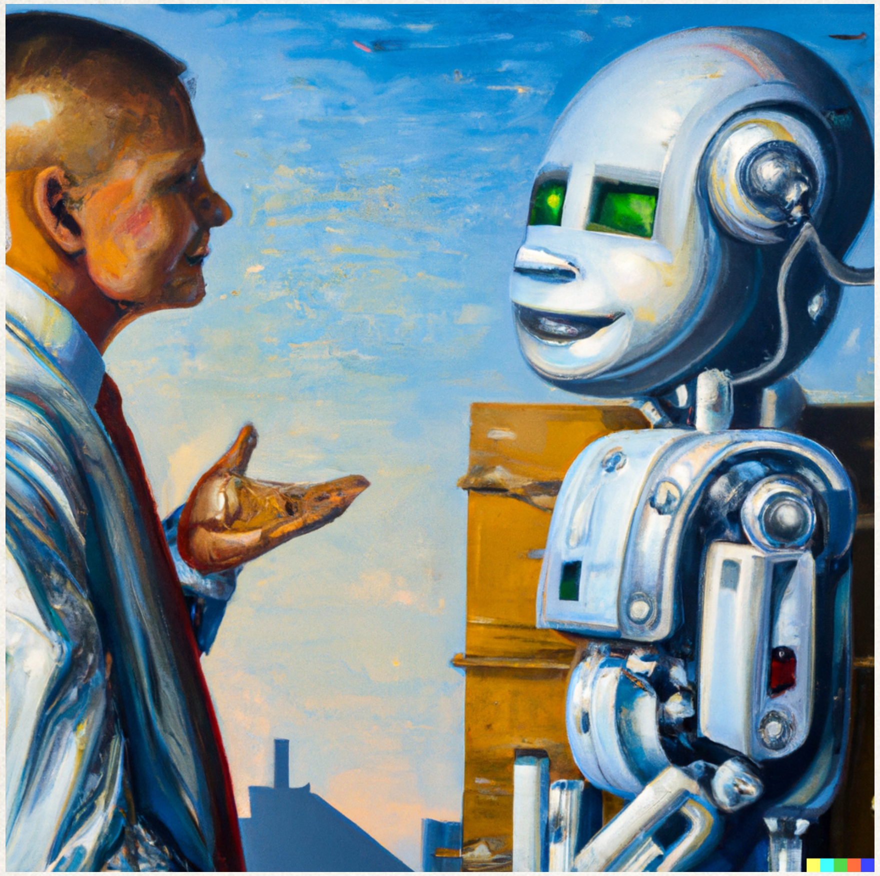


Automatically generated by DALL-E « an oil painting that shows a social conversation between a human and a robot »

How to integrate the socio-emotional component in conversational systems?

Chloe Clavel,

LTCI Telecom-Paris, IDS/S2A/Social Computing Team

https://clavel.wp.mines-telecom.fr/,

December 2022

Scope: Socio-conversational AI

- Machine learning for the analysis and generation of socio-emotional behaviors (**text**, voice, gestures, facial expressions, posture)
- **Applications:** Web analysis, Human-agent interactions
- **Application areas:** Societal trend analysis, Education, Human resources, Health, Customer relationship management, Space
- **Keywords:** Affective/Social Computing Natural Language Processing, Human-agent interactions, Conversational Artificial Intelligence, Explainable AI

