Emotion Recognition in Textual Conversations

Empowering NLP from IQ to EQ

Julius Berner Philipp Möhl

February 18, 2020

University of Vienna, A1 Telekom Austria Group, Brainference

Table of contents

Motivation - Why do we need emotion recognition?

Left and Right Brain

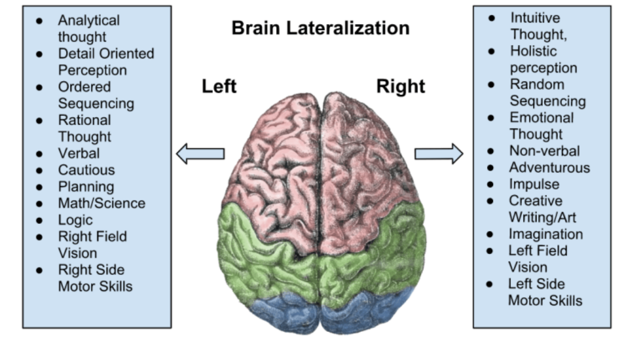


Figure 1: Brain Lateralization[9]

Example Customer Requests - Conversation Bot

Sadly my Mediabox delivery didn't contain a remote control. - SAD

I have got my Mediabox delivery. Is the remote control not contained anymore? - CURIOUS

You did not send me the remote control for my Mediabox! - ANGER

Machine Personas - Conversation Bot

We are sorry for your inconvenience. Please tell us your ID so we can help you. - **medium potential of cross selling basic product**

Please tell us your ID so we can look that up for you. - **high potential of cross selling premium product**

We are sorry that you are experiencing this inconvenience. Please tell us your ID so we can solve this issue for you. - **low potential of cross selling**

Assistance and Conversational Bots

Conversation with Meena	
Human:	*I like how cute cows are*
Meena:	*Ah, interesting. You like how they are cute, right?*
Human:	*yeah, and they're realy smart*
Meena:	*I heard they go to college*
Human:	*cows go to college?*
Meena:	*I heard that a cow went to Harvard.*
Human:	*what did the cow study?*
Meena:	*Bovine sciences.*
Human:	*do horses go to Harvard?*
Meena:	*Horses go to Hayvard.*
Human:	*that's a pretty good joke, I feel like you led me into it*
Meena:	*You were trying to steer it elsewhere, I can see it.*

Table 1: Meena by Google[8]

Content generation

The spaceship was flying erratically, something was wrong... The force of the anomaly soon halted the rocket's fall and caused the engine to grind to a halt, while the ship's hull began to fracture from the side as the force of the anomaly turned the shield generator into a debris of debris. The ship's contents settled. "You're fine, I'm sure..." "I can hear your heartbeat, you're fine, but it might take a moment for it to reach my ears..." As they sat, a passing shuttle floated past. Or the ones we were on, at least. Did it know about us? The missile fired right through the shuttle. They reached for the pilot, and realized their presence was painfully obvious. A blast ripped through the shuttle, and the pilot went airborne, crashing.

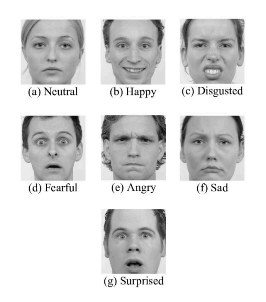
Source: Talk to Transformers[7]

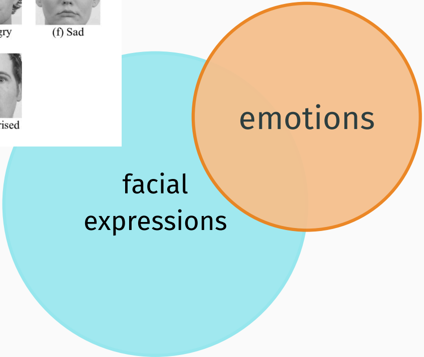
Challenges - Why textual emotion recognition is a hard task?

