Conversational Agents and Societal Impact

Marco Guerini
Fondazione Bruno Kessler
guerini@fbk.eu

Introduction
Conversational Agents

Conversational Agents AKA Dialog Agents

What we have in mind when we talk about CA:

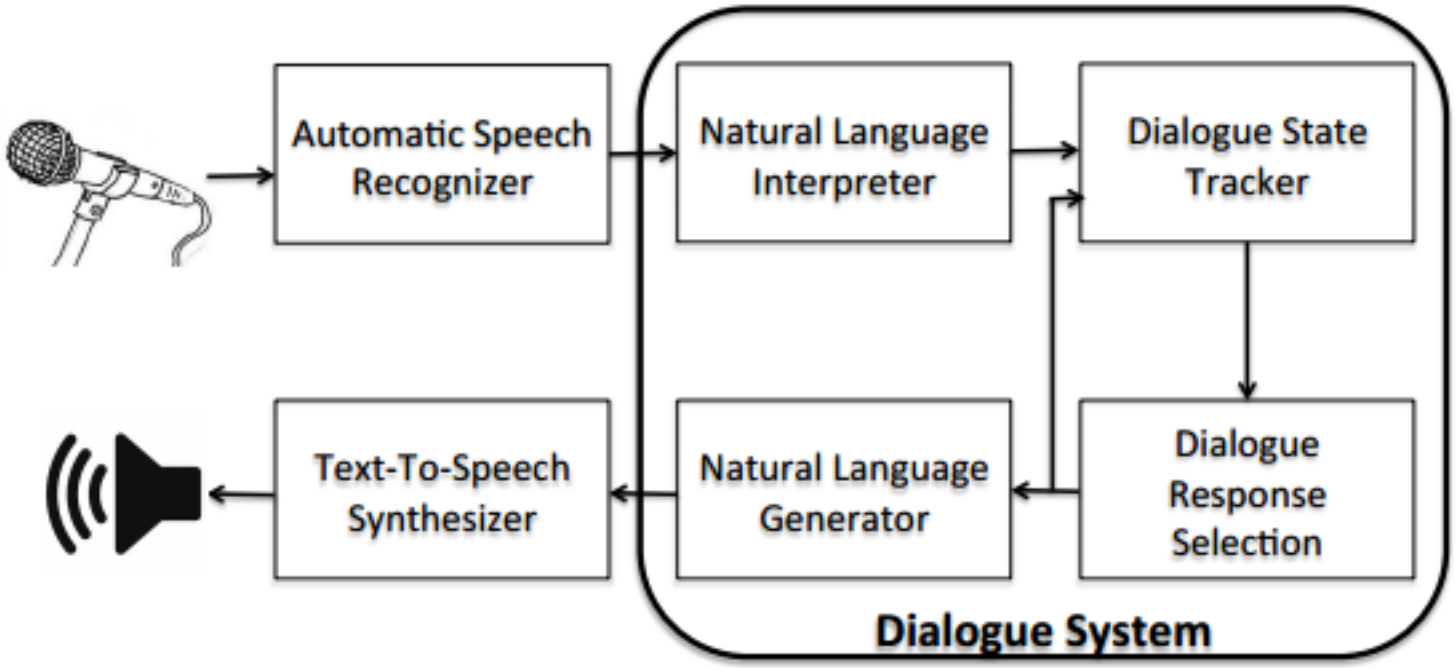
- Phone Personal Assistants, e.g. SIRI
- Home - Alexa
- Talking to your car
- Communicating with robots
- Clinical uses for mental health
- Chatting for fun

