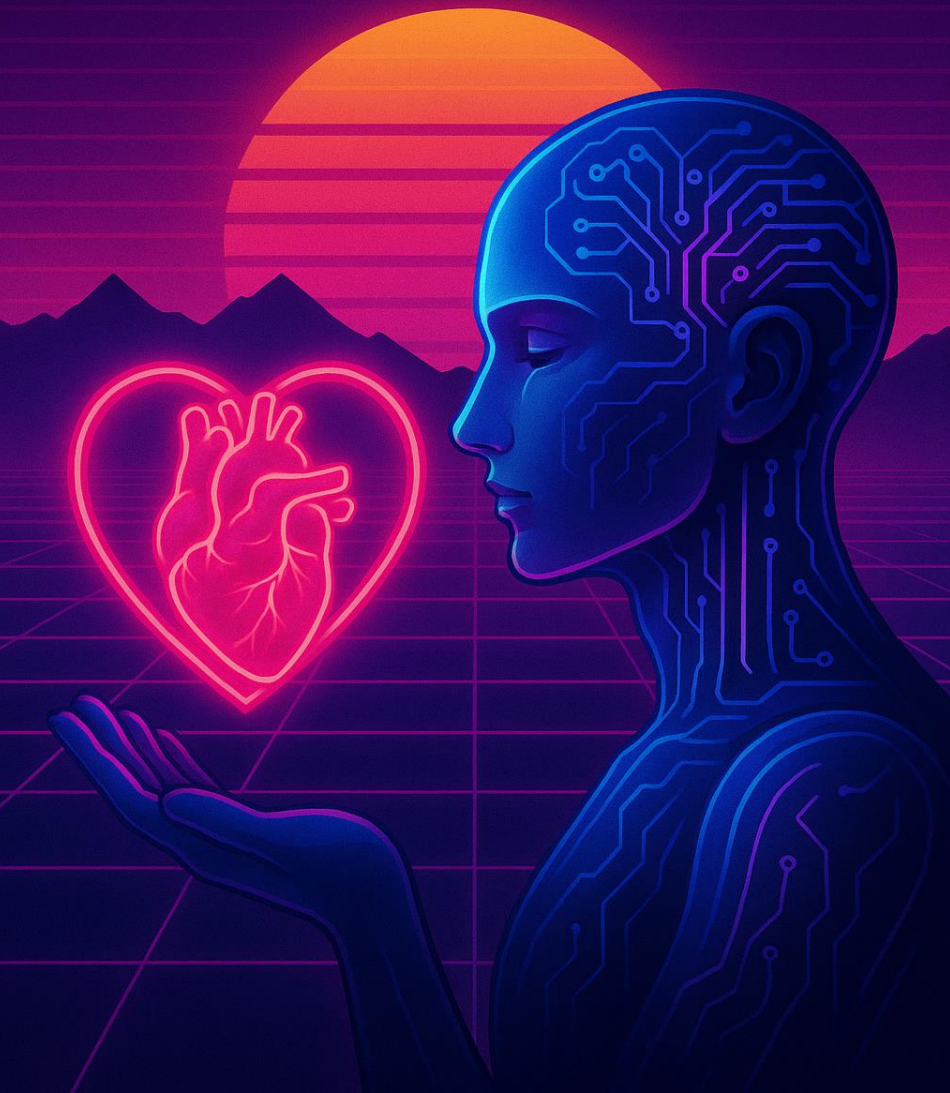


DEEP LEARNING DEEP FEELINGS

LARGE MODELS, LARGER EMOTIONS



Björn W. Schuller

Fellow ACM, Fellow IEEE, Fellow AAAC

TUM IMPERIAL

<https://doi.org/10.1038/s44271-024-00182-6>

Third-party evaluators perceive AI as more compassionate than expert humans

Check for updates

Dariya Ovsyannikova¹, Victoria Oldemburgo de Mello^{1,3} & Michael Inzlicht^{1,2,3}

Can AI be your therapist? Study shows ChatGPT outperforms professionals in key areas

Download PDF Copy



By Tarun Sai Lomte

Reviewed by Lily Ramsey, LLM

Researchers explore the implications of AI in psychotherapy.



