DEEP

NLP in DEEP

DEEP is a groundbreaking tool that utilizes Natural Language Processing (NLP) technology for supporting, categorizing, and analyzing texts to generate reports in a more efficient manner. DEEP is pioneering the practical application of Artificial Intelligence in the humanitarian sector, and it paves the way for the use of NLP technology in this field.

DEEP NLP FEATURES:

Assisted Tagging
A multi-label classification model tailored for common humanitarian analysis frameworks and trained on DEEP data with direct access in DEEP and a pluggable interface for other applications.

Automatic Summarization
Automating the summarization step for creating reports using annotated data.

Topic Modeling
Using it to identify the topics present in a large corpus of text data and for analyzing large volumes of text data and identifying meaningful patterns and insights embedded in the upcoming analysis module.

Automatic Geo-tagging
Optimizing a model that extracts geographic information from the text making it more accurate and appropriate to the humanitarian sector.

Ongoing research
DEEP's potential extends beyond its current features, as it has ongoing research projects, including on AI ethics and further automated features which will make it even more efficient in handling large volumes of text data:

Entry extraction
This model selects a subset of passages that contain relevant information from the given document; these entries do not necessarily follow the common units of text, such as sentences and paragraphs and can appear in various lengths.

Fully automated pipeline

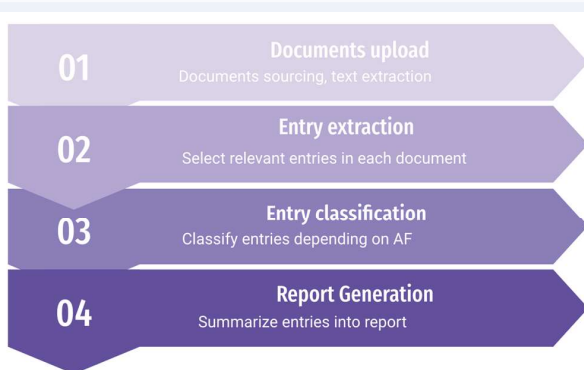|	
01	**Documents upload** — Documents sourcing, text extraction
02	**Entry extraction** — Select relevant entries in each document
03	**Entry classification** — Classify entries depending on AF
04	**Report Generation** — Summarize entries into report

Additional resources:
DEEP offers additional resources by providing important data for training and testing NLP models in the humanitarian context, available in Hugging Face https://huggingface.co/nlp-thedeep:

Humanitaran BERT
HumBert is the first language model adapted to humanitarian topics. It is an XLM-Roberta model trained on humanitarian texts - approximately 50 million textual examples (roughly 2 billion tokens) from public humanitarian reports, law cases and news articles. Although XLM-Roberta was trained in 100 different languages, it is fine-tuned and performs in three languages, English, French and Spanish.

HUMSET
HumSet is a novel and rich multilingual dataset curated by humanitarian analysts and covers various disasters around the globe that occurred from 2018 to 2021. The dataset consists of approximately 17K annotated documents from 46 humanitarian response projects.

If you have questions or would like to contact us, send an email to: nlp@thedeep.io

theDEEP.io