Classical distinction: two classes of CAs

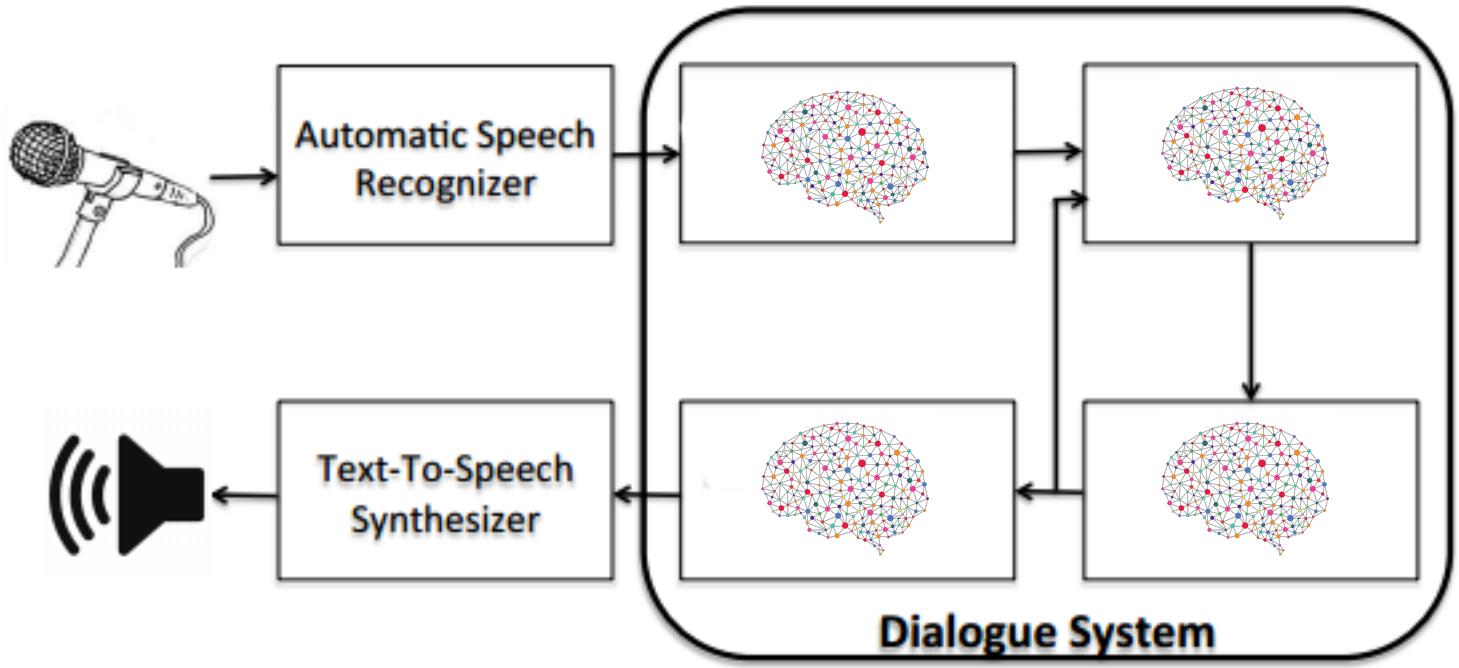
Task Oriented

Chatbots

Typical Modular Architecture

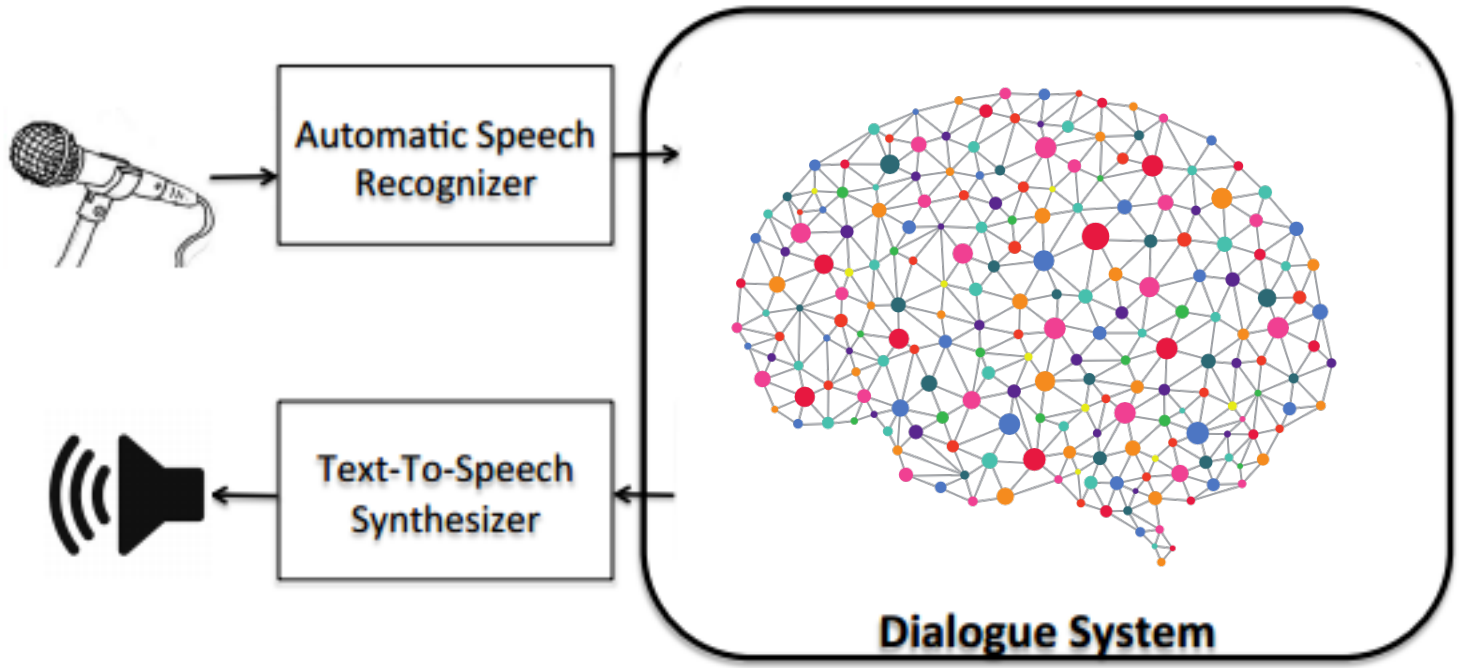


Typical Neural Modular Architecture



Artificial Neural Networks are computing systems vaguely inspired to real brains (interconneted computing nodes). They can learn several tasks from input-output data pairs.

Typical end-to-end Architecture



Advantages and Limits

- Neural Model are very good at generalizing …
- Need lots of data.

- Either simulated data via crowdsourcing or
- Real data from real interactions

Model of Conversational Agents

Retrieval Based	Generative Based
repository of predefined responses	generate new responses from scratch
Short Conversation	**Long Conversation**
single response to a single input	multiple turns: need to keep track of what said before
Closed Domain	**Open Domain**
possible inputs and outputs are limited (goal based)	user can take the conversation anywhere
No Personality	**Personality and Persona**
Just do the task - no underlying coherent traits	CA has his own style, preferences, story.

Section 1
Typical Task Oriented CA

Frame & Slot structure for task oriented CAs

A data structure that represents the info needed to fulfill the task. The dialogue is the activity carried out to collect relevant info fro the user (i.e. to fill the slots)

Show me morning flights from Boston to SF on Tuesday.

SHOW:
- FLIGHTS:
  - ORIGIN:
    - CITY: Boston
    - DATE: Tuesday
    - TIME: morning
  - DEST:
    - CITY: San Francisco

A close future

The future of work will be a hybrid environment that involves both human and machine intelligence working in conjunction toward the same shared goals (e.g. industry 4.0).

What about societal impact and social good?

Societal Impact of CA technologies

The Hatemeter Project

The project

- Project Hatemeter aims at increasing the efficiency and effectiveness of NGO/CSOs in preventing and tackling Islamophobia
- Developing and testing an ICT tool that automatically monitors and analyses Internet and social media data, and produces computer-assisted responses and hints to support counter-narratives.

An additional Class of CAs

There's an additional class of CAs - not much inspected so far - useful for our analysis:

- No Frames, no slots
- still goals to pursue (e.g. motivational systems)
- What about societal goals?

Task Oriented **Goal Oriented** Chatbots

< Socially > Augmented Humans

- Special class of highly influential users exploiting thousands of **bots**, for enhancing their online influence.
- Digital augmentation represents the online counterpart of the human desire to acquire power within social systems.
- Augmented humans generate deep information cascades, to the same extent of news media and other broadcasters.