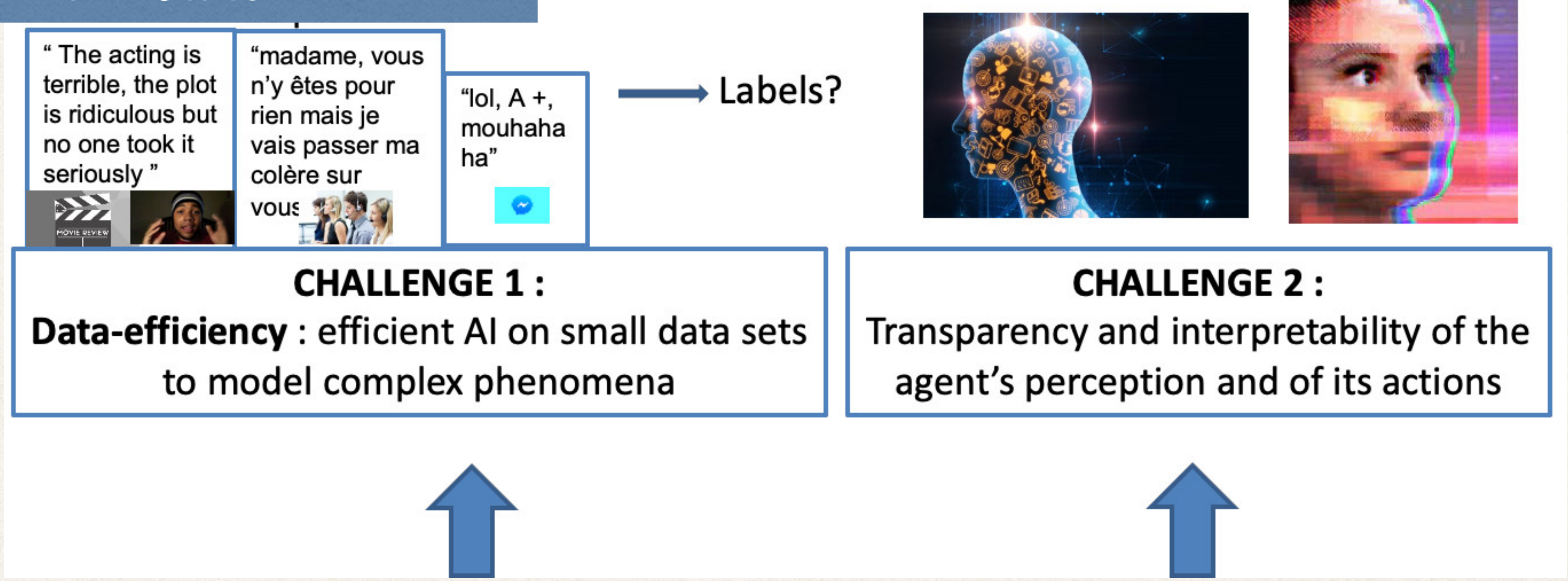
Figure 1: The JUSThink activity setup.

ITN European project ANIMATAS

Scientific challenges

Socio-emotional phenomena (ex. trust, frustration, engagement , etc.) are difficult to define and annotate

Social and ethical impact of making the machine able to understand and reproduce socio-emotional phenomena

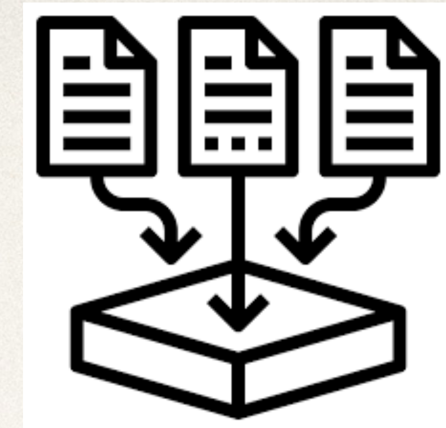


Our approach:
integrating human and social sciences at the heart of machine learning

