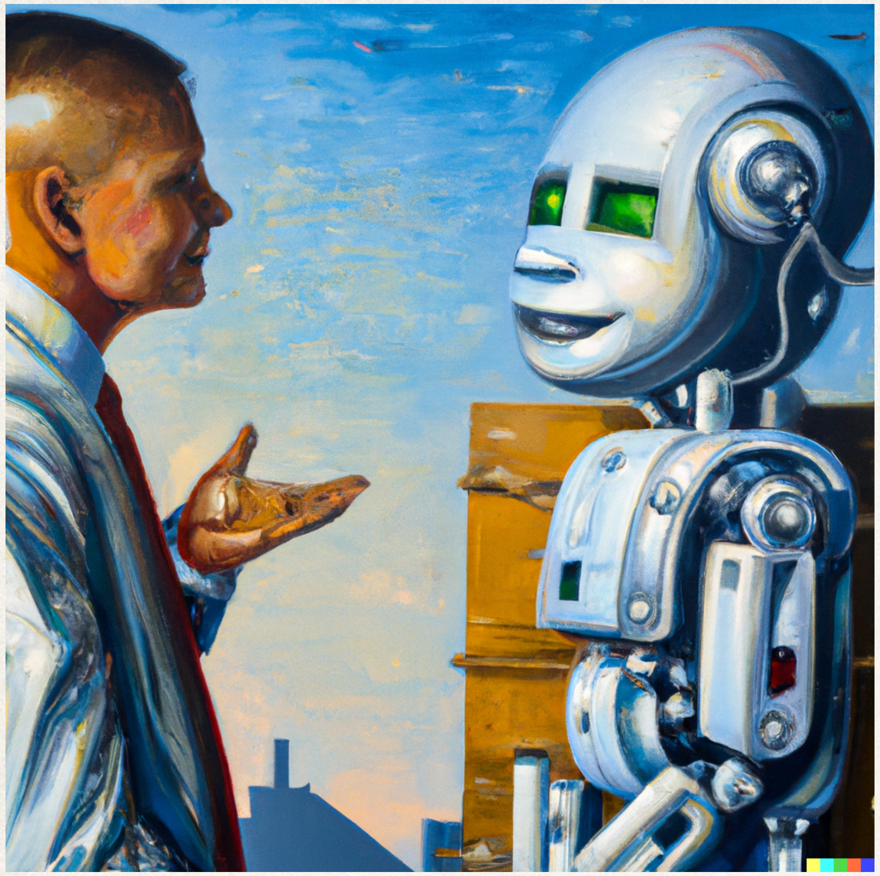


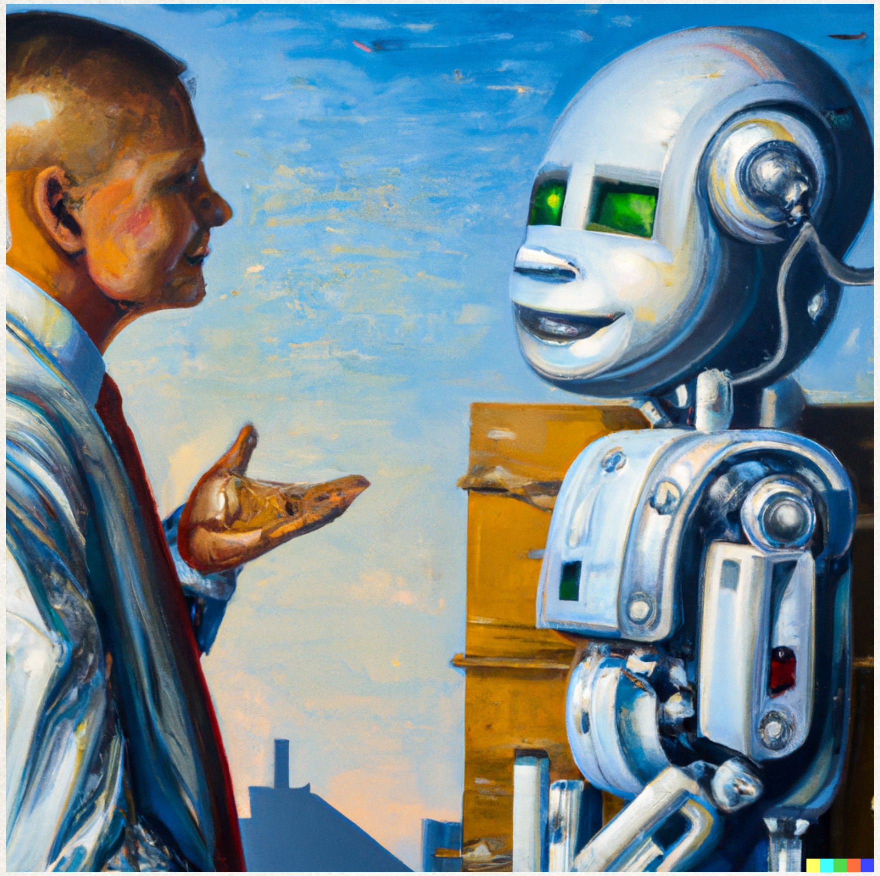
Socio-Conversational AI
Integrating the social component in interactions using neural models

Chloe Clavel,

Polytechnic Institute of Paris, Telecom-Paris, LTCI, Social Computing Team

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

February 2023



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

Socio-Conversational AI
Integrating the social component in interactions using neural models

Chloe Clavel,

Polytechnic Institute of Paris, Telecom-Paris, LTCI, Social Computing Team

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

February 2023

Scope: Socio-conversational AI

Socio-emotional phenomena: catch-all term, that I will use here and that gathers both emotion, social stance, sentiment, mood, trust, engagement, stance, conversation strategies etc.

Scope: Socio-conversational AI

Socio-emotional phenomena: catch-all term, that I will use here and that gathers both emotion, social stance, sentiment, mood, trust, engagement, stance, conversation strategies etc.

- **Machine learning models** of **socio-emotional phenomena** in **interactions**:

Scope: Socio-conversational AI

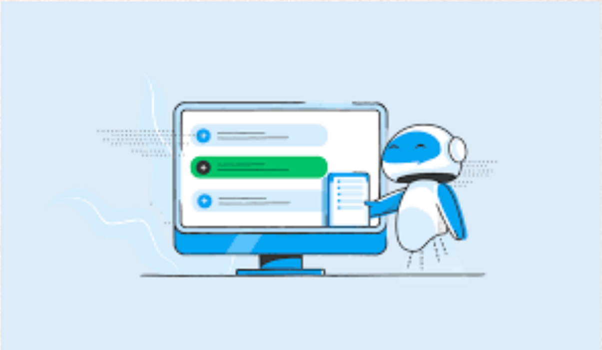
Socio-emotional phenomena: catch-all term, that I will use here and that gathers both emotion, social stance, sentiment, mood, trust, engagement, stance, conversation strategies etc.

- **Machine learning models** of **socio-emotional phenomena** in **interactions**:
  - Human-human (ex: social networks, interviews)

Scope: Socio-conversational AI

Socio-emotional phenomena: catch-all term, that I will use here and that gathers both emotion, social stance, sentiment, mood, trust, engagement, stance, conversation strategies etc.

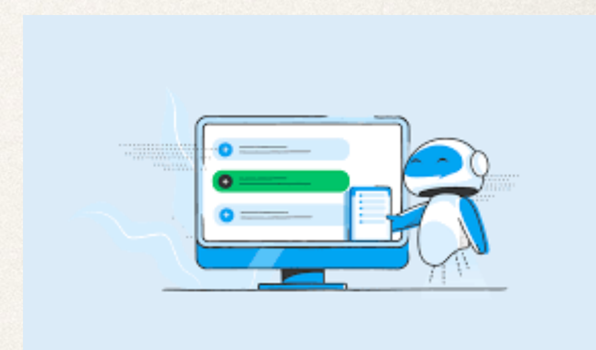
- **Machine learning models** of **socio-emotional phenomena** in **interactions**:
  - Human-human (ex: social networks, interviews)
  - Human-agent (ex: chatbot, voice assistant, robot, virtual characters)



Scope:
Socio-conversational AI

Socio-emotional phenomena: catch-all term, that I will use here and that gathers both emotion, social stance, sentiment, mood, trust, engagement, stance, conversation strategies etc.

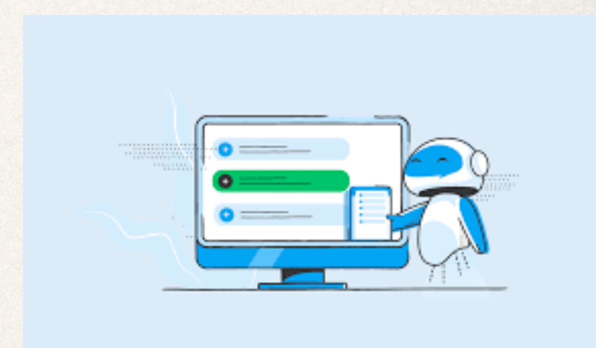
- **Machine learning models** of **socio-emotional phenomena** in **interactions**:
  - Human-human (ex: social networks, interviews)
  - Human-agent (ex: chatbot, voice assistant, robot, virtual characters)
- **Mono and Multimodal** models : **text**, **voice**, gestures, facial expressions, posture



Scope: Socio-conversational AI

Socio-emotional phenomena: catch-all term, that I will use here and that gathers both emotion, social stance, sentiment, mood, trust, engagement, stance, conversation strategies etc.