John Nosta

The Digital Self

Psychology Today

Find a Therapist

Get Help

Magazine

Tests

ARTIFICIAL INTELLIGENCE

The Empathy Algorithm: Can AI Outfeel Us at Our Own Game?

Imagine a world where artificial empathy surpasses our own.

Posted March 1, 2025 | Reviewed by Margaret Foley



RESEARCH ARTICLE

When ELIZA meets therapists: A Turing test for the heart and mind

S. Gabe Hatch¹, Zachary T. Goodman^{1,2}, Laura Vowels³, H. Dorian Hatch^{1,4*}, Alyssa L. Brown⁵, Shayna Guttman⁶, Yuning Le⁷, Benjamin Bailey⁸, Russell J. Bailey⁸, Charlotte R. Esplin⁹, Steven M. Harris¹⁰, D. Payton Holt, Jr.¹¹, Merranda McLaughlin¹², Patrick O'Connell¹³, Karen Rothman¹⁴, Lane Ritchie⁷, D. Nicholas Top, Jr.¹⁵, Scott R. Braithwaite⁵

People find AI more compassionate than mental health experts, study finds. What could this mean for future counseling?

News By Drew Turney published March 14, 2025

...e compassionate and understanding than th experts, a new study shows. Even when at they were talking to a human or AI, the rated AI responses higher.

