# UNHCR project initial plan

**Current situation:**

Literature that falls into relevant category is quite broad and perhaps a bit difficult to narrow down at this point what is useful and what is not.

**Tasks**

1. Clear problem definition what to focus on   
   - E.g. hate speech, racism or just negative sentiment targeting the UNHCR?   
   - This should take only like 1 meeting
2. Literature review   
   - Some literature already reviewed but more extensive lit review based on decided focus from part 1
3. What data to collect:   
   - E.g. Tweets with particular keyword or content?
4. Exploration of data    
   - We will need a meeting based on initial findings and possibly would need to return to step 1 or 3
5. Testing and developing a proof-of-concept model   
   - Building the first script(s)   
   - E.g. clustering of content (unsupervised ML) and seeing if the clusters make sense and if clustering is useful   
   - Creation of a basic supervised ML model with limited data. E.g. can we find specific content and label it accurately
6. Developing the full-scale model    
   - At this point we will see what worked in step 5 and refine it and run it with a larger dataset  
   - We will start working on analyzing the findings so that we can produce a publication   
   - We will work on producing a system that can be used as a tool at UNHCR   
   - Note! This is most likely **not** going to be a production level piece of software, but something which can be developed further if it is proven to work and seen as valuable

**Tentative timeline**

Note, this really depends on how fast things work and how many iterations we need on each step.

August

* Narrowed down literature review
* Data collection
* Starting analysis of data

September

* Exploratory data analysis
* Additional data collection if necessary
* Starting to work on the PoC model

October

* Tuning and working on PoC model

November

* If things work, start developing the full-scale model
* Initial findings

December

* Testing and finalizing full scale model
* Starting work on paper and turning the model into a tool