- **Machine learning models** of **socio-emotional phenomena** in **interactions**:
  - Human-human (ex: social networks, interviews)
  - Human-agent (ex: chatbot, voice assistant, robot, virtual characters)
- **Mono and Multimodal** models : **text**, **voice**, gestures, facial expressions, posture
- Models for the **analysis** and for the **generation**



Applications

Applications

Societal trend analysis in social networks: stance about vaccine for covid, fallacy detection

- **ANR chair NoRDF**

Applications

Societal trend analysis in social networks: stance about vaccine for covid, fallacy detection

- ANR chair NoRDF

AI for human skill improvement - public speaking training: automatic analysis of speech content to give feedback

- ANR Revitalise

Applications

Societal trend analysis in social networks: stance about vaccine for covid, fallacy detection
- ANR chair NoRDF

AI for human skill improvement - public speaking training: automatic analysis of speech content to give feedback
- ANR Revitalise

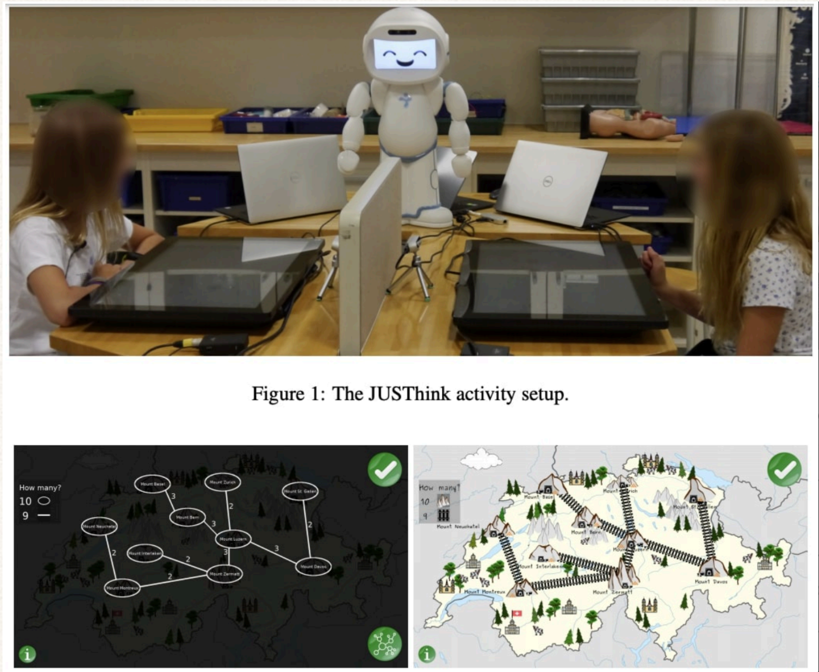
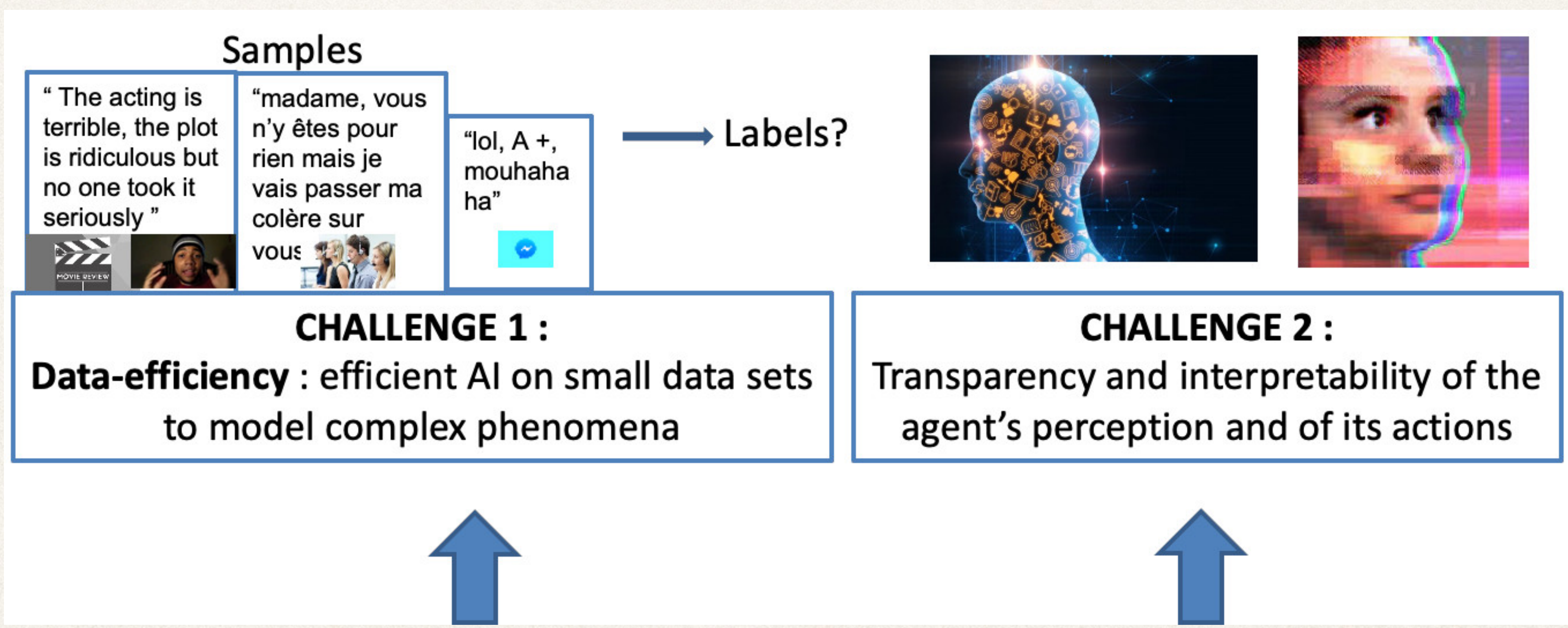


Figure 1: The JUSThink activity setup.

EDUCATION - social robots as partners of the learning process: automatic analysis of self-confidence
- European ITN ANIMATAS