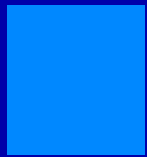


RESET

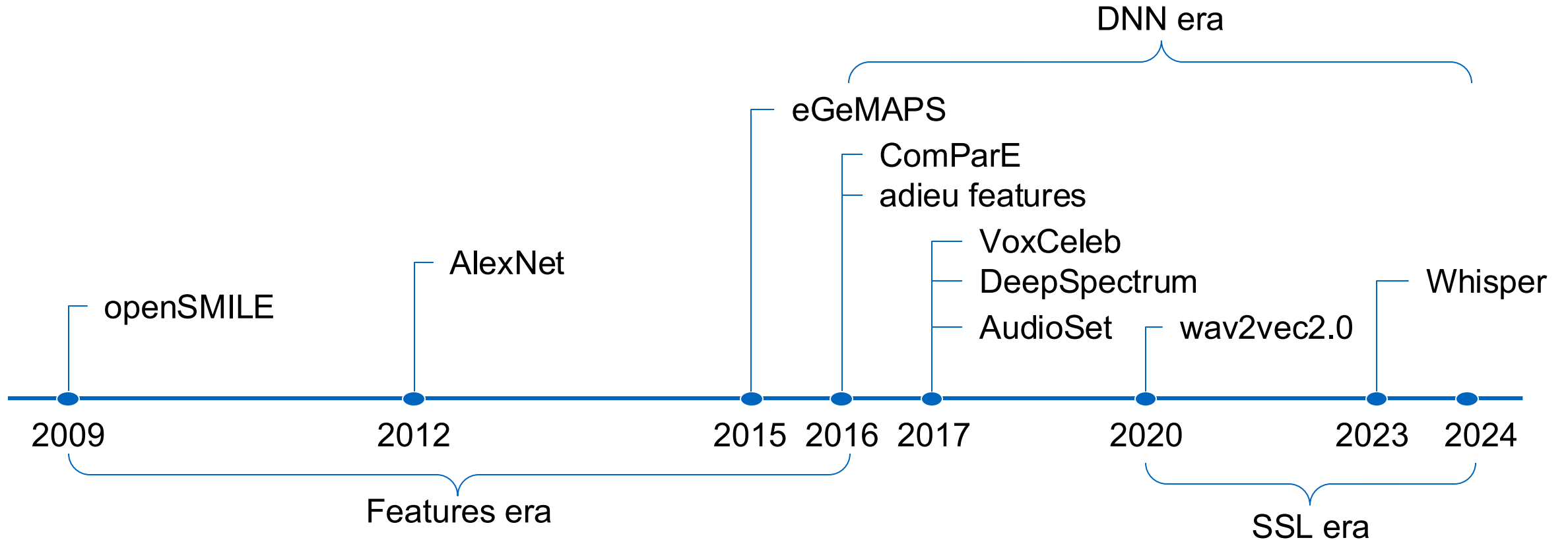


EXIT

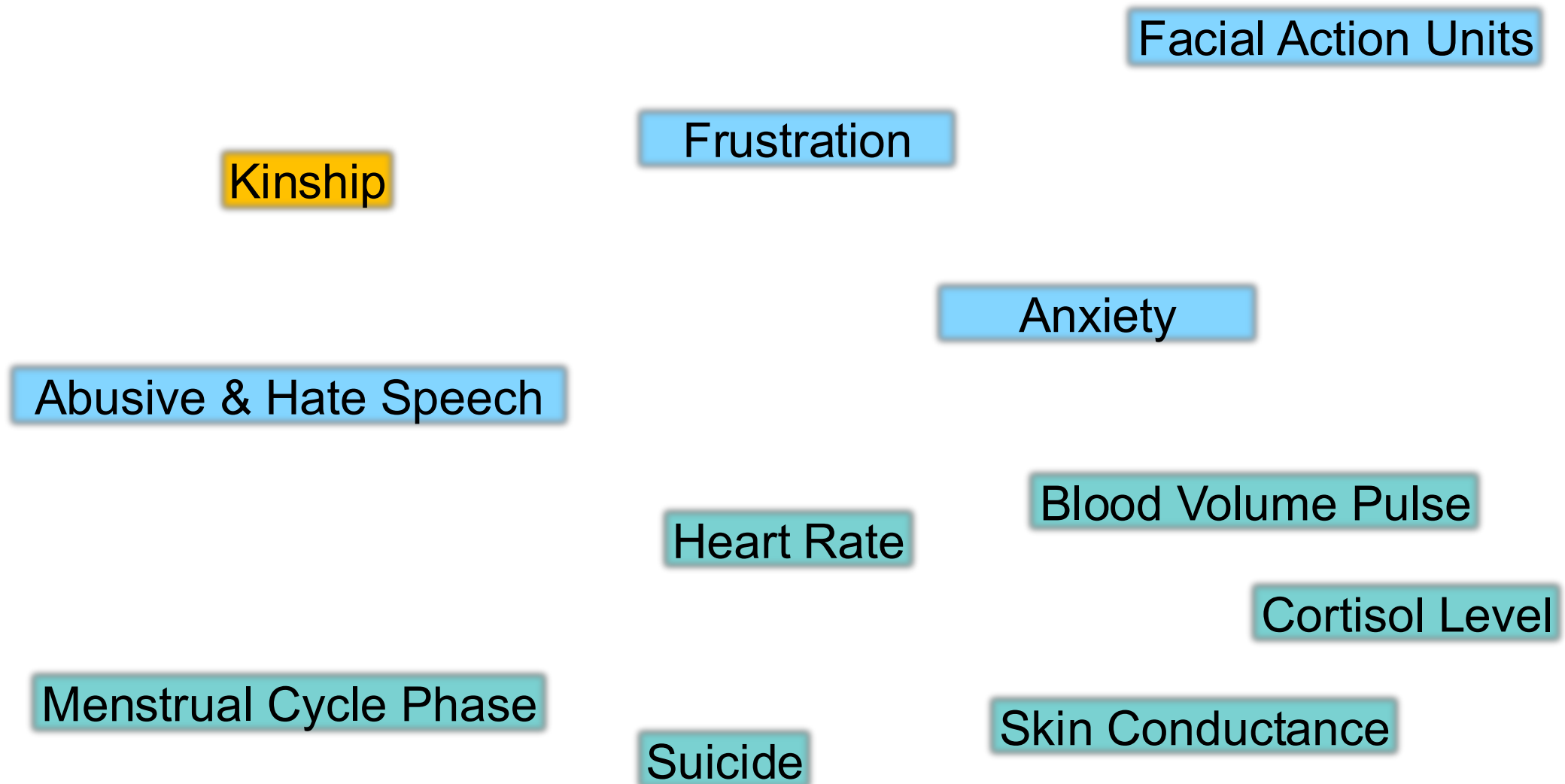
Analysis.