Channels and Data Sources [5, 11]

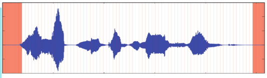
emotions

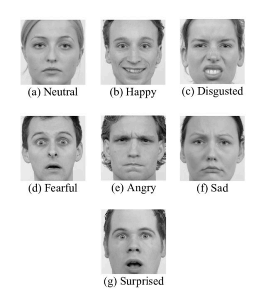
Channels and Data Sources [5, 11]


(a) Neutral (b) Happy (c) Disgusted
(d) Fearful (e) Angry (f) Sad
(g) Surprised



Channels and Data Sources [5, 11]



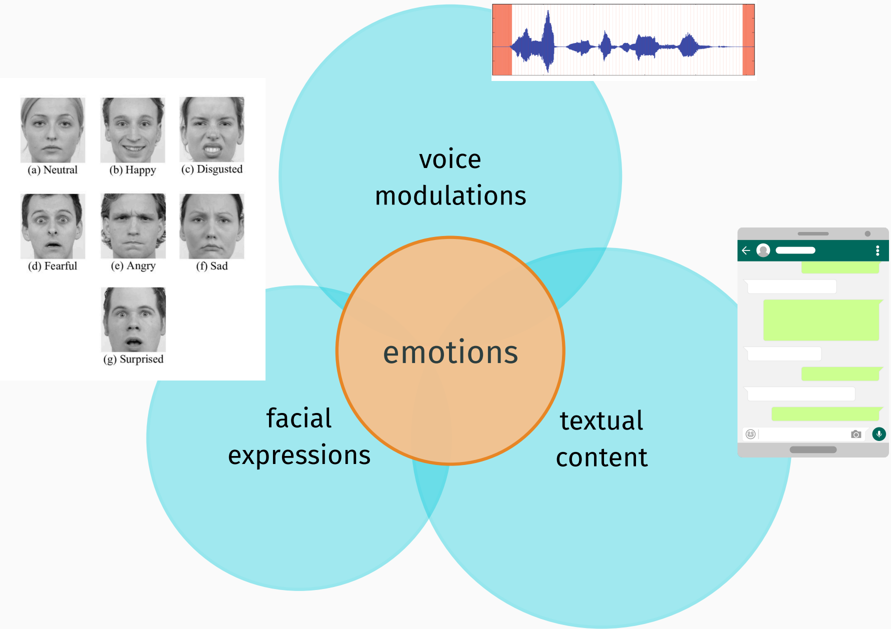


voice modulations

emotions

facial expressions

Channels and Data Sources [5, 11]



Emotion Model

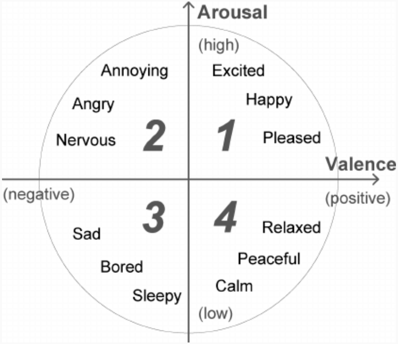


Figure 2: 2D valence-arousal emotion space - Russell 1980 [14]

dimensional model vs. categorical model

fine-grained model vs. availability of data

Emotion Model

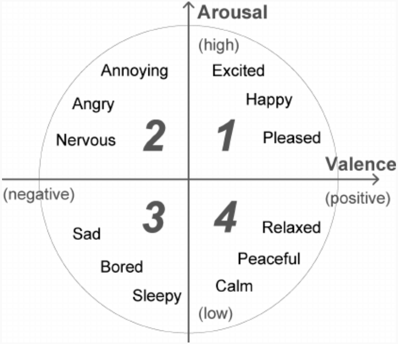


Figure 2: 2D valence-arousal emotion space - Russell 1980 [14]