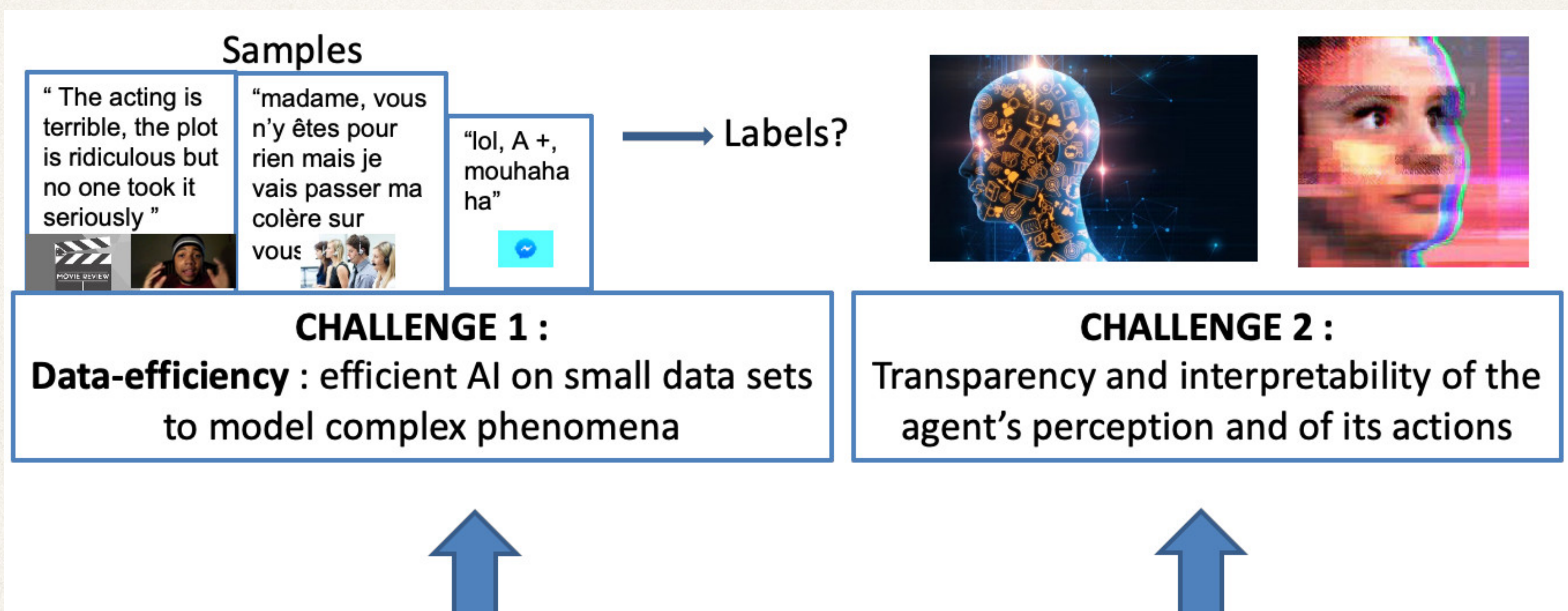
Scientific challenges

Scientific challenges



Scientific challenges

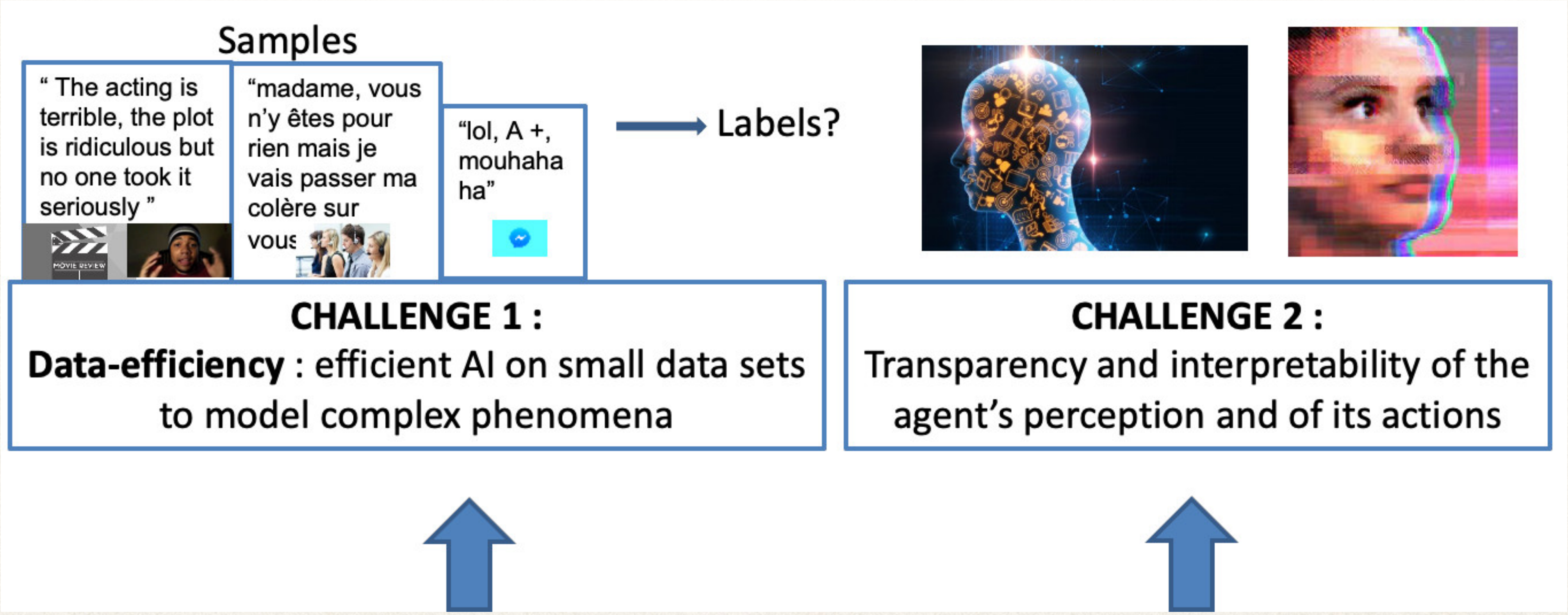
Socio-emotional phenomena (ex. trust, frustration, engagement , etc.)
are difficult to define and annotate +
difficult consensus



Scientific challenges

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

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



Scientific challenges

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

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

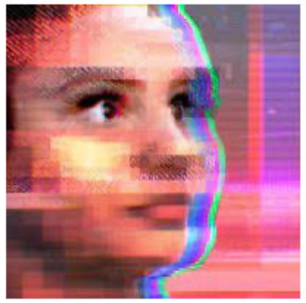
Samples

" The acting is terrible, the plot is ridiculous but no one took it seriously "

"madame, vous n'y êtes pour rien mais je vais passer ma colère sur vous"

"lol, A +, mouhaha ha"

→ Labels?



CHALLENGE 1 :
Data-efficiency : efficient AI on small data sets to model complex phenomena

CHALLENGE 2 :
Transparency and interpretability of the agent's perception and of its actions

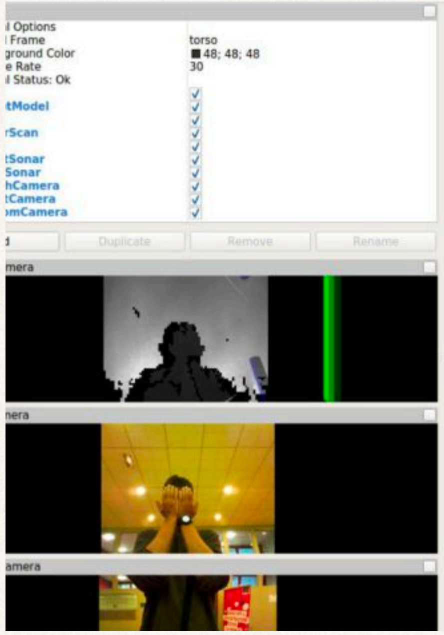
Our approach:
integrating human and social sciences at the heart of machine learning

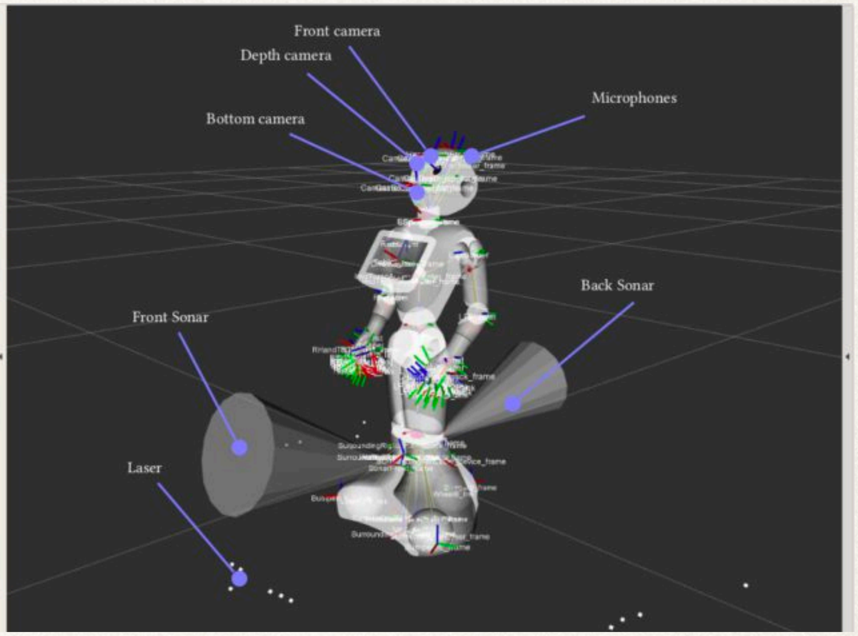
Two chapters in this presentation

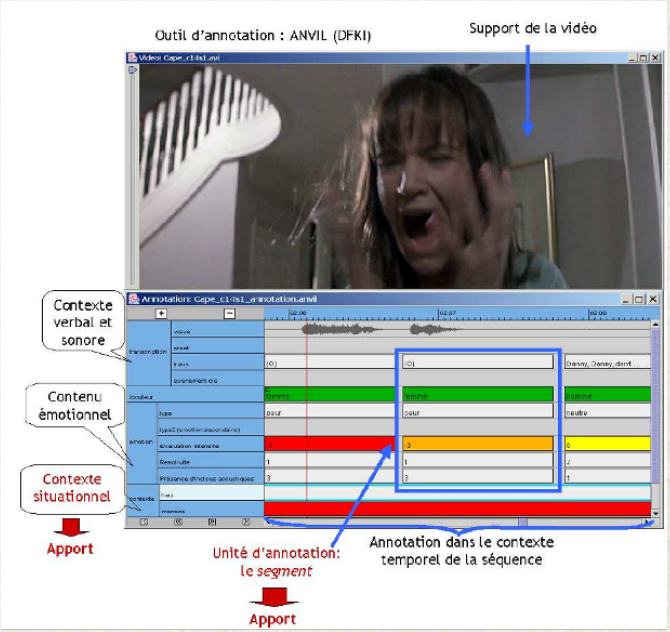
- Prologue: collecting and annotating data for supervised machine learning models
- Chapter 1: data/label efficient socio-emotional models
- Chapter 2: explainable socio-emotional neural models

Collecting new spontaneous socio-emotional data

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







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

Providing new coding scheme and annotation tools

Providing new coding scheme and annotation tools

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

Providing new coding scheme and annotation tools

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

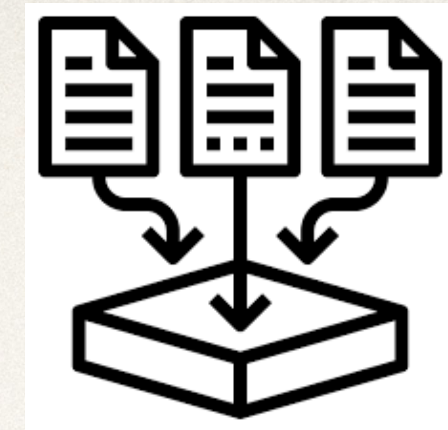
Rollet & Clavel. "Talk to you later" Doing social **robotics** with conversation analysis. Towards the development of an automatic system for the prediction of **disengagement**, Interaction Studies 2020

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

Text in Multimodal Data

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

Providing new coding scheme and annotation tools



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

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.