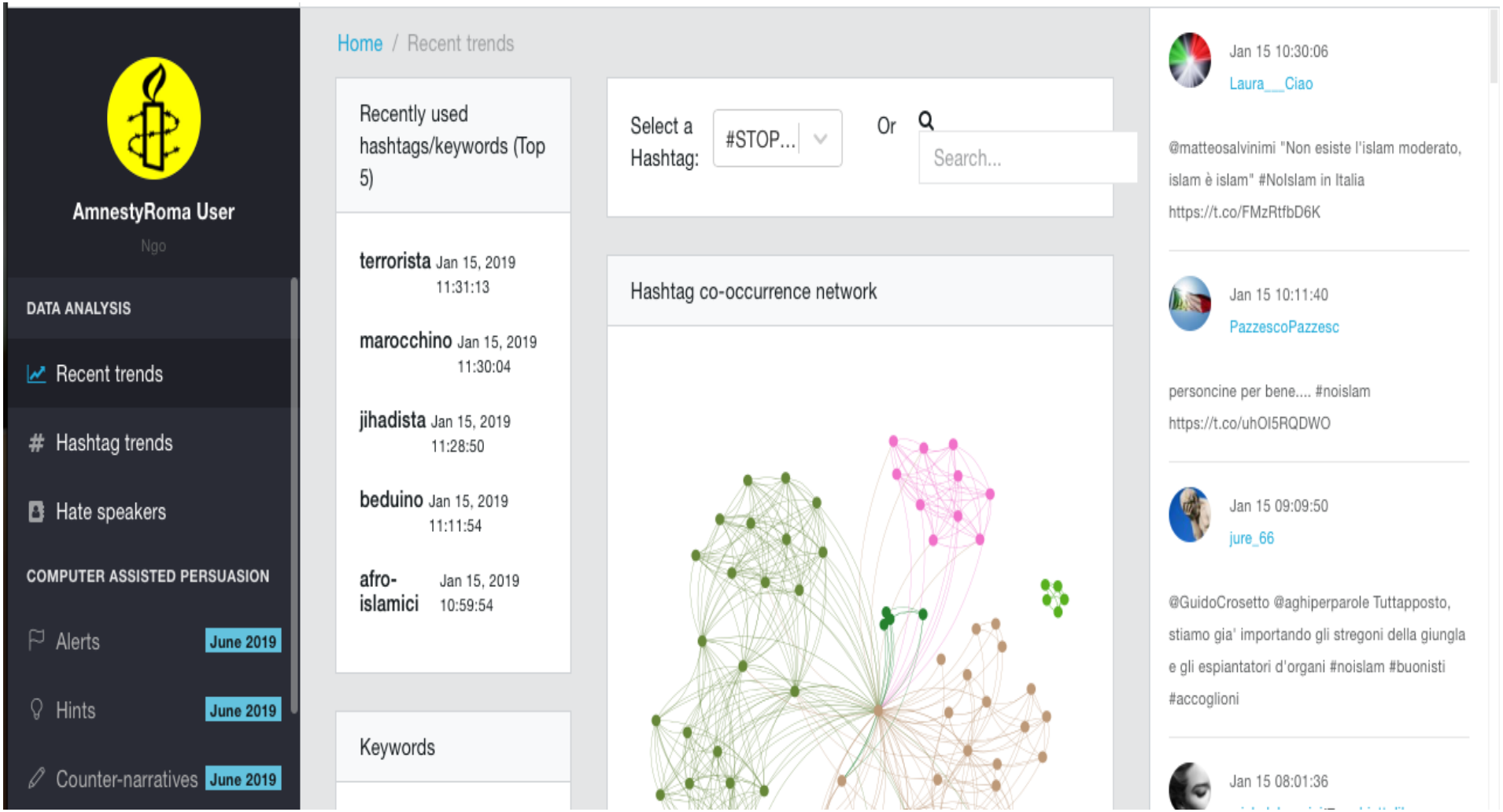
Such augmented humans already act on Social Media Platforms, often to spread misinformation and hate.

Tools for Computer Assisted Persuasion

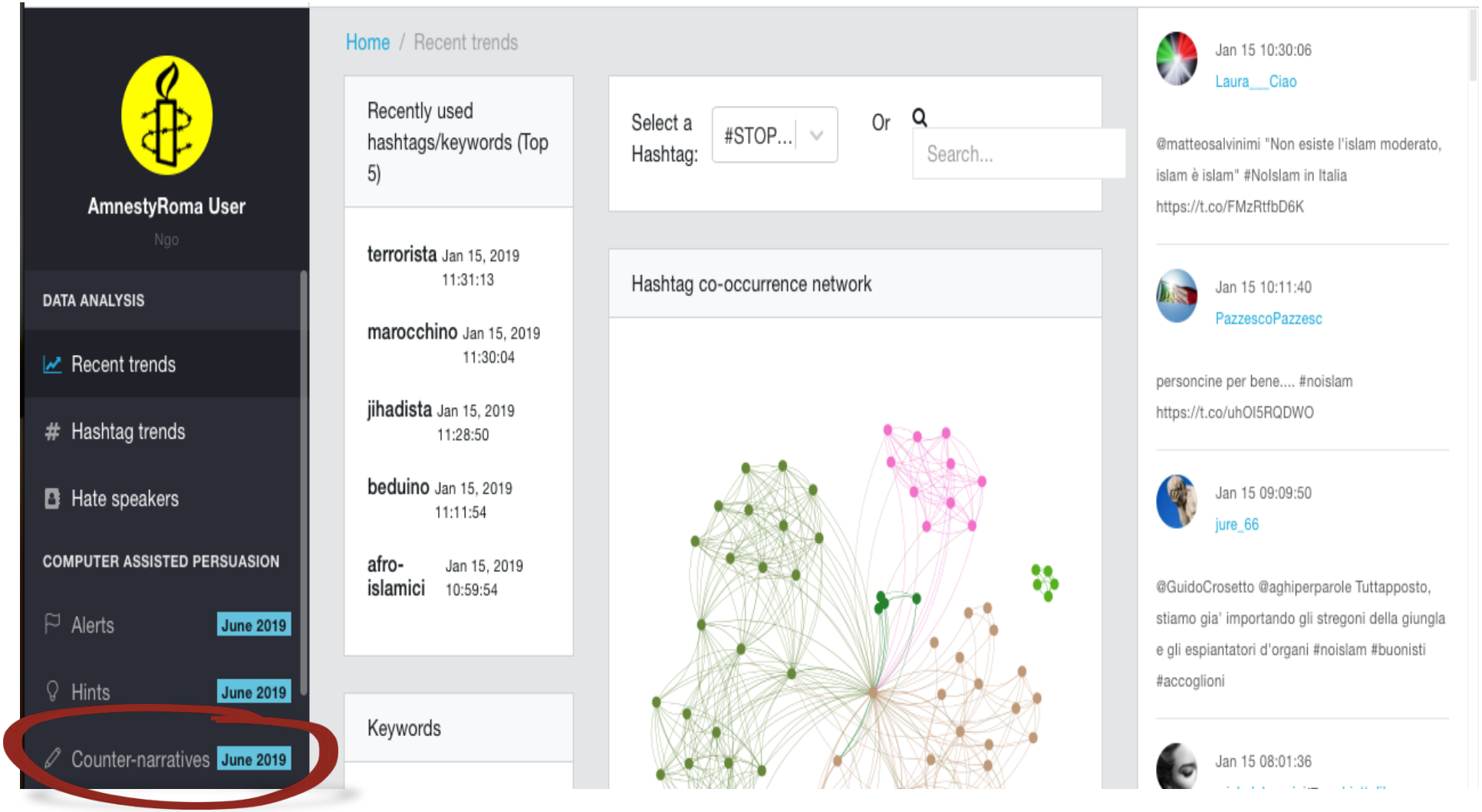
- Implementation of a suite of **Computer Assisted Persuasion (CAP) tools** for NGOs to analyze and monitor hate speech online
- **Automatic Generation of 'responses'** to hate posts the operator can choose from and possibly edit
- **Generation:** Expert-Based and/or Data-Driven