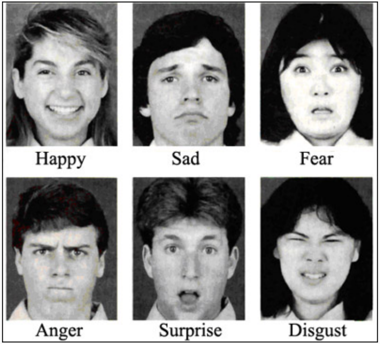


Figure 3: Six Basic Emotions - Ekman 1992 [3]

dimensional model vs. categorical model

fine-grained model vs. availability of data

Emotion Model

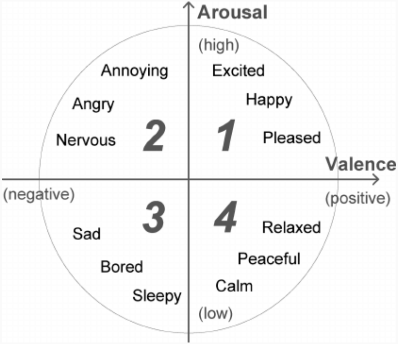


Figure 2: 2D valence-arousal emotion space - Russell 1980 [14]

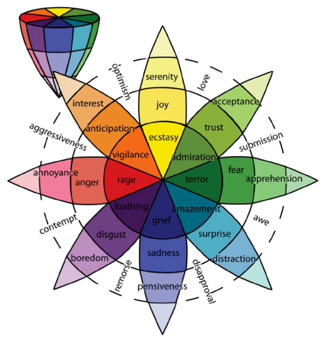


Figure 3: Wheel of emotions - Plutchik 1980 [12]

dimensional model vs. categorical model

fine-grained model vs. availability of data

Challenges - Context

[...]

A: Trump won the election?

B: Yeah!

Challenges - Context

[...]

A: Trump won the election?

B: Yeah!

Challenges - Context

[...]

A: Trump won the election?

B: Yeah!

Challenges - Context

[...] ← context is crucial for emotion recognition

A: Trump won the election?

B: Yeah!

Challenges - Datasets

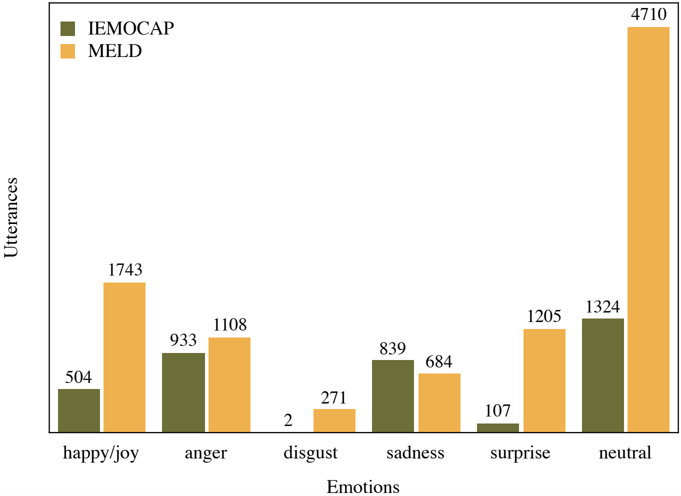


Figure 4: MELD and IEMOCAP multimodal multi-party datasets [13]

Challenges - Emotion Shifts

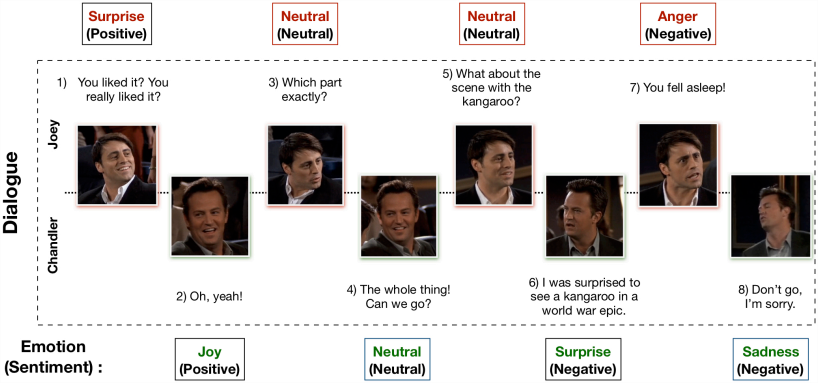


Figure 5: example dialogue from MELD dataset [10]

Current State - What solutions exist?