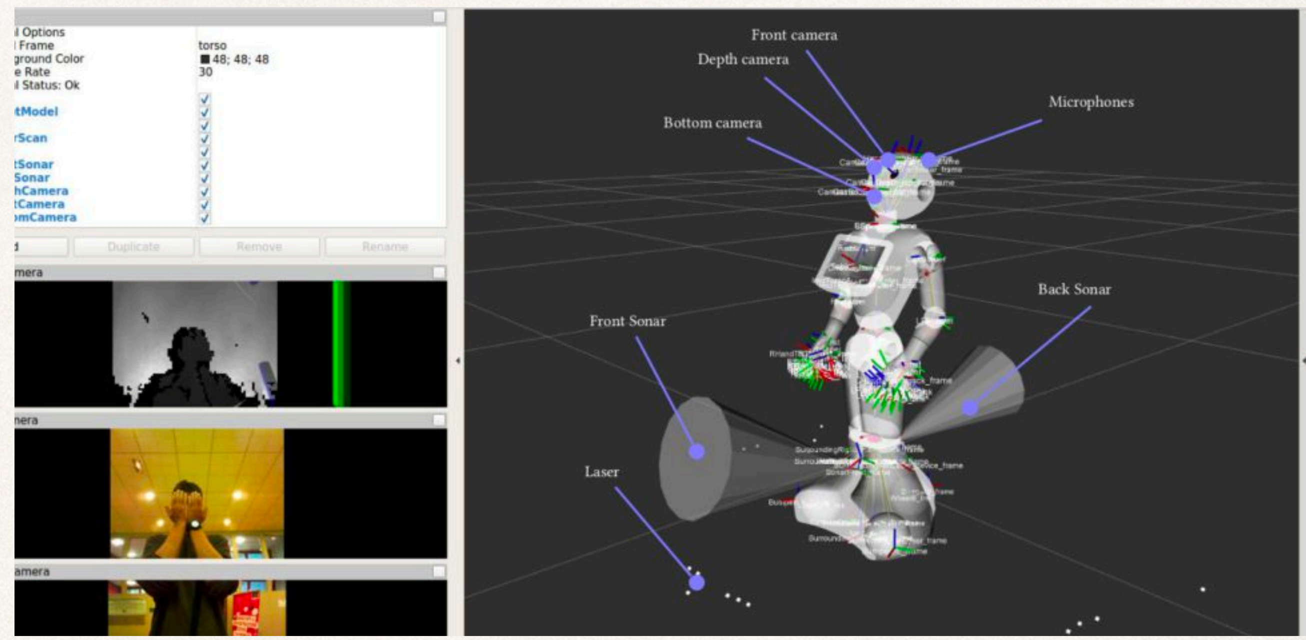
Two chapters in this presentation

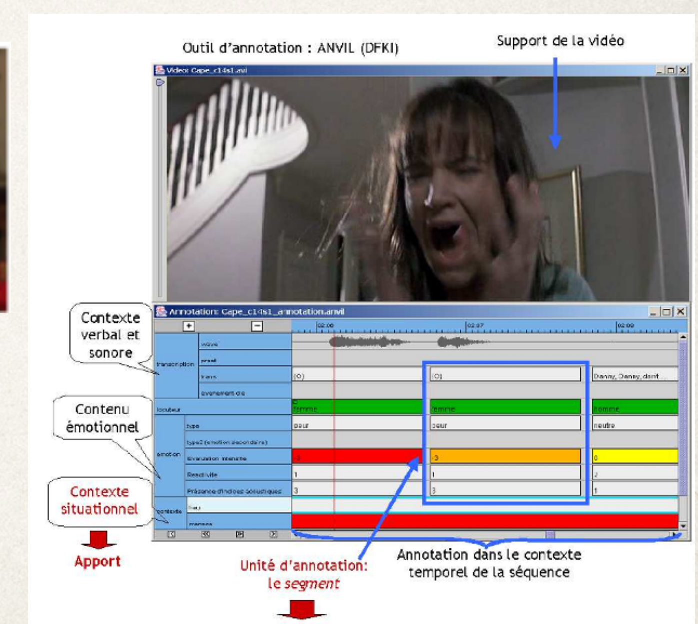
- Prologue: collecting and annotating data for supervised machine learning models
- Chapter 1: data/label efficient socio-emotional models
- Chapter 2: explainable socio-emotional models

Collecting new socio-emotional data



- Human-robot interactions (ex. UE-HRI)
- Human-human interactions (ex. SAFE movie corpus)
- Monologues (ex. Political addresses - POTUS corpus)





Available at https://clavel.wp.imt.fr/corpora/

Annotating data: Providing new coding scheme and annotation tools

Langlet et al.. A **Web-Based Platform** for Annotating **Sentiment-Related Phenomena** in **Human-Agent Conversations**. IVA 2017

Guibon et al. EZCAT: an Easy **Conversation Annotation Tool**. In LREC 2022.