Augmented humans to fight hate online.

Hatemeter Platform



Hatemeter Platform



Conversation Technologies

Counter Messaging

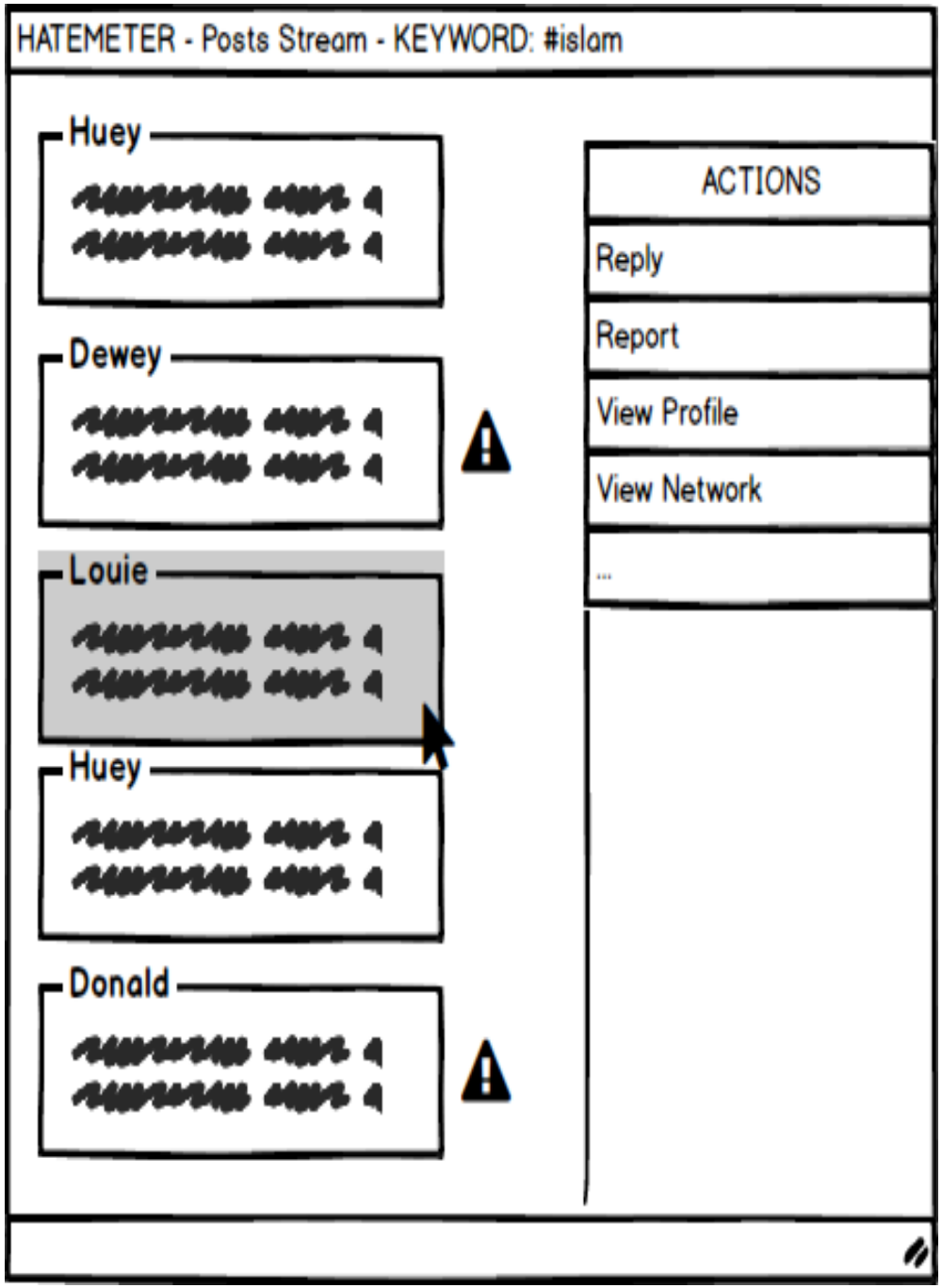
Bot or suggestion tools

- Technologies for the suggestion tool or for bots are the same.
- Can be very simple or very complex.