Theoretical Framework

Data

- sequence $[(u_1, p_1), (u_2, p_2), ...]$
  - utterances u_i
  - spoken by party p_i
- perhaps: background information
  - personalities
  - topic

Task

- identify emotion e_i of each utterance u_i
  - $e_i \in \{E_1, \ldots, E_N\}$
  - $e_i \in \mathbb{R}^d$

real time classification vs. beneficial use of future utterances

Theoretical Framework

Data

- sequence $[(u_1, p_1), (u_2, p_2), ...]$
  - utterances u_i
  - spoken by party p_i
- perhaps: background information
  - personalities
  - topic

Task

- identify emotion e_i of each utterance u_i
  - $e_i \in \{E_1, \ldots, E_N\}$
  - $e_i \in \mathbb{R}^d$

real time classification vs. beneficial use of future utterances

Theoretical Framework

Data

- sequence $[(u_1, p_1), (u_2, p_2), ...]$
  - utterances u_i
  - spoken by party p_i
- perhaps: background information
  - personalities
  - topic

Task

- identify emotion e_i of each utterance u_i
  - $e_i \in \{E_1, \ldots, E_N\}$
  - $e_i \in \mathbb{R}^d$

real time classification vs. beneficial use of future utterances

SemEval 2019 Task 3

	SemEval 2019 Task 3 - Data example
User:	*I just qualified for the Nabard internship*
Conversational Agent:	*WOOT! Thats great news. Congratulations!*
User:	*I started crying*
Label:	*happy*

Table 2: Examples showing influence of context in determining emotion of last utterance[1]

Context Aware Bert

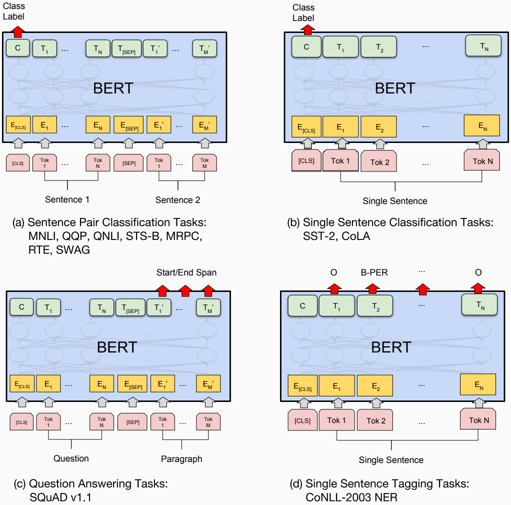


Figure 6: Bert Fine-tuning[2]

DialogueGCN - Context Approach

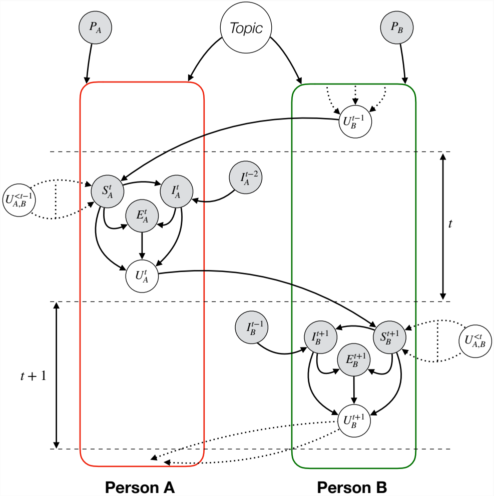


Figure 7: P represents personality,U represents utterance, S represents interlocutor state, I represents interlocutor intent, E represents emotion and Topic represents topic of the conversation[4]