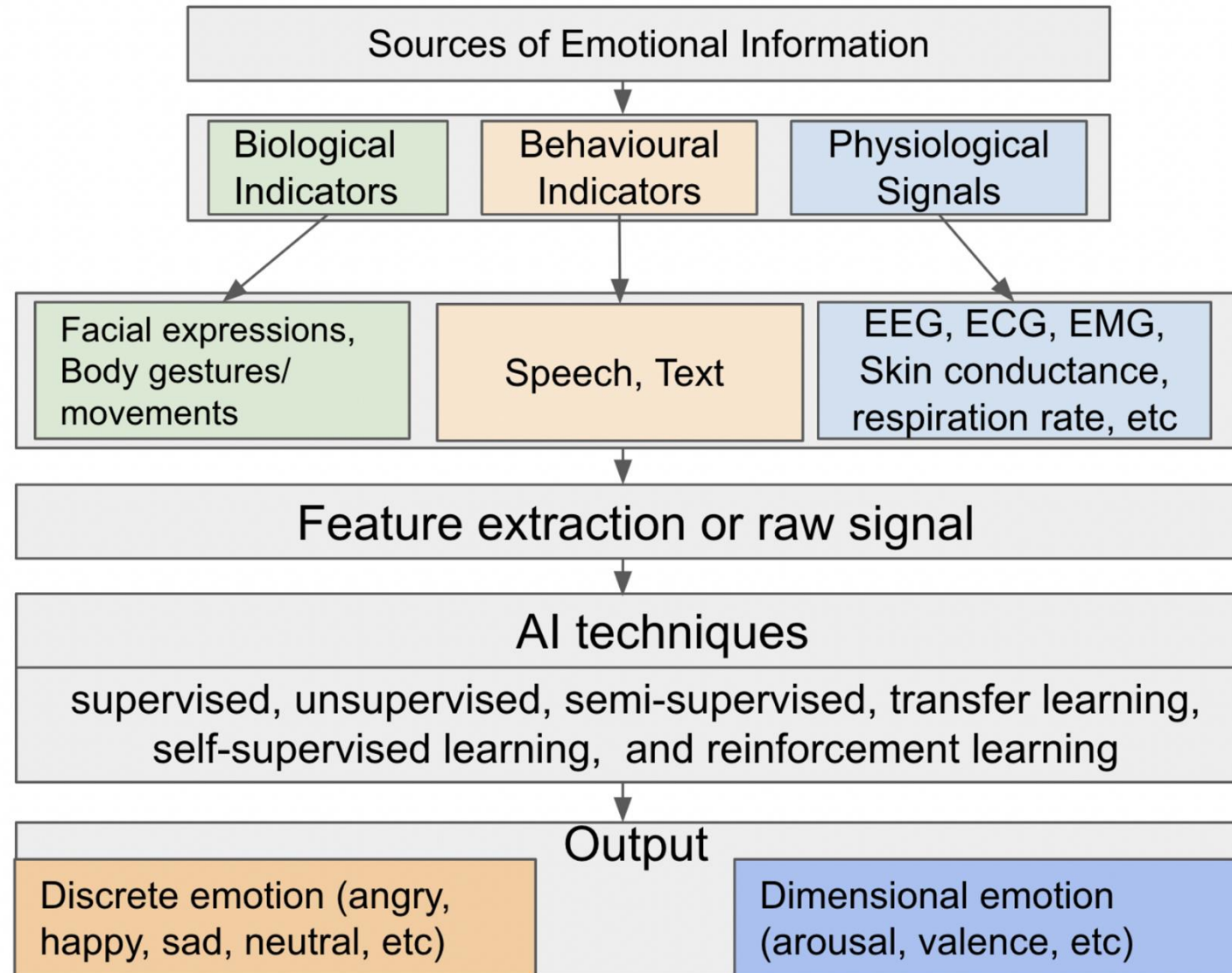
Speech Analysis.



Beyond Emotion?