Text in Multimodal Data

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**. : **emotions** In LREC 2022.

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

Chhun, et al. Of Human Criteria and Automatic Metrics: A Benchmark of the Evaluation of **Story** Generation (HANNA), in COLING, 2022: **surprise, engagement**

Text only

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

Develop research in order to make the most of this labelled data

Chapter 1: data/label efficient socio-emotional models

e.g., transferring what has been learned
on a corpus of certain socio-emotional phenomena….
to other socio-emotional phenomena occurring in slightly different data.

Overview: data/label efficient socio-emotional models

Reasoning models

ex: agent's gesture generation [Ravenet et al., AAMAS 2018

Theoretical models from psychology, linguistics, conversation analysis
(ex. Cognitive models for gesture generation, linguistics for hybrid approaches)

Overview: data/label efficient socio-emotional models

Reasoning models

ex: agent's gesture generation [Ravenet et al., AAMAS 2018

Hybrid approaches

Encoding:

- **Linguistics-driven features [Raphalen et al., ACL 2022]**
- Pre-training objectives {Colombo et al., EMNLP 2021]

Decoding : model label dynamics [Chapuis et al., AAAI 2020]

Theoretical models from psychology, linguistics, conversation analysis
(ex. Cognitive models for gesture generation, linguistics for hybrid approaches)

Overview: data/label efficient socio-emotional models

Reasoning models	Hybrid approaches	Data augmentation
ex: agent's gesture generation [Ravenet et al., AAMAS 2018	Encoding: • **Linguistics-driven features [Raphalen et al., ACL 2022]** • Pre-training objectives {Colombo et al., EMNLP 2021] Decoding : model label dynamics [Chapuis et al., AAAI 2020]	Generating new data using: • Logical rules for entailment data [Helwé et al., F. EMNLP 2022] • Extreme value theory for rare sentiment data [Jalalzai et al., Neurips 2020]

Theoretical models from psychology, linguistics, conversation analysis
(ex. Cognitive models for gesture generation, linguistics for hybrid approaches)

Overview: data/label efficient socio-emotional models

Reasoning models

ex: agent's gesture generation [Ravenet et al., AAMAS 2018

Hybrid approaches

Encoding:

- **[Linguistics-driven features [Raphalen et al., ACL 2022]]**
- Pre-training objectives {Colombo et al., EMNLP 2021]

Decoding : model label dynamics [Chapuis et al., AAAI 2020]

Data augmentation

Generating new data using:

- Logical rules for entailment data [Helwé et al., F. EMNLP 2022]
- Extreme value theory for rare sentiment data [Jalalzai et al., Neurips 2020]

Transfer / Few-shot learning

- Weight transfer [Deng et al., ACII 2022] Guibon et al., AAAI 2023]
- Meta learning [Guibon et al., MetaLearn 2021]
- **[Prototypical Network [Guibon et al., EMNLP 2021]]**

Theoretical models from psychology, linguistics, conversation analysis
(ex. Cognitive models for gesture generation, linguistics for hybrid approaches)

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 dictionary)

Knowledge-Driven Features (KDF)

Rule-based *vs.* **Hybrid model** *vs.* BERT fine-tuned?

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

Rule-based *vs.* **Hybrid model** *vs.* BERT fine-tuned?

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

Hybrid model: KDF + ML models

Rule-based *vs.* **Hybrid model** *vs.* BERT fine-tuned?

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

Hybrid model: KDF + ML models

Data: peer-tutoring interactions (23000 utterances)

Rule-based *vs.* Hybrid model *vs.* BERT fine-tuned?

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

Hybrid model: KDF + ML models

Data: peer-tutoring interactions (23000 utterances)

SentBERT

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

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

Rule-based *vs.* Hybrid model *vs.* BERT fine-tuned?

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

Hybrid model: KDF + ML models

Data: peer-tutoring interactions (23000 utterances)

SentBERT

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

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

Take home message:

When we want to detect a phenomenon that is well known by linguists but not available in a sufficient quantity in existing labelled corpora, using hybrid model makes the most sense!

Gaël Guibon, Matthieu Labeau, Hélène Flamein, Luce Lefeuvre and Chloé Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

Focus on few-shot learning for data/label-efficiency

Gaël Guibon, Matthieu Labeau, Hélène Flamein, Luce Lefeuvre and Chloé Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

Focus on few-shot learning for data/label-efficiency

- Task:
  - detect **emotions** and their **evolution in a conversation flow** (sequence labeling)

Focus on few-shot learning for data/label-efficiency

- Task:
  - detect **emotions** and their **evolution in a conversation flow** (sequence labeling)
- Data: a live chat customer service

Operator: Did you make the simulation using the promo code?
Visitor: I did it 5 minutes ago
Operator: Ok, you have to wait 30min Visitor: but as said before, I didn't finished the "simulation" because I had to pay a 10€ ticket even th
Visitor: ....even though the right one is 11.5€
Operator: And the code will be available again

Gaël Guibon, Matthieu Labeau, Hélène Flamein, Luce Lefeuvre and Chloé Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

Focus on few-shot learning for data/label-efficiency

- Task:
  - detect **emotions** and their **evolution in a conversation flow** (sequence labeling)
- Data: a live chat customer service
  - **Language specificities** (unfinished sentences, specific lexical field)

Operator: Did you make the simulation using the promo code?
Visitor: I did it 5 minutes ago
Operator: Ok, you have to wait 30min Visitor: but as said before, I didn't finished the "simulation" because I had to pay a 10€ ticket even th
Visitor: ....even though the right one is 11.5€
Operator: And the code will be available again

Focus on few-shot learning for data/label-efficiency

- Task:
  - detect **emotions** and their **evolution in a conversation flow** (sequence labeling)
- Data: a live chat customer service
  - **Language specificities** (unfinished sentences, specific lexical field)
  - Only **a few data are labelled** in emotions (1000 conversations)

Operator: Did you make the simulation using the promo code?
Visitor: I did it 5 minutes ago
Operator: Ok, you have to wait 30min Visitor: but as said before, I didn't finished the "simulation" because I had to pay a 10€ ticket even th
Visitor: ....even though the right one is 11.5€
Operator: And the code will be available again

Focus on few-shot learning for data/label-efficiency

- Task:
  - detect **emotions** and their **evolution in a conversation flow** (sequence labeling)
- Data: a live chat customer service
  - **Language specificities** (unfinished sentences, specific lexical field)
  - Only **a few data are labelled** in emotions (1000 conversations)
- Objective:
  - **Train a model with only a few set of annotated samples**

Operator: Did you make the simulation using the promo code?
Visitor: I did it 5 minutes ago
Operator: Ok, you have to wait 30min Visitor: but as said before, I didn't finished the "simulation" because I had to pay a 10€ ticket even th
Visitor: ....even though the right one is 11.5€
Operator: And the code will be available again

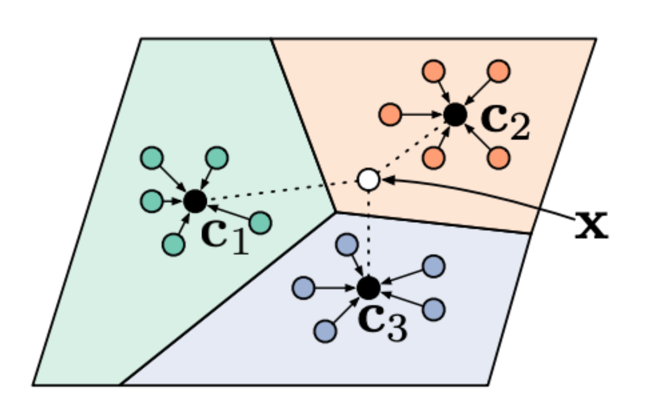
Few-shot learning using **Prototypical networks**

G. Guibon, M. Labeau, H. Flamein, L. Lefeuvre and C. Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

Few-shot learning using **Prototypical networks**

G. Guibon, M. Labeau, H. Flamein, L. Lefeuvre and C. Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

- **Prototypical networks** - learns a metric space in which classification can be performed by computing distances to prototype representations of each class.