- Nowadays are mainly data-driven
- The more data they are provided with, the better they work.

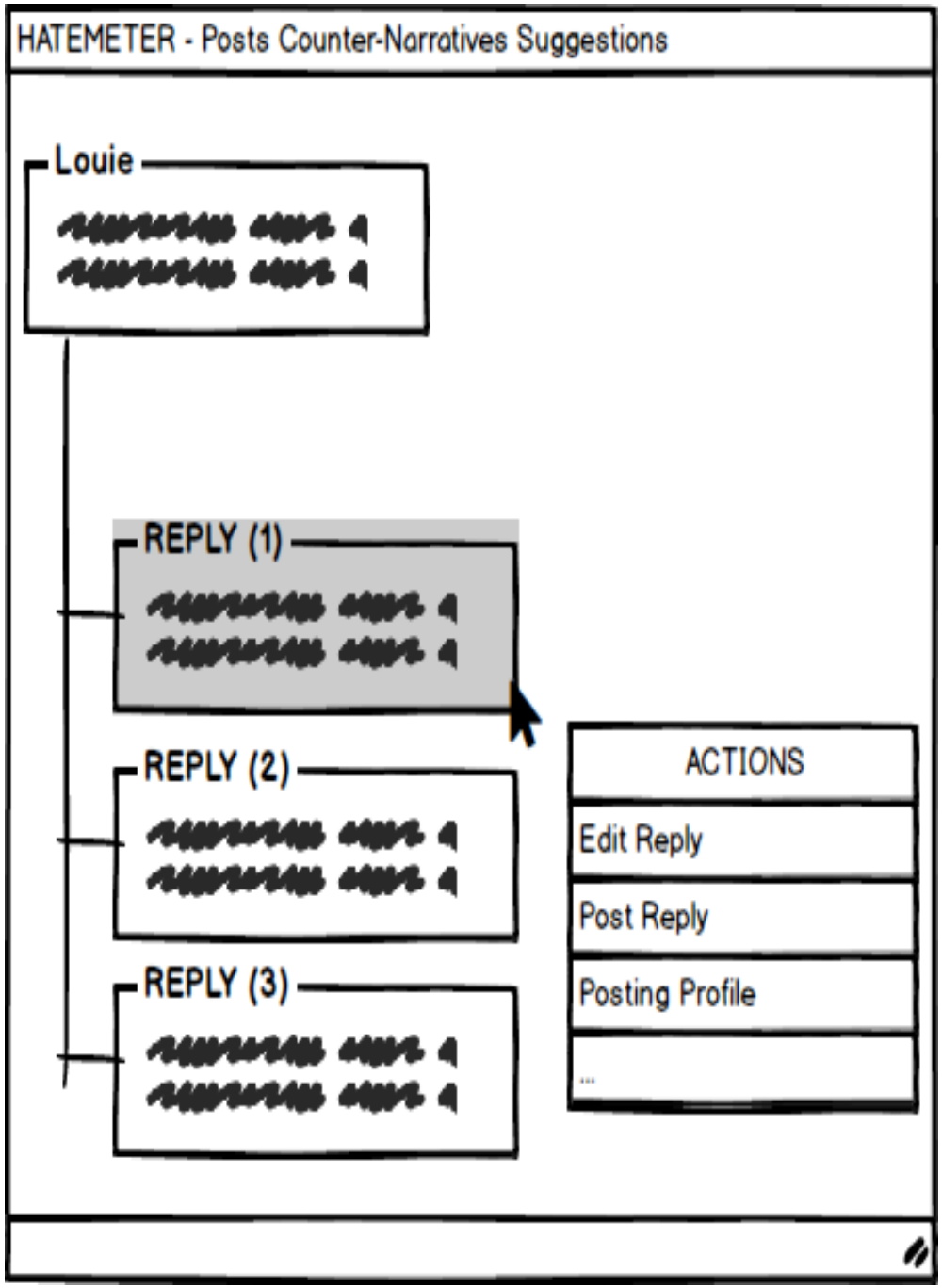
Example of expert-based guidelines

Exploit and automatize experts' suggestions with **text generation / recommendation tools**