Hulcelle et al., TURIN : A coding system for **Trust** in hUmanRobot INteraction ACII 2021

Janssoone, et al. « The POTUS Corpus, a database of weekly addresses for the study of **stance** in politics and virtual agents. » LREC 2020

Rollet & Clavel. "Talk to you later" Doing social **robotics** with conversation analysis. Towards the development of an automatic system for the prediction of **disengagement**, Interaction Studies 2020

Clavel et al., **Fear**-type emotions recognition for future audio-based surveillance systems. Speech Communication, 2008.

Theoretical models from psychology, linguistics, conversation analysis (ex. Psychological models for emotion and engagement, socio-linguistic definition of trust)

Example of a coding schema: Trust in human-robot interaction

Hulcelle et al., TURIN : A coding system for **Trust** in hUmanRobot INteraction ACII 2021

- **Theoretical framework:** Interactional sociology (*Goodwin. 1981. Conversational organization. Interaction between speakers and hearers*)

- **Hypothesis:** observability of trust within the interaction, via the participants' behaviors
- **List of observable behaviors** (e.g., communication modes based on benevolence such as self-disclosure)

- **Coding scheme :** units of annotation, (sub)categories of observable behaviors
- **Annotate** an existing human-robot interaction dataset (Vernissage dataset)

Chapter 1: data/label efficient socio-emotional models

Overview: data/label efficient socio-emotional models

Reasoning models	Hybrid approaches	Data augmentation	Meta-learning
• Agent's gesture generation • Detection of users ' likes	• **Knowledge-driven features** • Interactional dynamics at encoder and decoder levels	• **Logical rules for generating entailment data : see poster session!** • Extreme value theory for generating rare sentiment data	• Transfer socio-emotional information learned on one corpus (e.g,. tweets) to a second corpus (e.g. chat)
• [Ravenet et al., AAMAS 2018 • [Langlet & Clavel, ACL 2015]	• **[Raphalen et al., ACL 2022;** Barriere et al., ICASSP 2018] • [Chapuis et al., AAAI 2020; Colombo et al., EMNLP 2021]	• **[Helwé et al., F. EMNLP 2022]** • **[Jalalzai et al., Neurips 2020]**	• [Guibon et al., EMNLP 2021, AAAI 2022] • [Deng et al., ACII 2022]

Theoretical models from psychology, linguistics, conversation analysis
(ex. Cognitive models for gesture generation, Heider balance theory for likes and dislikes)

Focus on an hybrid model for detecting hedges in peer-tutoring interactions

Raphalen, Clavel and Cassell. « You might think about slightly revising the title »: identifying hedges in peer-tutoring interactions. ACL 2022

Hybrid model: knowledge-driven textual features + machine learning models

Ex. "You might think about asking questions at the end of this presentation » vs. « Ask questions ! »

Descriptions of **hedges** (a pragmatic competence, dedicated to mitigating the social imposition of a proposition) from linguistic theories: Rowland (2007), Fraser (2010) and Brown and Levinson (1987),