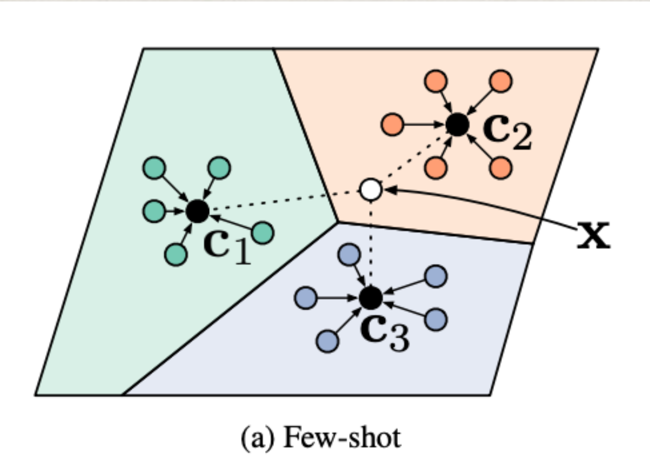
(a) Few-shot

From Snell, 2017

Few-shot learning using **Prototypical networks**

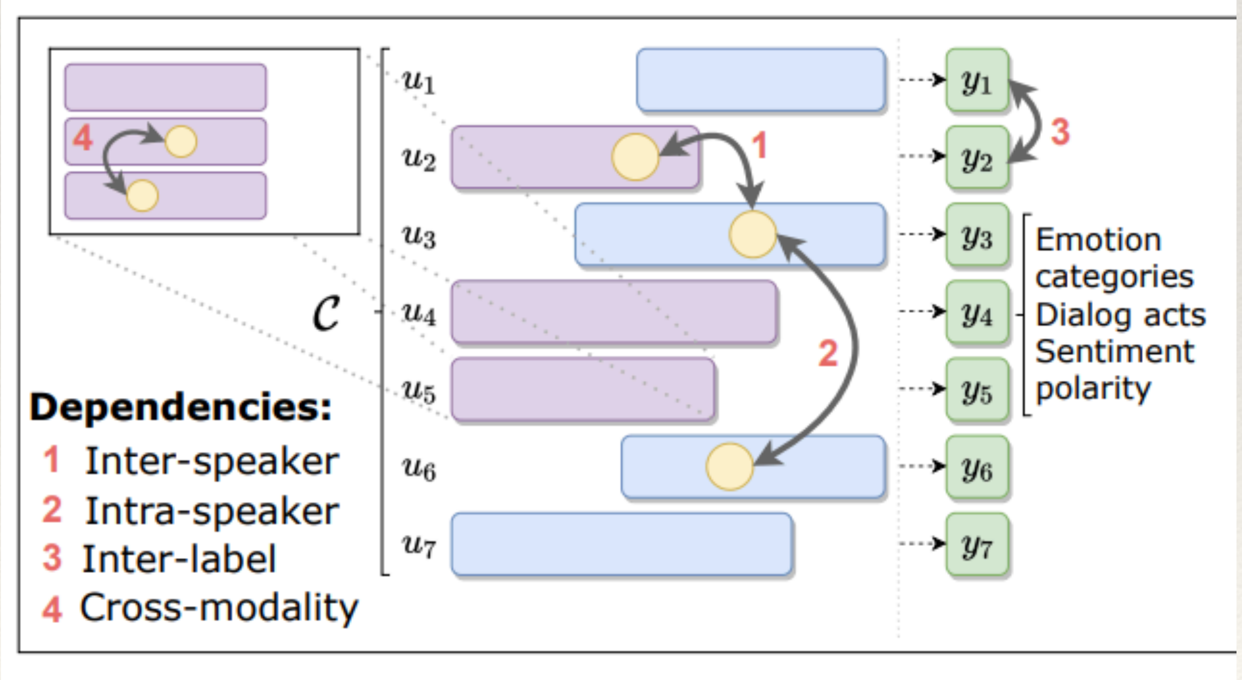
G. Guibon, M. Labeau, H. Flamein, L. Lefeuvre and C. Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

- **Prototypical networks** - learns a metric space in which classification can be performed by computing distances to prototype representations of each class.



From Snell, 2017

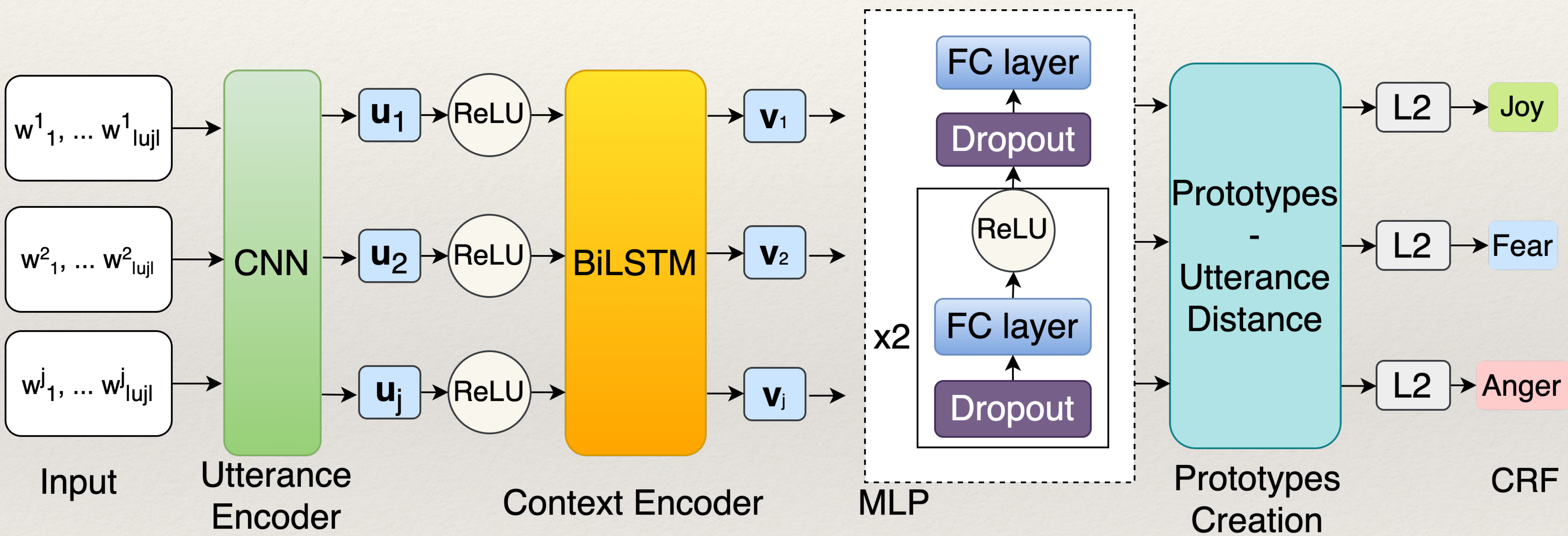
- **Leverage Social Science:** a conversation is a co-construction over time by two or more interlocutors [Clark, 1996, Schegloff, 2007]-> **ProtoSeq:** integrate **conversational dynamics** when building prototypes (utterance and emotional label dependencies)



Conversational dynamics as pictured in [Clavel, Labeau, Cassell, 2022]

ProtoSeq: integrate conversational dynamics

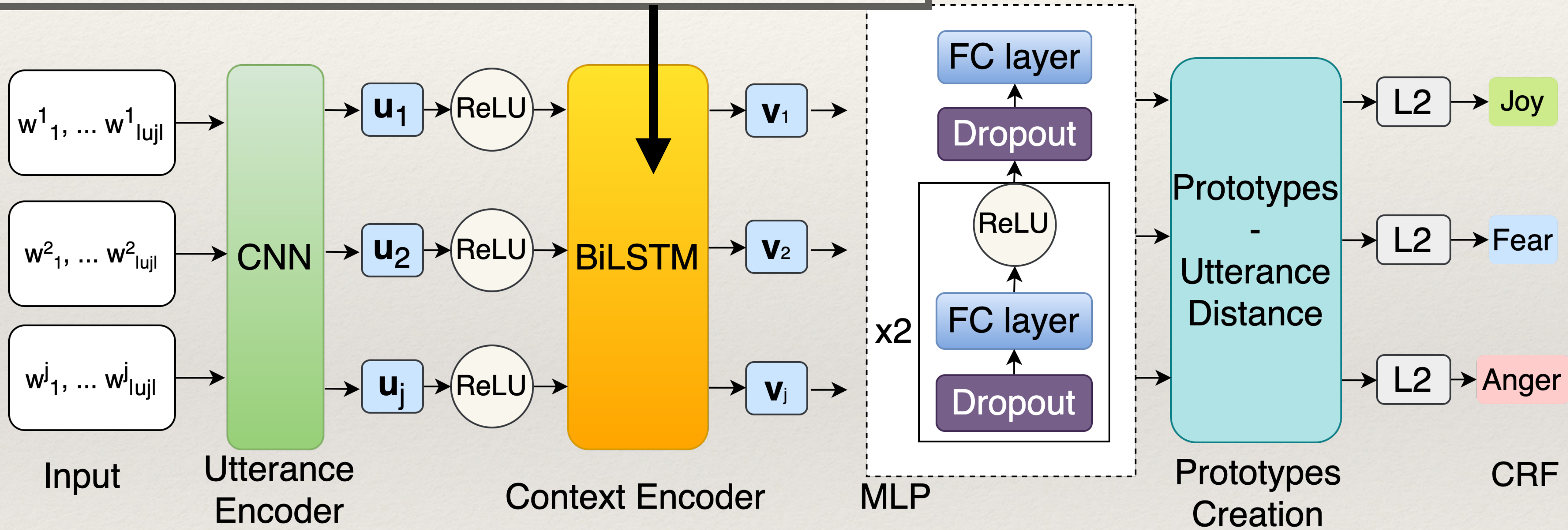
G. Guibon, M. Labeau, H. Flamein, L. Lefeuvre and C. Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)