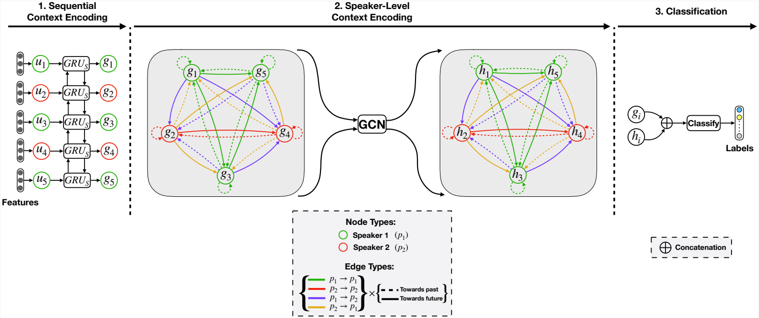


Figure 8: DialogueGCN Model, [4]

DialogueGCN - 3 Parts

DialogueGCN [4]

3 Parts

- Sequential Context Encoding ← (Glove with 1D CNN to Features & Bi-directional RNN)
- Speaker-Level Context Encoding ← (2-layer GCN, $f(H^i, A) = \sigma(AH^iW^i)$)
- Classification ← (Similarity-based Attention & Fully Connected)

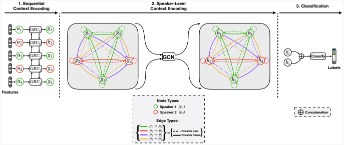


Figure 9: Model Architecture

DialogueGCN - 3 Parts

DialogueGCN [4]

3 Parts

- Sequential Context Encoding $\leftarrow$ (Glove with 1D CNN to Features & Bi-directional RNN)
- Speaker-Level Context Encoding $\leftarrow$ (2-layer GCN, $f(H^i, A) = \sigma(AH^iW^i)$)
- Classification $\leftarrow$ (Similarity-based Attention & Fully Connected)

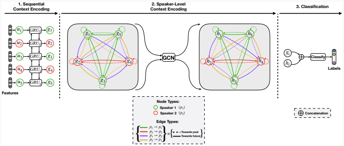


Figure 9: Model Architecture

DialogueGCN - 3 Parts

DialogueGCN [4]

3 Parts

- Sequential Context Encoding $\leftarrow$ (Glove with 1D CNN to Features & Bi-directional RNN)
- Speaker-Level Context Encoding $\leftarrow$ (2-layer GCN, $f(H^i, A) = \sigma(AH^iW^i)$)
- Classification $\leftarrow$ (Similarity-based Attention & Fully Connected)

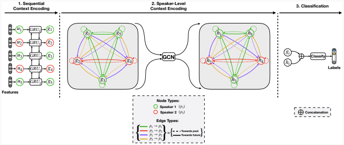


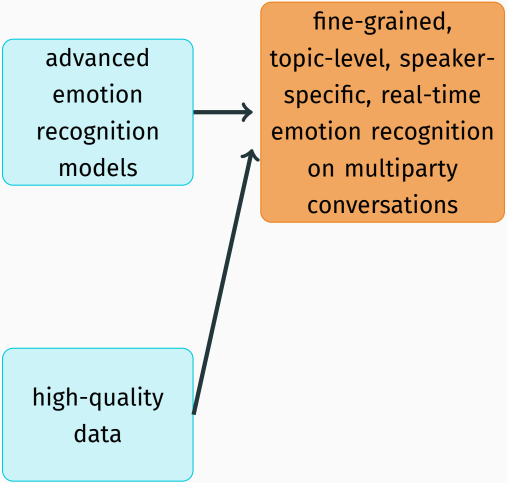
Figure 9: Model Architecture

Conclusion - What are future perspectives?