- A stream of messages coming from the monitored sources
- Some messages are signaled for potential danger
- The operator can click on a message to activate a list of actions
- One of these activate the counter-messaging tools



- The counter messaging tool provide suggestions on possible responses
- The operator can select a response and possibly edit it before posting.
- The control and the final decision is always in the hands of the operator.



NicheSourcing: Expert Based data collection

- More than 500 hours of data collection with NGOs
- More than 100 operators
- More than 14 thousand pairs
- Three Languages (EN – FR- IT)
- Operator demographics collected
- Counter-Narrative type annotation

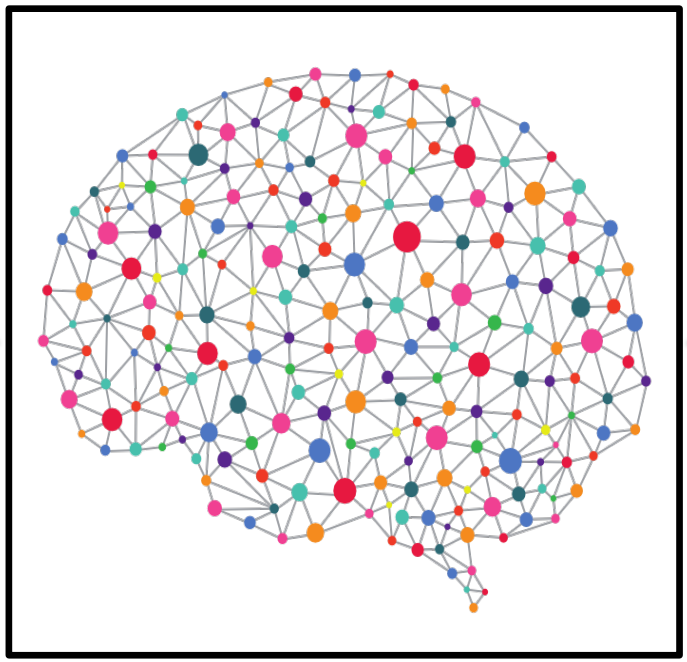
Data pairs

To each hate speech we paired a counter-narrative written by the NGO operators

Every Muslim is a potential terrorist.	Every Muslim is also a potential peacemaker, doctor, philanthropist... What's your point?
The veil is contrary to secularism.	On the contrary, secularism allows every citizen to freely profess his faith.
…	…
…	…

How it works - Learning

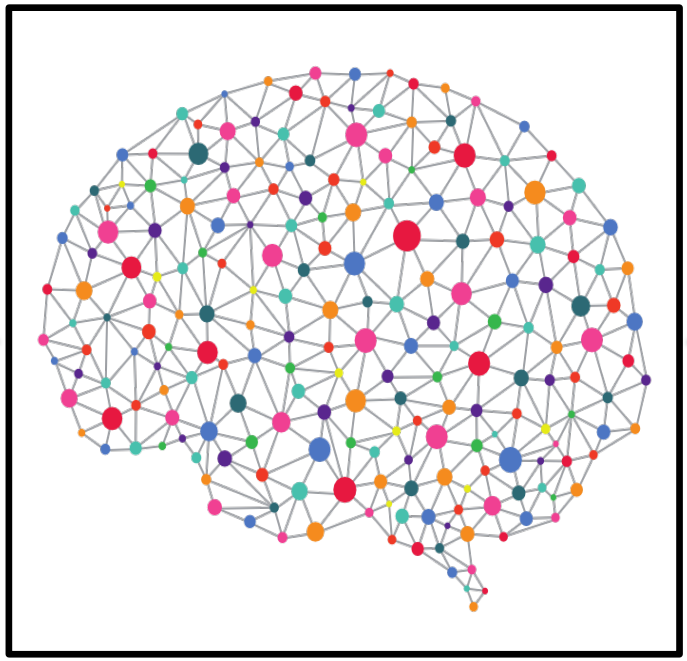
Every Muslim is a potential terrorist.



Every Muslim is also a potential peacemaker, doctor, philanthropist...