Linguistic patterns

<table>
<tr><th>Class</th><th></th></tr>
<tr><td>j.</td><td>(?!what).*(i|we) ?(don't|didn't|did)? ?(not)?
(guess|guessed|thought|think|believe|believed|suppose|supposed)
?(whether|if|is|that|it|this)?.*</td></tr>
<tr><td>Subj.</td><td>.*(i|i'm|we) ?(was|am|wasn't)? ?(not)? (sure|certain).*</td></tr>
<tr><td>Subj.</td><td>.*(i feel like you).*</td></tr>
<tr><td>Subj.</td><td>.*(you (might|may) (believe|think)).*</td></tr>
<tr><td>Subj.</td><td>.*(according to|presumably).*</td></tr>
<tr><td>Subj.</td><td>.*(i|you|we) have to (check|look|verify).*</td></tr>
<tr><td>Subj.</td><td>.*(if i'm not wrong|if i'm right|if that's true).*</td></tr>
<tr><td>Subj.</td><td>.*(unless i).*</td></tr>
<tr><td>Apol.</td><td>.*(i'm|i|we're) (am|are)? ?(apologize|sorry).*</td></tr>
<tr><td>Apol.</td><td>(?!.*(be|been|was) like excuse me)((excuse me|sorry)[w ,']+|[w ,']+(excuse me|sorry))</td></tr>
<tr><td>Prop.</td><td>.*(just|a little|maybe|actually|sort of|kind of|pretty
much|somewhat|exactly|almost|little bit|quite|
regular|regularly|actually|almost|as it were|basically|
probably|can be view as|crypto-|especially|essentially|
exceptionally|for the most part|in a manner of speaking|
in a real sense|in a sense|in a way|largely|literally|
loosely speaking|kinda|more or less|mostly|often|
on the tall side|par excellence|particularly|
pretty much|principally|pseudo-|quintessentially|
relatively|roughly|so to say|strictly speaking|
technically|typically|virtually|approximately|
something between|essentially|only).*</td></tr>
<tr><td>Prop.</td><td>.*(i|i'm|you|it's) (am|are) (apparently|surely)[,]?.*</td></tr>
<tr><td>Prop.</td><td>.*(it) (looks|seems|appears)[,]?.*", ".* (or|and) (that|something|stuff|so forth)</td></tr>
</table>

+ Linguistic resources (LIWC)

Knowledge-Driven Features (KDF)

Focus on an hybrid model for detecting hedges in peer-tutoring interactions

Raphalen, Clavel and Cassell. « You might think about slightly revising the title »: identifying hedges in peer-tutoring interactions. ACL 2022

Data: peer-tutoring interactions (23000 utterances)

Models	KD Feat. (KDF)	Pre-Trained Emb. (PTE)	KDF + PTE
Rule-based (3-classes)	67.6	Ø	Ø
MLP (3-classes)	68.5 (1.6)	35.8 (3.1)	64.8 (1.1)
Attention-CNN (3-classes)	Ø	64.5 (3.0)	Ø
LSTM (3-classes)	65.1 (5.7)	39.8 (8.0)	65.2 (5.1)
BERT (3-classes)	Ø	**70.6 (2.3)**	Ø
LGBM (3-classes)	$\boxed{\textbf{79.0 (1.3)}}$	35.0 (2.2)	**70.1 (1.4)**

Knowledge-Driven Features

BERT

Best results (F1 score) obtained with KDF and LGBM
(Light Gradient Boosting Machine).

Chapter 2: explainable socio-emotional models

Overview: explainable socio-emotional models

Post-modelling explainability

- SHAP analysis of features that matter for hedge detection
- Analysis of attention mechanisms of neural networks in order to identify *attention slices*

- [Raphalen et al., ACL 2022]
- [Hemamou et al., Trans. on Affective Computing, 2021]

« BERTology » : Analyzing BERT pre-trained representations

- Information about fillers
- Information about stances

- [Dinkar et al., EMNLP 2020]
- [Gari Soler et al., COLING 2022]

Outputs interpreted from literature of psychology, linguistics, conversation analysis

Attention slices for explainability

L. Hemamou; A. Guillon; J.C. Martin; C. Clavel, Multimodal Hierarchical Attention Neural Network: Looking for Candidates Behaviour which Impact Recruiter's Decision. IEEE TaffC 2021

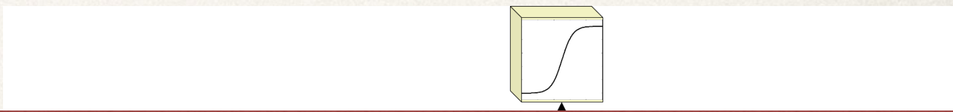
Research question: What are the social signals that are impacting recruiters decision during a job interview?