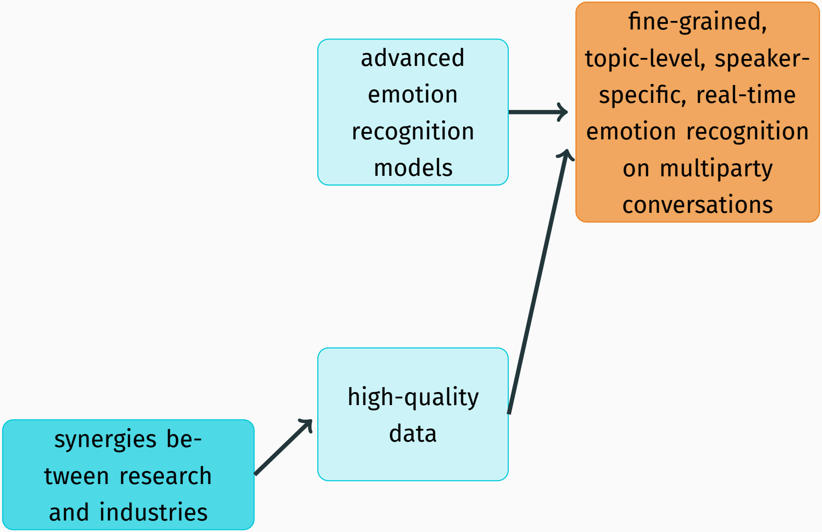
The future of emotion recognition

fine-grained, topic-level, speaker-specific, real-time emotion recognition on multiparty conversations

