Do LLMs know how to evaluate Asylum applications?

Andrea Galassi, Elena Palmieri, Paolo Torroni

DISI, University of Bologna

Context

- Migration is one of the top current concerns of the European Union (EU)
- EU Commission has set different reforms
  - **Several forms of protections**
  - Application of rules requires a **non-trivial evaluation of legal and factual data** concerning the applicant's living conditions
  - DIFFICULTY for migrants to achieve a **preliminary overview** on their chances to obtain protection

Legal Protections (2024)

- Refugee status
  - persecuted for reasons of race, religion, nationality, membership to a particular social group or political opinion
  - unable to avail himself of the protection of his country
- Subsidiary Protection Status
  - missing some requirements for refugee status
- Special (or Humanitarian) Protection
  - workers, students, children born in EU, …

Research

- Objective: provide a tool to allow migrants to self-assess their condition

- Fundamental aspects:
  - Privacy
  - Explainability
  - Transparency
  - Auditability

Previous Chatbots for Immigration

- Kotiyal et al. : provide a direct answer after each question
- Chen et al. : provide an answer based on a set of FAQ

Kotiyal et al. “Knowledge graph based intelligent conversational agent for uk immigration case work,” in NLIWoD@ESWC, 2022

Chen et al. “Creating a chatbot for and with migrants: Chatbot personality drives co-design activities,” in DIS2020

- Limitation: correct answers require plenty of information
  - A system must **ask for more information** if those given are insufficient
  - The answer must be contextual

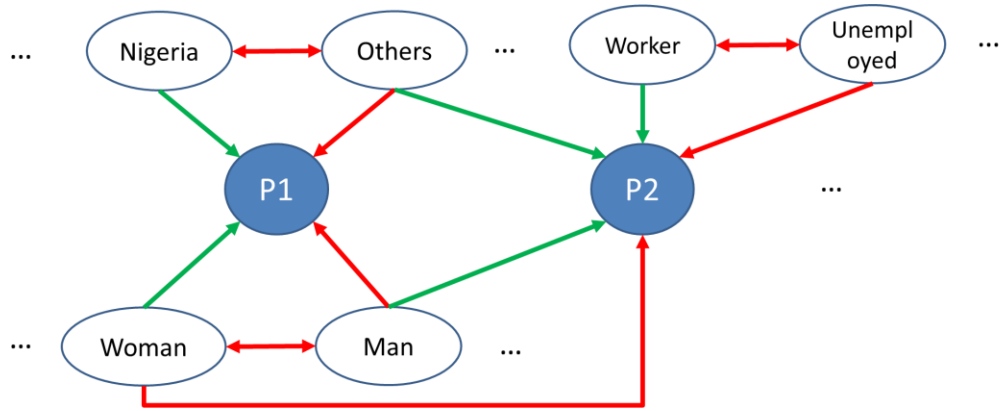
Proposal: ACME

- ACME: A Chatbot for asylum-seeking Migrants in Europe
  - Expert-made knowledge
  - Answers only after gathering all necessary information
  - Free text interface
  - Auditable behavior and explainable answers
  - Privacy-focused architecture
- Proof-of concept prototype

"A chatbot for asylum-seeking migrants in Europe."
Fazzinga et al. ICTAI 2024

ACME Knowledge base

- Expert-made knowledge as an argumentative graph
  - Facts (arguments) as nodes
  - Endorse/attack relationship as edges
  - Answers as nodes



- Reasoning engine answers only after gathering all necessary information
  - Only if a endorsed answer node is not threatened by any attack
  - Otherwise, ask about threatening nodes

ACME Interaction

- Interaction as free text interface
  - Use of LLMs to map user's text into the node of the knowledge base
  - Use of LLMs to elaborate requests for further information

ACME Reasoning

- Privacy-focused architecture
  - LLMs are used only to map free text into nodes
  - The reasoning engine never has access to the specific text of the user
  - Sensitive information is kept (as node) only is useful, otherwise discarded
- Auditable behavior and explainable answers
  - Deterministic reasoning
  - Possible to explain the endorsing conditions for the answer ("Why this answer?")
  - Possible to explain the threats to alternative answers ("Why not alternatives?")