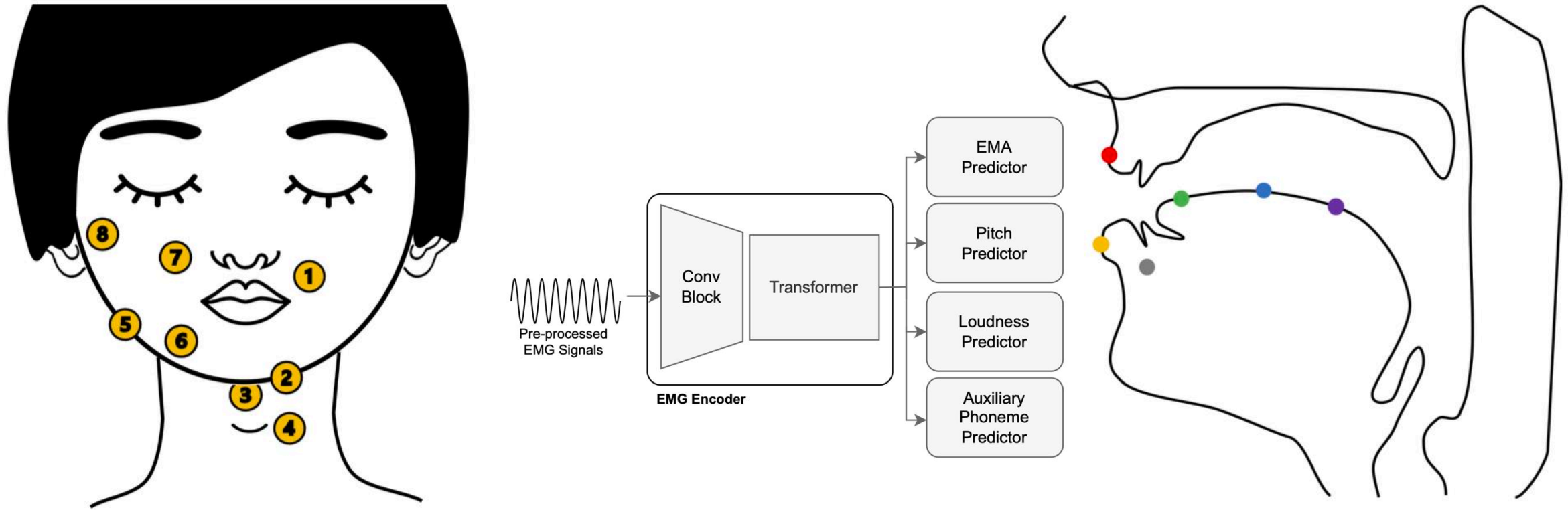


Multimodal...