How it works - Selection

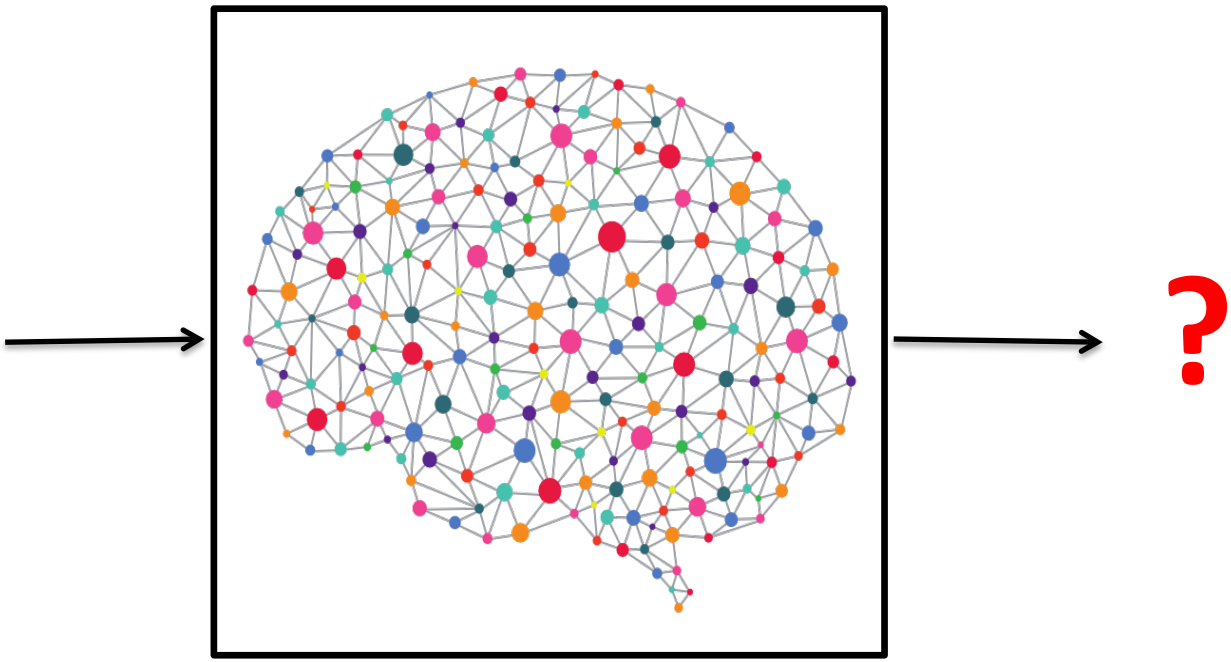
Every Muslim is a potential terrorist.



Every Muslim is also a potential peacemaker, doctor, philanthropist...

How it works - Generation

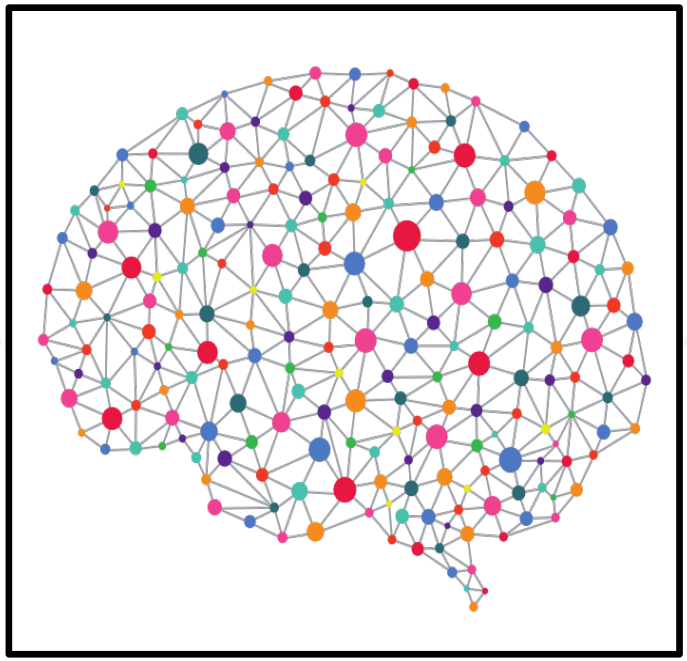
Every Muslim is a potential terrorist.



?

How it works - Generation

Every Muslim is a potential terrorist.



On the contrary, every Muslim can be a potential peacemaker, doctor, philanthropist...

Conclusions

Conversational Agents: a fast growing and fast evolving field

Key elements: neural models and need for data

Societal impact and social good not much inspected so far

Thank you!

Contacts

Marco Guerini

Fondazione Bruno Kessler
Via Sommarive 18,
38123 Trento (Italy)

Mail: guerini@fbk.eu
Twitter: @m_guerini