Method:

Step 1 - build a neural model dedicated to reproduce the recruiters assessment

Step 2 - study **attention mechanisms** of the neural model in order to identify *attention slices* (salient moments in the assessment of job interviews)

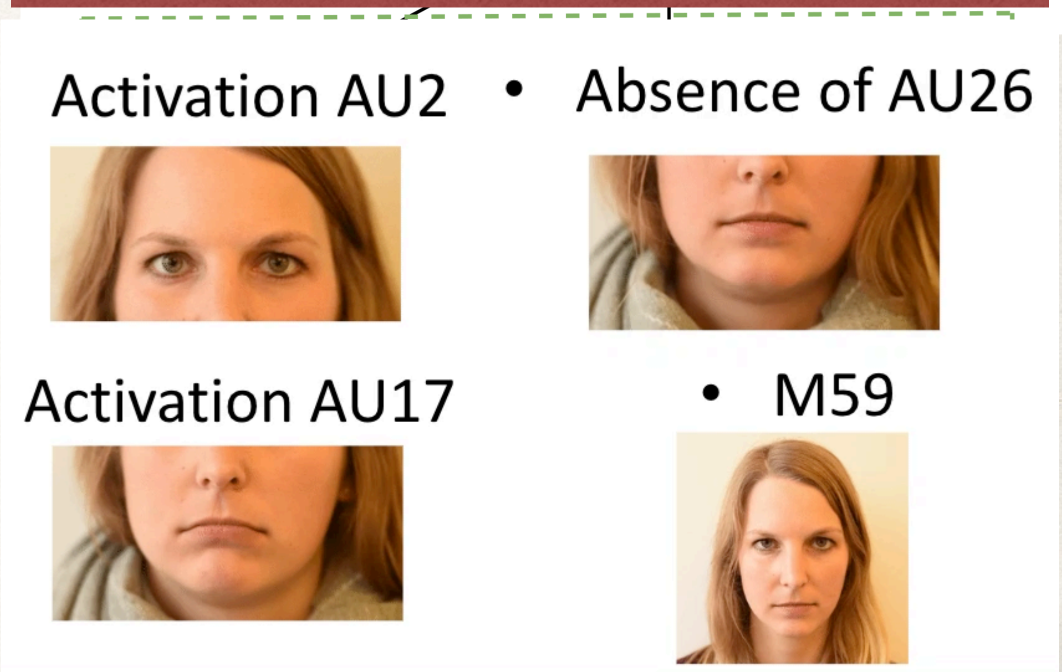
Step 3 - analyze the timing and the content of attention slices in terms of social cues



Attention slices tend to occur at the beginning and at the end of an answer

Social cues characterizing the attention slices

ex. breathing, fillers, activation of some action units:

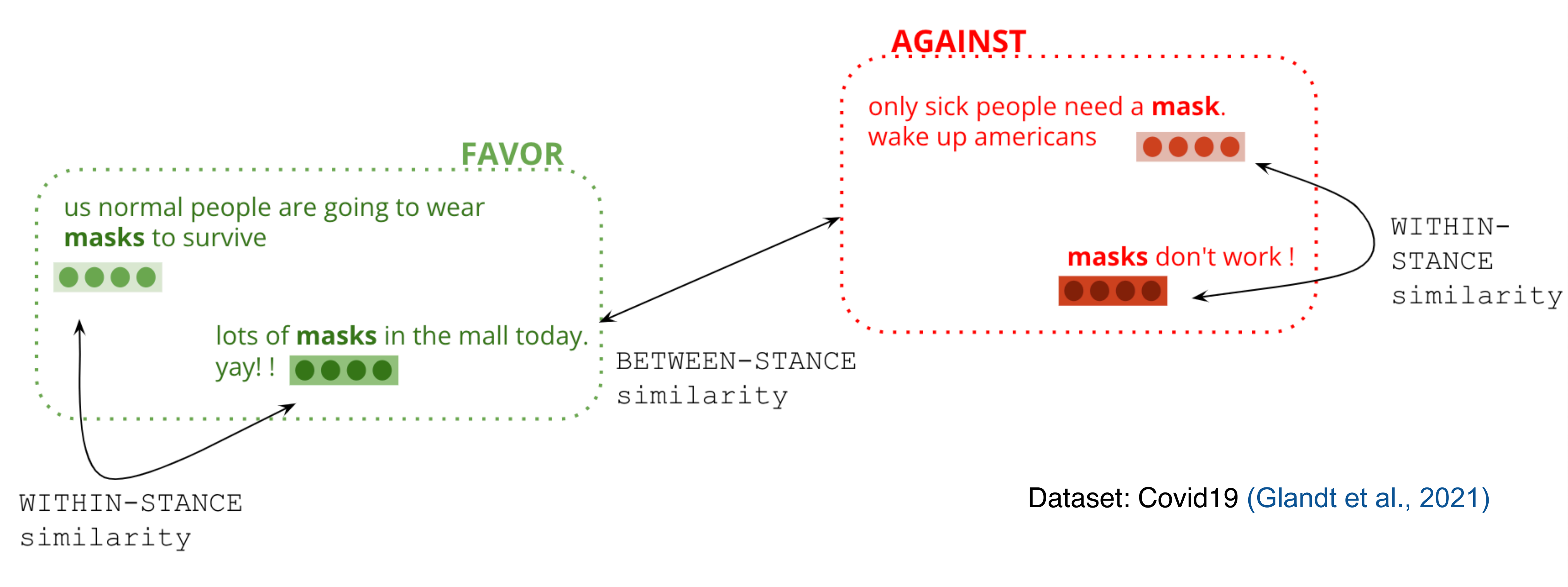


BERT word representations and stances

A. Garí Soler, M.Labeau and C. Clavel (2022). One Word, Two Sides: Traces of Stance in Contextualized Word Representations. COLING

Are BERT word representations sensitive to the opinion expressed ?

Method:



BERT word representations and stances

A. Garí Soler, M.Labeau and C. Clavel (2022). One Word, Two Sides: Traces of Stance in Contextualized Word Representations. COLING

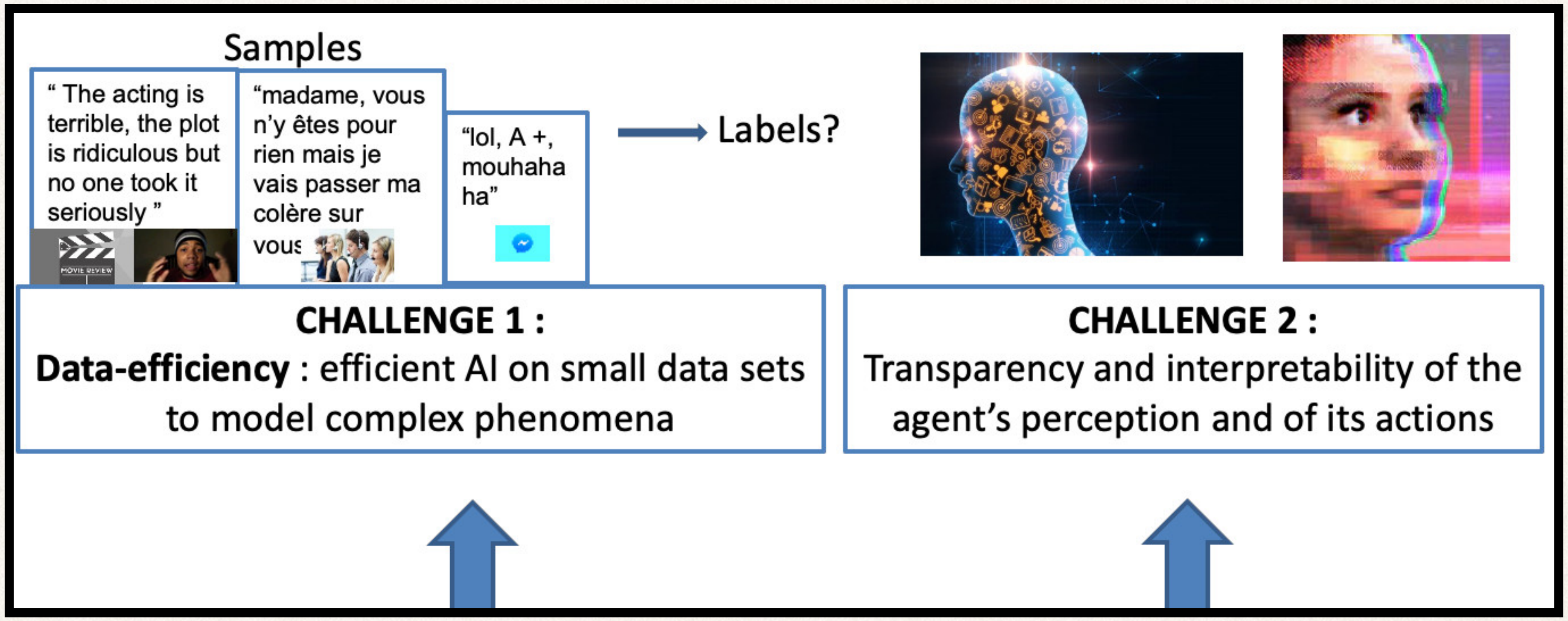
Are BERT word representations sensitive to the opinion expressed ?