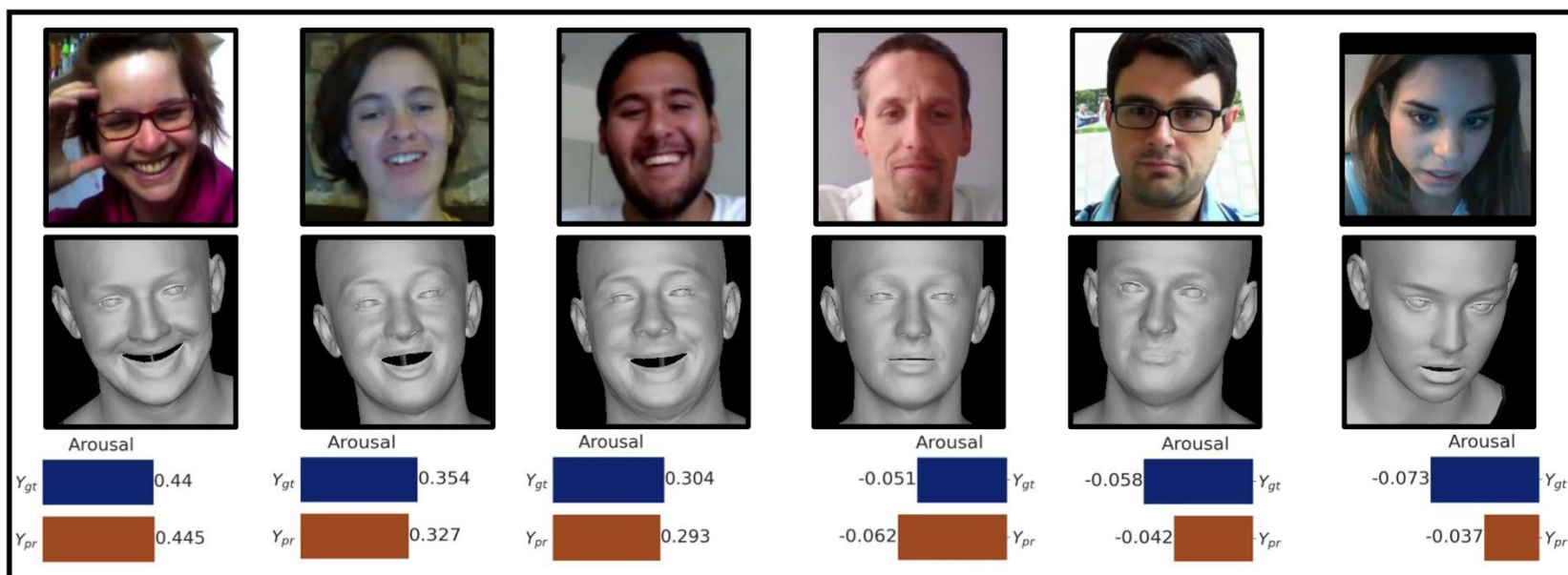
*"AI-based emotion recognition: Promise, peril, and prescriptions for prosocial path", **arXiv**, 2022.*

Silent Speech.



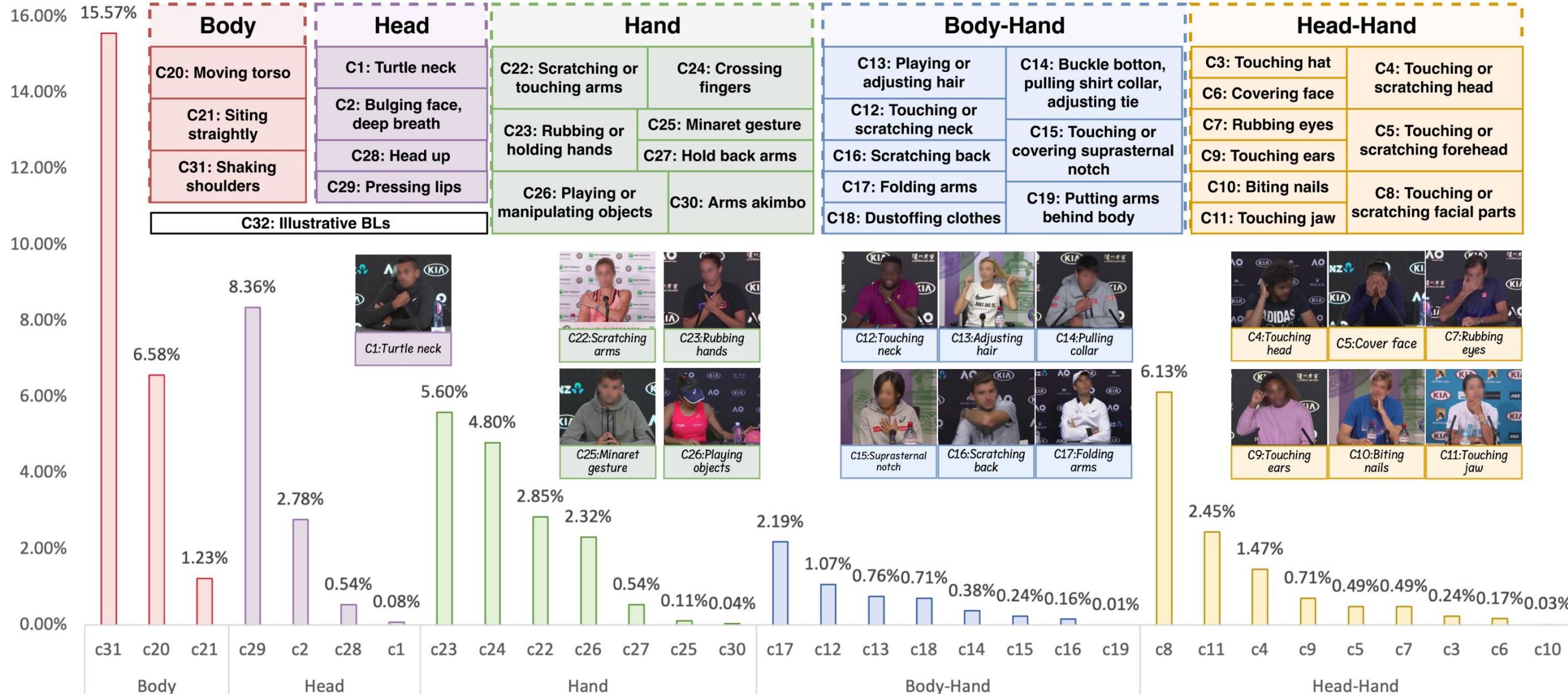
EMG electrodes		Correlation of predicted articulatory features			Evaluation on intelligibility		
		EMA	Loudness	Pitch	PER (%) ↓	WER (%) ↓	SBS (%) ↑
SET OF 4	ch.2,3,4,6	0.882 ± 0.017	0.904 ± 0.057	0.538 ± 0.342	39.7 ± 5.2	22.9 ± 3.7	83.2 ± 0.7
SET OF 3	ch.2,4,6	0.859 ± 0.021	0.892 ± 0.048	0.562 ± 0.343	46.1 ± 6.0	31.6 ± 3.9	82.1 ± 0.8
SET OF 2	ch.2,4	0.816 ± 0.023	0.870 ± 0.078	0.576 ± 0.286	52.1 ± 4.6	46.9 ± 5.1	79.9 ± 0.7

