluir participações em Cursos, Congressos, Seminários, Encontros, Estágios, Jornadas, Simpósios e outras atividades similares. b. Tais eventos encontram cobertura legal nos seguintes diplomas: (1) Decreto-Lei no 177/2009 de 04Ago, estabelece o regime da carreira especial médica, aplicável aos médicos cuja relação jurídica de emprego público seja constituída por contrato de trabalho em funções públicas, no no 3 do Art.o 25o, permite a frequência de cursos de formação complementar ou de actualização profissional, com vista ao aperfeiçoamento, que pode ser autorizado m

---

[Documento: PDE 4-46-00.pdf | Categoria: Outros]
ma casuística e não integrada, Direti

==========
PERGUNTA: Que documento fala sobre redes sociais?

--- PROMPT PARA O QWEN ---

Responde à pergunta com base apenas no contexto fornecido.

Pergunta:
Que documento fala sobre redes sociais?

Contexto:
[Documento: PEMGFA_GER_006_GUIA DO ADIDO DE DEFESA.pdf | Categoria: Outros]
PEMGFA/GER 006 CAPÍTULO 12 COMUNICAÇÃO INSTITUCIONAL 1201. REDES SOCIAIS Atualmente as redes sociais fazem parte do dia-a-dia das pessoas e das organizações em todo o planeta. Com o crescimento do Facebook, Instagram, Twitter, etc permitindo acessos por computadores e também por tecnologias móveis como telefones, tablets, IPhones, etc é pois uma incontornável realidade. Para que possamos capitalizar no uso das redes sociais, no que se refere às atividades dos militares das Forças Armadas Portuguesas nas mais diversas missões, os Adidos de Defesa devem enviar ao GLADM propostas de publicações, desejavelmente com fotografias, sempre que as circunstâncias forem consideradas relevantes, oportunas ou convenientes, designadamente, artigos alusivos às atividades protocolares, seminários, visitas de altas entidades nacionais e estrangeiras. 1202. USO DO CORREIO ELETRÓNICO a. Nota introdutória A melhor forma de comunicar com alguém é presencialmente. Mais de 70% da mensagem que pretendemos transmitir advém da nossa linguagem não-verbal. Os restantes 30% são naturalmente provenientes da nossa expressão oral (tom, intensidade e volume). Contudo, a comunicação escrita, em detrimento de um enco

---

[Documento: Livro_VDS2018_definitivo_RCAN.pdf | Categoria: Outros]
Moqbel, 2012). Segundo Ellis

==========
PERGUNTA: Quais são os procedimentos do interrogatório?

--- PROMPT PARA O QWEN ---

Responde à pergunta com base apenas no contexto fornecido.

Pergunta:

Quais são os procedimentos do interrogatório?

Contexto:
[Documento: E0202A1P_Esp 028 - I Armas Pesadas Morteiros CFS.pdf | Categoria: Outros]
o dos procedimentos - Supervisão

[Documento: MD 240-03_Modelo de Referencial de Curso.pdf | Categoria: Outros]
de recolha e análise de dados. (a) O procedimento metodológico, explica os passos inerentes aos processos de avaliação descrevendo quando e como se vai realizar a avaliação. A metodologia e instrumentos de recolha da informação nas fontes acima indicadas podem ser: 1 Testes. 2 Circuitos de Avaliação. 3 Observação direta. 4 Questionários. 5 Relatórios. 6-3

[Documento: PTE 005 01 Auditoria de Abonos e Descontos.pdf | Categoria: Outros]
CAPÍTULO 3 – O PROCESSO DE AUDITORIA 301. Fases do processo de auditoria O Processo de auditoria contempla as quatro fases a seguir apresentadas. FASES DA AUDITORIA Compreensão dos processos, das entidades intervenientes e do meio envolvente através do desenvolvimento do Estudo Preliminar. Avaliação do risco e elaboração do Programa de Auditoria. Avaliação do controlo interno e dos procedimentos, obtenção e Execução avaliação das evidências de auditoria e estabelecimento das Observações e Conclusões Preliminares. Escrituração das observações e opiniões, formulação de Relato recomendações e elaboração do Relatório de Auditoria. Acompanhamento do acolhimento das recomendações e Seguimento eventualmente a r

==========
PERGUNTA: Que regras existem sobre administração de rede local?

--- PROMPT PARA O QWEN ---

Responde à pergunta com base apenas no contexto fornecido.

Pergunta:
Que regras existem sobre administração de rede local?

Contexto:
[Documento: PAD 230-01 RGSUEOE SET22.pdf | Categoria: Outros]
tação das NAT nesta área, emanadas pela respetiva entidade competente. 356. Competências do administrador de Rede local a. A tarefa principal do administrador de Rede local é gerir os sistemas informáticos locais, de acordo com as Normas em vigor no Exército e em estreita coordenação com a entidade competente do Exército para esta matéria. b. É dever do administrador de Rede local cumprir e divulgar as Normas técnicas bem como sensibilizar de forma permanente os utilizadores para os aspetos de segurança da informação e de operação dos meios e sistemas de informação: cuidados com o equipamento, manuseamento de matérias classificadas, políticas de segurança, sistema

---

[Documento: (REVOGADA) NAT 02.03.04 - Normas Tecnicas e procedimentos do Admin da Rede Local das UEO14JAN2014_ocr.pdf | Categoria: Outros]
istentes na U/E/O e fornece-lo ao CTE, sempre que solicitado ou em caso de alterações; (c) No processo de implementação de pontos de acesso wireless, regulamentados em Norma Técnica própria (NEP/NAT 02.03.06), é responsabilidade do Administrad

```
or Local: 1. Solicitar autorização e parecer técnico à DCSI; 2. Se autorizada, ga
rantir que se cumprem as normas de implementação e exploração de pontos de acesso
wireless, (3) No âmbito da configuração de equipamentos ligados ao domínio exe
```

Meter meter os ficheiros na indice_documentos.faiss dataset_chunks.json mesma pasta do app.py (que vai funcionar como backend do nosso "Chat"), no meu caso eu crei um pasta denominada Modelos_bot_exercito mas o idela é estar na mesma pasta