Conclusions:

- Differences in similarity between concurring and conflicting stances are small, but significant.
- Words with the highest differences tend to be central to the topic: potentially useful for detecting points of discordance.

Dataset	Target	Most different	Least different
SemEval 2016	Feminist Movement	woman, men, equality, gender	come, leave, believe, go, take, tell
SemEval 2016	Atheism	religion, #god, believe, #freethinker	man, think, go, take, make, come
ArgQ	Zoos	animal, zoo, live, habitat	life, allow, make, provide, keep, take
ArgQ	Nuclear weapons	weapon, country, use, war	maintain, keep, life, mean, make, world

Epilogue



Social Sciences

In the supervision of machine learning models
In the design of features used by machine learning models
At encoder and decoder levels of neural architectures
In the design of data augmentation and meta-learning approaches
For the interpretation of the models

Merci !

Collaborateurs pour les travaux détaillés ici (dans l'ordre d'apparition):

Marc Hulcelle (PhD student), Nicolas Rollet (I3, Telecom-Paris), Giovanna Varni (Trento University), Yann Raphalen (ex PhD student), Justine Cassell (CMU & Inria Paris), Léo Hemamou (PhD, ICIMS), Jean-Claude Martin (LISN), Aina Gari Soler (post-doc),
Matthieu Labeau (LTCI, Telecom-Paris)

Autres travaux évoqués:
Catherine Pelachaud (ISIR), Brian Ravenet (LISN), Emile Chapuis (ex PhD student), Pierre Colombo (ex. PhD student), Hamid Jalalzai (ex. PhD student), Anne Sabourin (Université Paris Cité), Chadi Helwé (PhD student), Fabian Suchanek (LTCI, Telecom-Paris), Gaël Guibon (ex post-doc), Luce Lefeuvre (SNCF), Tanvi Dinkar (ex PhD student), …

We're hiring!

Télécom Paris has a new permanent (tenure) faculty position (**Associate Professor/ "Maître de conférences"**) in the area of machine learning for social computing.
Applicants from the following sub-research areas are welcome:

- Neural models for behaviour recognition and generation
- Natural language and speech processing
- Dialogue, conversational systems and social robotics
- Reinforcement learning for dialogue
- Sentiment analysis in social interactions
- Bias and explainability in AI
- Model tractability, multi-task learning, meta-learning

Please contact **Matthieu Labeau** matthieu.labeau@telecom-paris.fr and **Chloé Clavel** chloe.clavel@telecom-paris.fr for more information