

FICHEIRO DE ACOMPANHAMENTO À DESCRIÇÃO DE DADOS DE INVESTIGAÇÃO
GRELHA DE INFORMAÇÃO

HANDLE	https://hdl.handle.net/10071/38671									
DESIGNAÇÃO	Ciberdúvidas da Língua Portuguesa: European Portuguese Linguistics Paraphrase Benchmark - 1000-Question Scaled Evaluation Set (5000 Queries)									
AUTOR(ES)	Moura, P., Batista, F., Lopes, A.									
DATA DE CRIAÇÃO	2026 Sep 17									
DESIGNAÇÃO DO PROJETO	This dataset contains a large scale paraphrase-based query benchmark built from 1000 original questions randomly sampled from Ciberdúvidas da Língua Portuguesa, an expert-curated Portuguese-language consultation service (https://ciberduvidas.iscte-iul.pt). For each of the 1000 original questions, five paraphrased reformulations were generated using the mistral-large-2512 model via the Mistral API, according to five distinct query profiles designed to simulate the range of phrasings a real user might submit: - Synthetic: short, direct, search-engine style queries - Formal: grammatically rigorous and highly detailed queries - Informal: casual, conversational tone queries - Professor: queries using technical pedagogical/linguistic terminology - Student: direct, comprehension-focused queries									
LICENÇA	CC-By Attribution 4.0 International									
DOI	https://doi.org/10.5281/zenodo.22819697									
FICHEIROS	Files (1.6 MB) <table><tr><th>Name</th><th>Size</th><th></th></tr><tr><td>ciberduvidas_5000_paraphrases.json</td><td>1.6 MB</td><td>Download all</td></tr><tr><td colspan="3">md5:48a6fbd282380d29ea1b9a49de4815db</td></tr></table>	Name	Size		ciberduvidas_5000_paraphrases.json	1.6 MB	Download all	md5:48a6fbd282380d29ea1b9a49de4815db		
Name	Size									
ciberduvidas_5000_paraphrases.json	1.6 MB	Download all								
md5:48a6fbd282380d29ea1b9a49de4815db										