ProtoSeq: integrate conversational dynamics

G. Guibon, M. Labeau, H. Flamein, L. Lefeuvre and C. Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

Utterance representations by capturing **dependencies with surrounding context** using **Bi-LSTM**

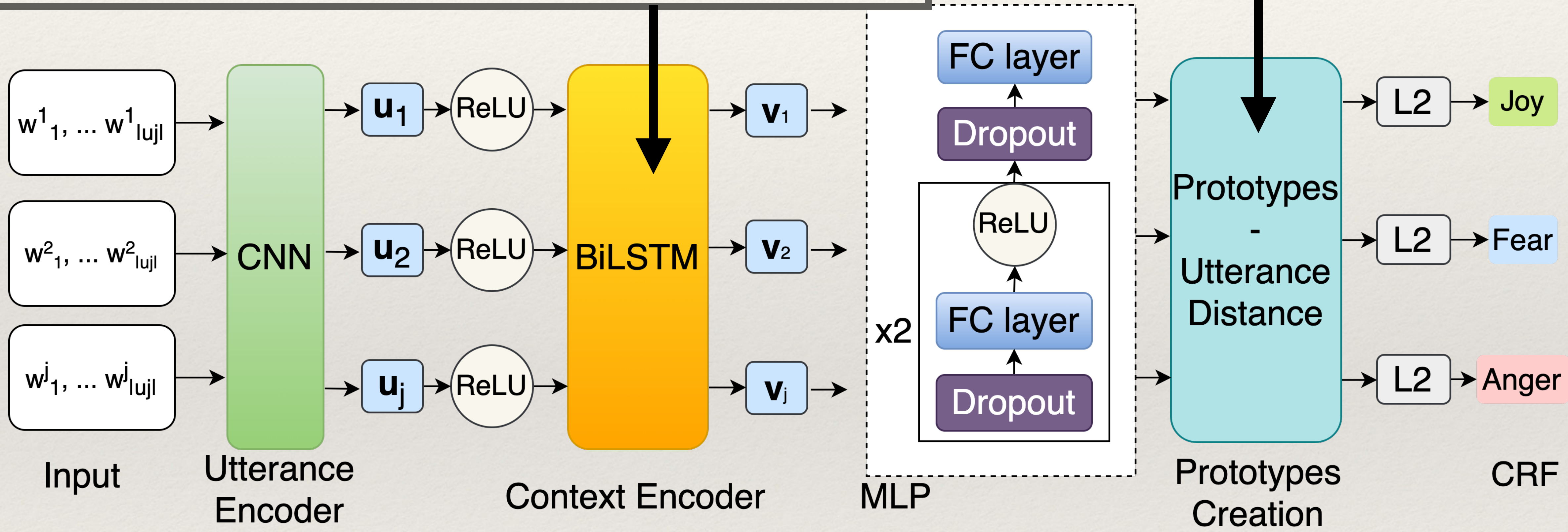


ProtoSeq: integrate conversational dynamics

G. Guibon, M. Labeau, H. Flamein, L. Lefeuvre and C. Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

Utterance representations by capturing **dependencies with surrounding context** using **Bi-LSTM**

$$\mathbf{c}_k \leftarrow \frac{1}{N_{\mathcal{C}}} \sum_{(u_j, y_j) \text{ with } y_j = k} MLP(\mathbf{v}_j)$$

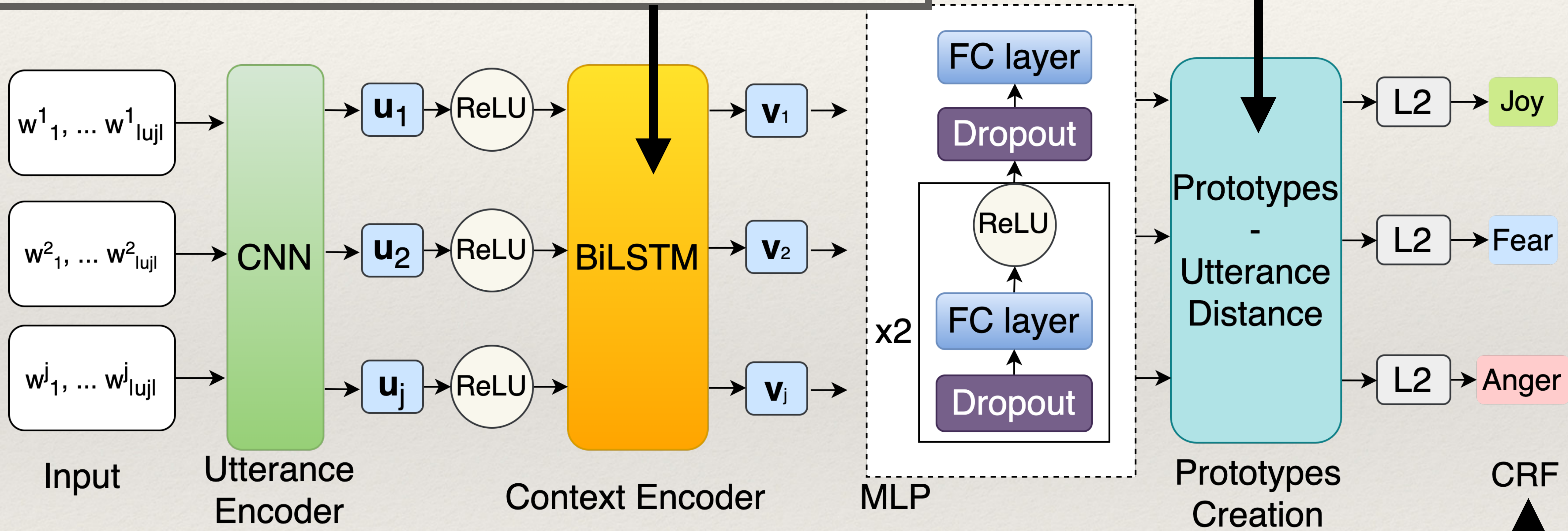


ProtoSeq: integrate conversational dynamics

G. Guibon, M. Labeau, H. Flamein, L. Lefeuvre and C. Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

Utterance representations by capturing **dependencies with surrounding context** using **Bi-LSTM**

$$\mathbf{c}_k \leftarrow \frac{1}{N_{\mathcal{C}}} \sum_{(u_j,y_j) \text{ with } y_j=k} MLP(\mathbf{v}_j)$$