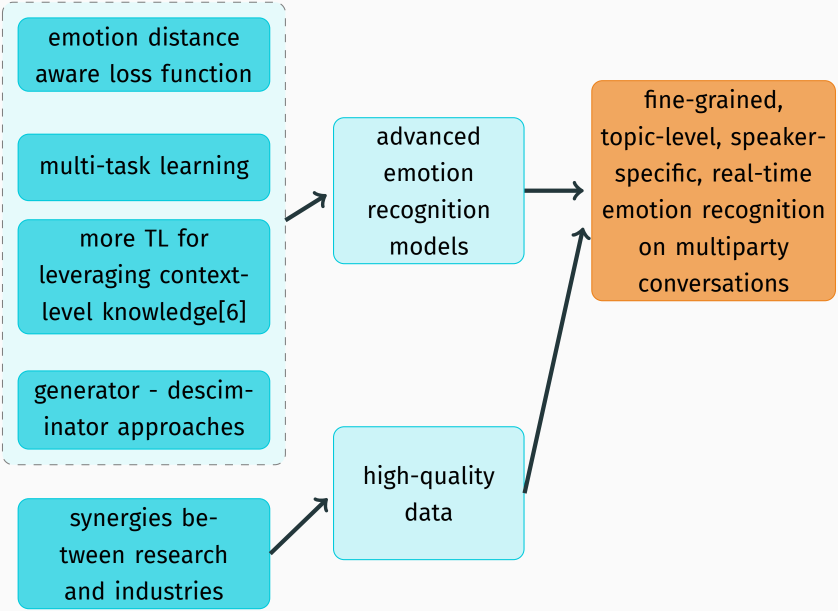
The future of emotion recognition



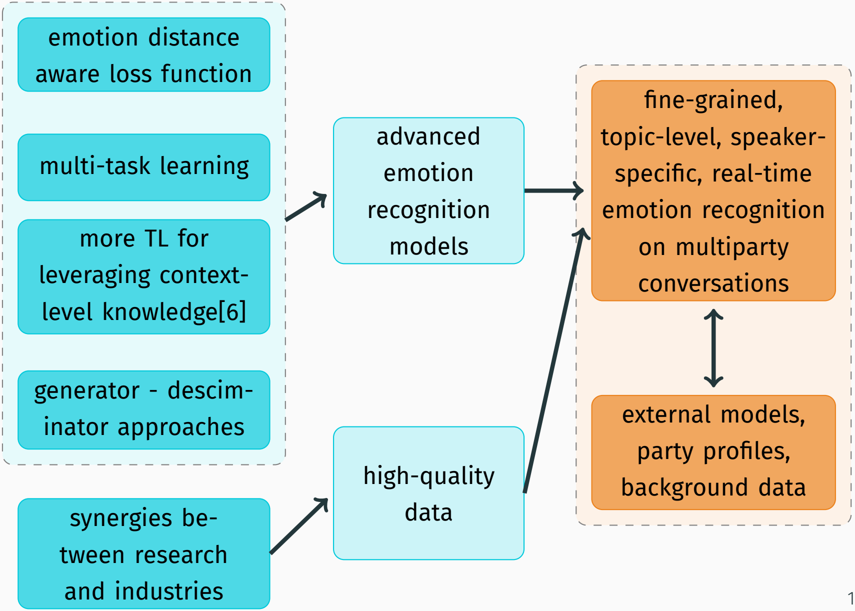
The future of emotion recognition



The future of emotion recognition



The future of emotion recognition



Let’s stay in touch

Philipp Möhl
linked.in/philipp-möhl-1b551b163

Julius Berner
linked.in/julius-berner

Questions?

References

[1] A. Chatterjee, K. N. Narahari, M. Joshi, and P. Agrawal. "SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text". In: *Proceedings of the 13th International Workshop on Semantic Evaluation*. Minneapolis, Minnesota, USA: Association for Computational Linguistics, June 2019, pp. 39–48. DOI: `10.18653/v1/S19-2005`. URL: `https://www.aclweb.org/anthology/S19-2005`.

[2] J. Devlin, M. Chang, K. Lee, and K. Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *CoRR* abs/1810.04805 (2018). arXiv: `1810.04805`. URL: `http://arxiv.org/abs/1810.04805`.

[3] P. Ekman. "Basic emotions". In: *Handbook of cognition and emotion* 98.45-60 (), p. 16.

[4] D. Ghosal, N. Majumder, S. Poria, N. Chhaya, and A. Gelbukh. "DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation". In: *EMNLP/IJCNLP*. 2019.

[5] P. Harár, R. Burget, and M. K. Dutta. "Speech emotion recognition with deep learning". In: *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*. 2017, pp. 137–140. DOI: `10.1109/SPIN.2017.8049931`.

[6] D. Hazarika, S. Poria, R. Zimmermann, and R. Mihalcea. "Emotion Recognition in Conversations with Transfer Learning from Generative Conversation Modeling". In: *ArXiv* abs/1910.04980 (2019).

[7] A. D. King. "Talk to Transformers". In: 2019. URL: `https://talktotransformer.com`.

[8] D. A.M.-T.L.D.R.S.J.H.F.R.T.Z.Y.A.K.N.Y.L.Q. V. Le. "Towards a Human-like Open-Domain Chatbot". In: *arXiv preprint arXiv:2001.09977* (2020).

[9] "Left and Right Hemisphere of the Brain". In: Oct. 2019. URL: `https://human-memory.net/left-and-right-hemisphere-of-the-brain/`.

[10] N. Majumder et al. "Dialoguernn: An attentive rnn for emotion detection in conversations". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 6818–6825.

[11] V. Mavani, S. Raman, and K. Miyapuram. "Facial Expression Recognition Using Visual Saliency and Deep Learning". In: Oct. 2017, pp. 2783–2788. DOI: `10.1109/ICCVW.2017.327`.

[12] R Plutchnik and H Kellerman. "Emotion". In: *A psychoevolutionary synthesis, New York* (1980).

[13] S. Poria et al. "MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations". In: *ACL*. 2018.

[14] J. A. Russell. "A circumplex model of affect." In: *Journal of personality and social psychology* 39.6 (1980), p. 1161.