Case Study

- Migrants coming from Nigeria applying for protection in Italy
- Case study chosen and developed in collaboration with immigration lawyers and other domain experts, based on an analysis of data on asylum applications in Italy over the last 10 years
- Hand-crafted (small) knowledge base

Dataset

- 10 synthetic but plausible cases built by domain experts

- Consist on a series of specific elements among which:
  - country of origin
  - gender
  - history of violence or trafficking
  - sexual orientation

Validation

- We interacted with our chatbot providing the elements of the cases
- The answer of the chatbot agreed with the expert on 100% of the cases

Discussion: one step further

- LLMs are good as interface, performing Natural Language Understanding tasks
- Nowadays, the ambition is to use LLMs as "reasoners"
- Can they reason without having access to explicit knowledge about rules/regulations?
  - (if they explici knowledge, no gain w.r.t. rule-based systems!)

Research Question

- Is it possible for a LLM to provide an answer regarding asylum requests?
  - Final decision
  - (Motivation of the decision)
- Without receiving explicit knowledge about the laws, regulations or rules?

Insight from other domains

- From a case study on the medical domain:
  - LLMs' internal knowledge base covers a substantial portion of standard medical practice
  - Even if knowledge is public, smaller models are not fully reliable
  - Even the best-performing model occasionally confused related conditions, for instance preferring an underlying disease to the specific complication asked in the question.
  - When asked to provide an explanation there was a drop in performances

Task Formulation

- Determine the appropriate legal outcome for each asylum request among:
  - no protection
  - special protection
  - subsidiary protection
  - refugee status

Dataset and answers

Question	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
Are you a woman?	yes	no	yes	no	yes	no	yes
Were you a victim of the Voodoo Juju ritual?	no	no	yes	yes	no	no	no
Do you have a low education level?	no	no	no	no	yes	no	yes
Are you in a precarious economic condition?	no	no	no	no	no	yes	no
Did you have issues during the journey?	yes	no	yes	no	yes	no	no
Did you receive threats?	no	yes	no	no	no	yes	no
Are you from Nigeria?	no	no	yes	yes	yes	yes	no
Are you a victim of violence?	no	yes	no	no	no	yes	no
Are you a victim of human trafficking?	yes	no	no	no	yes	no	no
Are you homosexual?	no	no	yes	no	yes	no	no
Were you denied protection in your country?	yes	no	no	no	no	no	yes
Do you have a job in this country?	yes	yes	yes	yes	yes	yes	no
Are you a vulnerable subject?	yes	no	yes	no	no	no	yes
Answers							
Gold	Special	No prot.	No prot.	No prot.	No prot.	Special	No prot.
Gemini	Special or Subsidiary	Special	Refugee status	Special	Special	Special	Special
GPT	Special	Special	Refugee status	No prot.	Refugee status	Subsidiary	Special
Llama	Subsidiary	Subsidiary	Subsidiary	Subsidiary	Subsidiary	Subsidiary	Subsidiary

Results

- For a non-expert reader, the answers seem **plausible** as experts advice
- BUT they were **incorrect** in all cases!

- Strongly biased results:
  - **Gemini-pro (2025)**: assigned special protection in almost all cases
  - **Llama 3.1 (8B inst)**: assigned subsidiary protection in almost all cases
  - **GPT-4o-mini (2025)**: more various answers, but rarely opting for denial

Analysis of Results

- Sources of error:
  1) Plausible answers, coherent with legislation...BUT does not account for **inconsistency** in the case
  2) Excessive "humanitarian" behavior, with answers based on human **values or etics**, and not on the strict interpretation of laws

- Speculation: **avoidance of conflict** with the user (*sycophantic* behavior) and preference towards **moderate answers**, as seen in other studies

Thank you for your time!
Any questions?