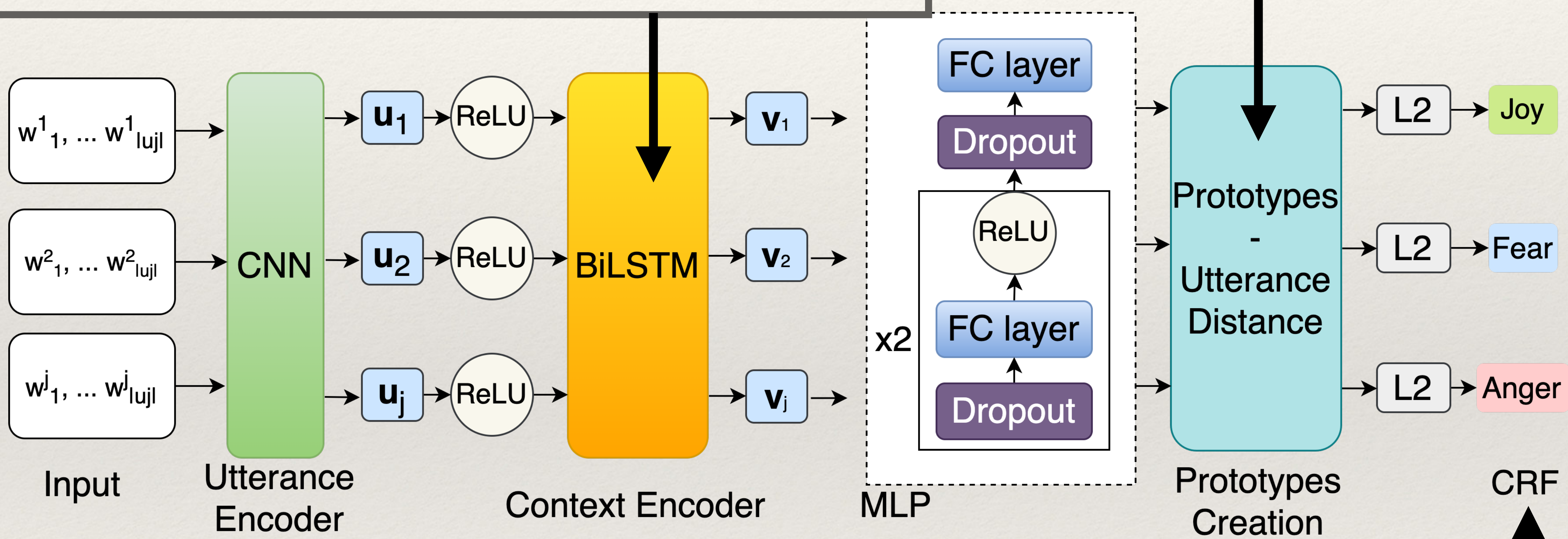
Emotion label dependencies: CRF layer on top of label prediction

ProtoSeq: integrate conversational dynamics

G. Guibon, M. Labeau, H. Flamein, L. Lefeuvre and C. Clavel, Few-Shot Emotion Recognition in Conversation with Sequential Prototypical Networks, EMNLP (2021)

Utterance representations by capturing **dependencies with surrounding context** using **Bi-LSTM**

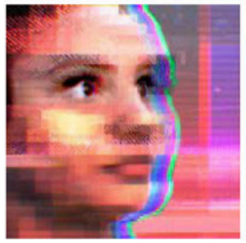
$$\mathbf{c}_k \leftarrow \frac{1}{N_{\mathcal{C}}} \sum_{(u_j, y_j) \text{ with } y_j = k} MLP(\mathbf{v}_j)$$



Results: Sequential Protypical Networks : achieves 31.8% in micro f1-score (10 classes) compared to 26.1% (SOTA prototypical network method)

Emotion label dependencies: CRF layer on top of label prediction

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



CHALLENGE 2 :
Transparency and interpretability of the agent's perception and of its actions

Chapter 2: explainable socio-emotional neural models

explain the rationales behind the prediction made by neural models

Overview: explainable socio-emotional neural models

Overview: explainable socio-emotional neural models

Post-modelling
explainability: dissect the model

Overview: explainable socio-emotional neural models

Post-modelling explainability: dissect the model

- SHAP : analysis of features that matter for hedge detection [Raphalen et al., ACL 2022]
- **Analysis of attention mechanisms of neural networks in order to identify *attention slices* [Hemamou et al., Trans. on Aff. Comp., 2021]**

Overview: explainable socio-emotional neural models

Post-modelling explainability: dissect the model

- SHAP : analysis of features that matter for hedge detection [Raphalen et al., ACL 2022]
- **Analysis of attention mechanisms of neural networks in order to identify *attention slices* [Hemamou et al., Trans. on Aff. Comp., 2021]**

↓

Outputs interpreted from literature of psychology, linguistics, conversation analysis

Overview: explainable socio-emotional neural models

Post-modelling explainability: dissect the model

« BERTology » : Analyzing BERT pre-trained representations

- SHAP : analysis of features that matter for hedge detection [Raphalen et al., ACL 2022]
- **Analysis of attention mechanisms of neural networks in order to identify *attention slices* [Hemamou et al., Trans. on Aff. Comp., 2021]**

Outputs interpreted from literature of psychology, linguistics, conversation analysis

Overview: explainable socio-emotional neural models

Post-modelling explainability: dissect the model

« BERTology » : Analyzing BERT pre-trained representations

- SHAP : analysis of features that matter for hedge detection [Raphalen et al., ACL 2022]
- **Analysis of attention mechanisms of neural networks in order to identify *attention slices* [Hemamou et al., Trans. on Aff. Comp., 2021]**

- Information about fillers [Dinkar et al., EMNLP 2020]
- **Information about stances [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

[Attention is not not Explanation]
(Wiegreffe & Pinter, EMNLP-IJCNLP 2019)

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

Approach: use a prediction model to try to understand the rationales behind the recruiters' decision by assuming that the prediction model mimics the recruiters' decision

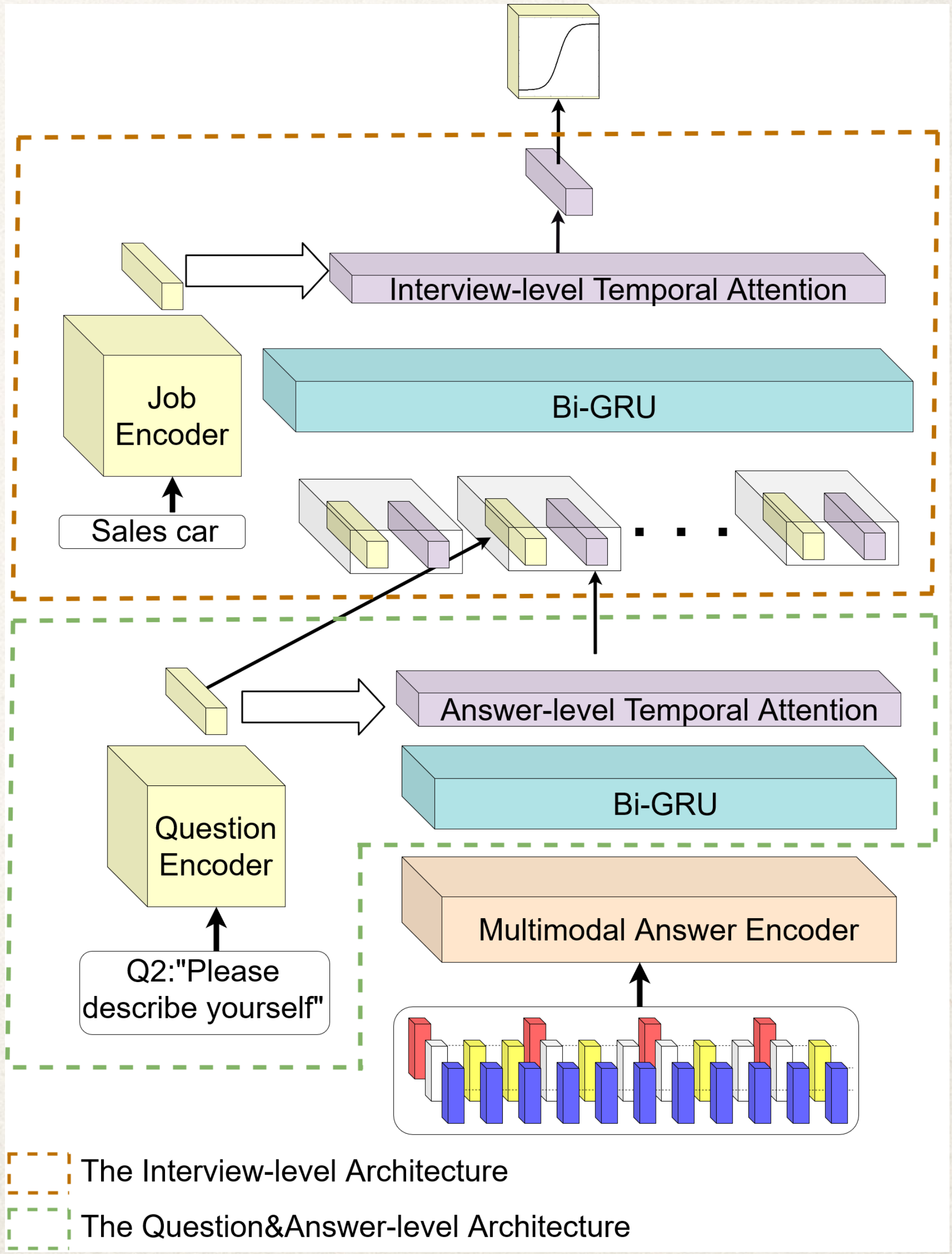
Attention slices for explainability

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

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

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

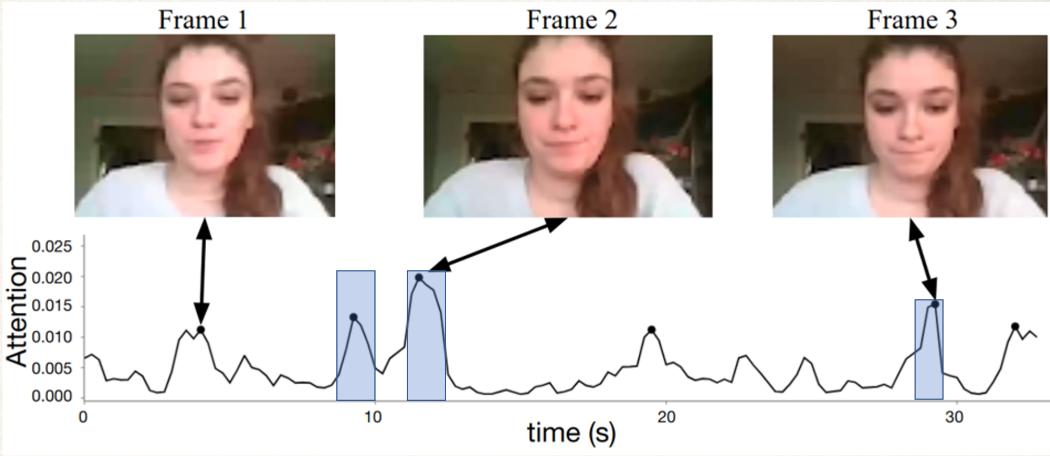


Attention slices for explainability

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

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

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



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

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

Step 2 - study attention mechanisms 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 And contain breathing, fillers, activation of some action units (confusion and emphasis), and specific vocabulary linked to competencies)

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

Final step : check whether it is consistent to what was found in human resource literature.

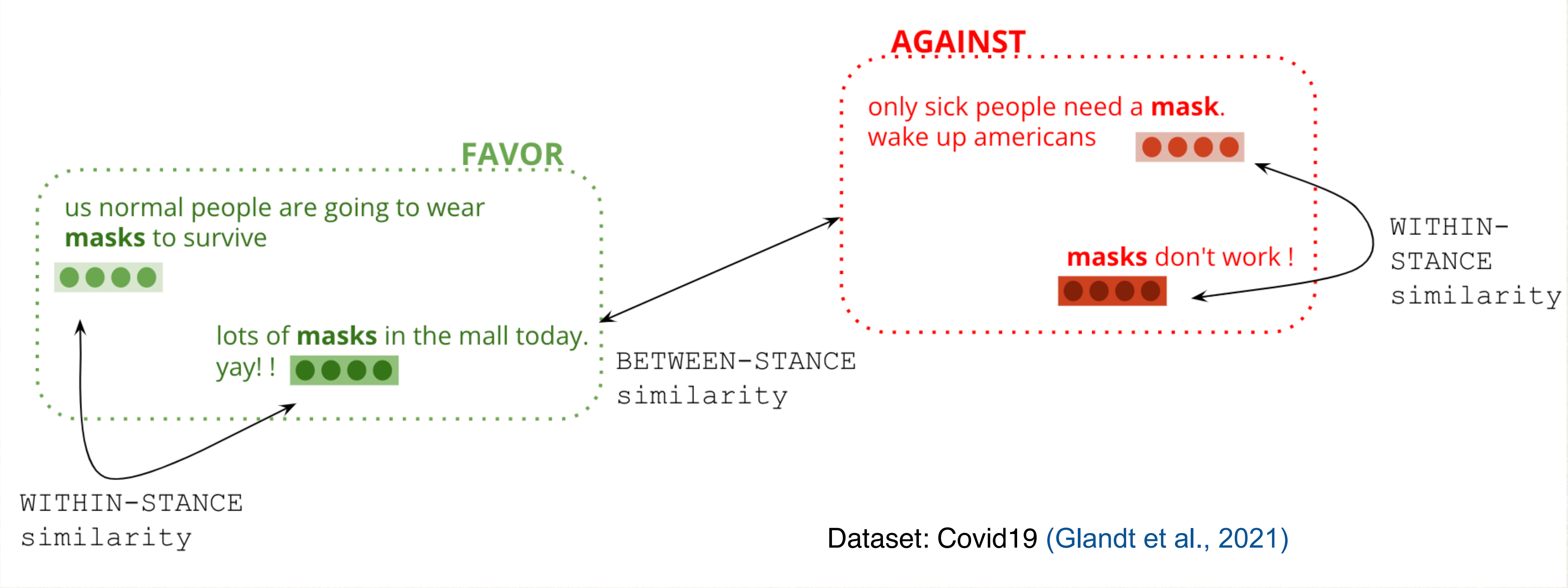
Take home message : *a step towards explainability -> we can try to dissect a model. It gives some interesting information to try to understand the decision BUT this is very local and we can not completely retrace the decision process such as it could be done when using reasoning models*

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:



$$sim(P,Q) = \frac{\sum_{w \in V_{PQ}} cos(\mathbf{w}_P, \mathbf{w}_Q)}{|V_{PQ}|}$$

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 stance expressed ?

Conclusions:

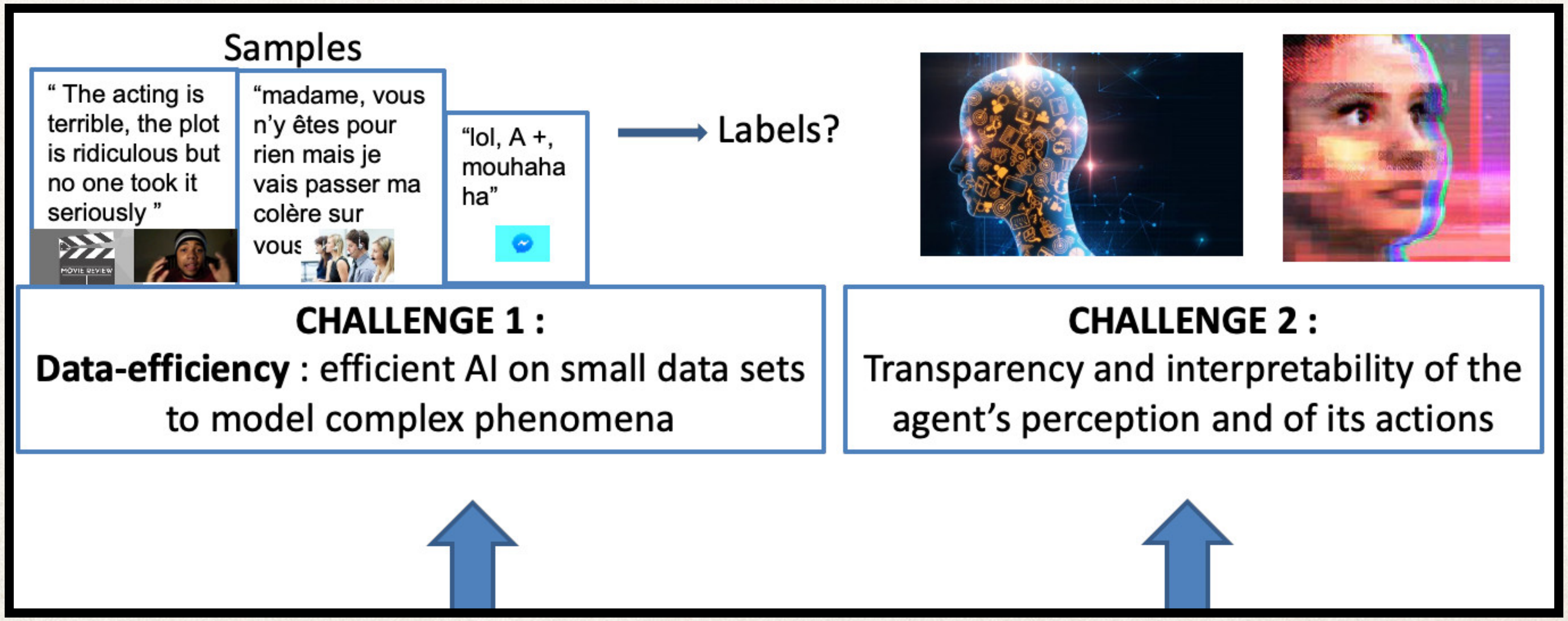
- Differences in similarity between concurring and conflicting stances are small, but significant.
- Words with the highest differences tend to be central to the topic: potentially useful for detecting points of discordance.

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/Take home message

My research: develop machine learning models for detecting and generating socio-emotional behaviors

My perspective: make benefit of social science research in order to contribute to performant, **tractable** and **explainable** neural models.



Epilogue/Take home message

The different ways of leveraging social science

Social Science

1. In the supervision of machine learning models (delineating the targeted socio-emotional behavior + build robust annotation scheme)
2. In the design of features used by machine learning models (ex: linguistic knowledge for hedge prediction)
3. In the design of transfer / few-shot learning approaches (ex: conversational dynamics in Protypical Network with ProtoSeq)
4. For the interpretation of the models , confronting the social science discovery to what the analysis of neural prediction models is showing (ex: attention slices and job interview analysis)

Thank you !

Collaborators who have contributed to the studies presented here (in the order of appearance):

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

Other mentioned studies:
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), Luce Lefeuvre (SNCF), Tanvi Dinkar (ex PhD student), …

Questions ?