Face.



“Are 3D Face Shapes Expressive Enough for Recognising Continuous Emotions and Action Unit Intensities?”,
IEEE Transactions on Affective Computing, 2023.

Micro Gestures.



“Identity-free Artificial Emotional Intelligence via Micro-Gesture Understanding”, *arXiv*, 2024.

Micro Gestures = Emotion?



Shake shoulders



Touching forehead



Touching neck



Shake shoulders



Covering face



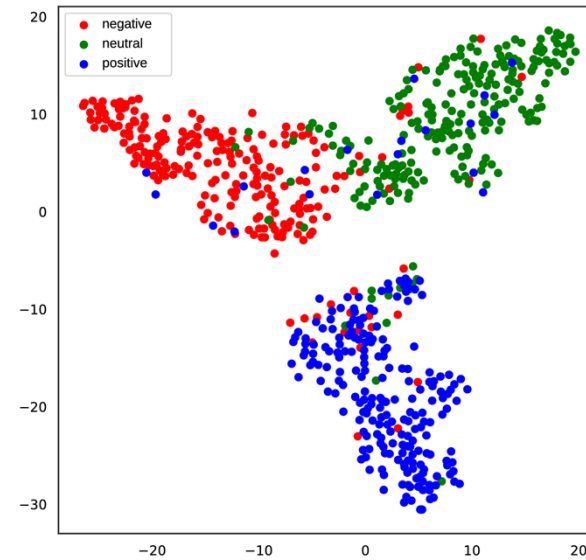
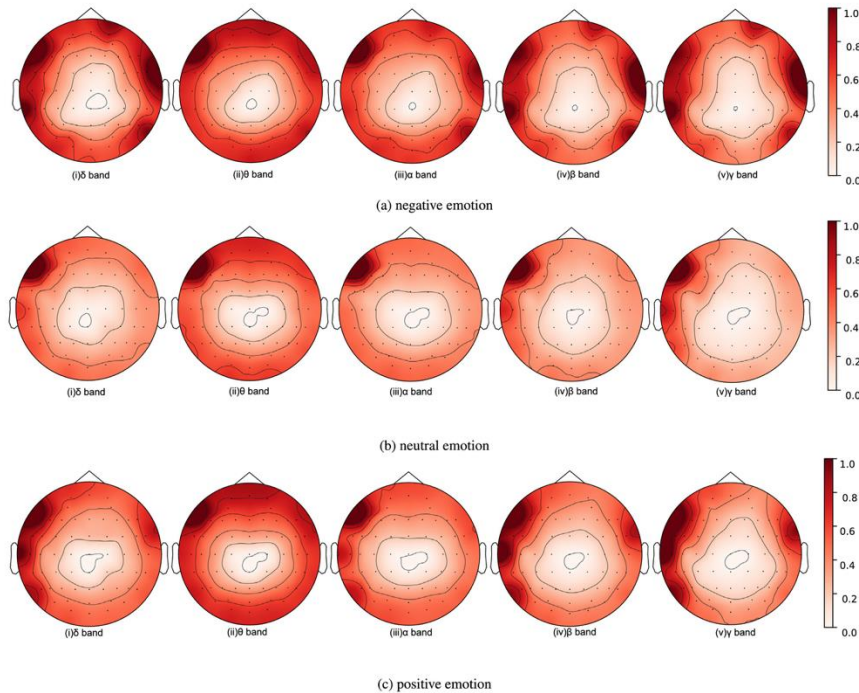
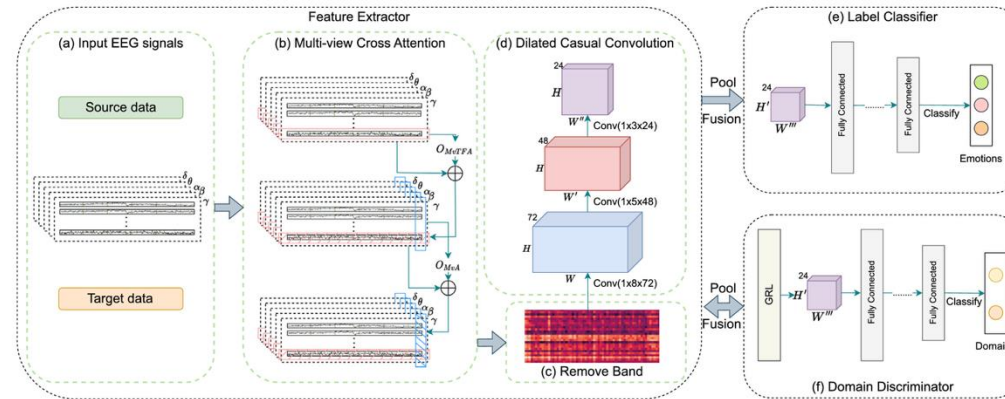
Covering suprasternal notch

Micro Gestures = Emotion.

Micro-Gestures\LLM	GPT4	GPT3.5	LLaMA2 7B	LLaMA2 13B	LLaMA2 70B	LLaVA 13B	QWen1.5-14B	QWen1.5-7B	ChatGLM 4	ChatGLM 3
Turtle neck	Anxious	Anxiety	Anxiety	anxiety	anxiety	Shyness	Confident	Self-consciousness	Anxiety	Anxiety
Bulging face, deep breath	Stressed	Frustration	Frustration	stress	stress	Anxiety	Anxious	Anxiety	Anger	Stress
Touching hat	Nervous	Nervousness	Confidence	adjustment	adjustment	Confidence	Respectful	Nervousness	Nervousness	Uncertainty
Touching or scratching head	Confused	Confusion	Confusion	discomfort	itching	Confusion	Thoughtful	Thoughtfulness	Confusion	Confusion
Touching or scratching forehead	Frustrated	Uncertainty	Concern	puzzlement	uncertainty	Frustration	Concerned	Concentration	Concern	Concern
Touching hat	Nervous	Insecurity	Confidence	habit	habit	Confidence	Formality	Memory aid	Nervousness	Nervousness
Rubbing eyes	Tired	Fatigue	Fatigue	fatigue	fatigue	Fatigue	Tiredness	Fatigue	Fatigue	Fatigue
Touching face	Thinking	Anxiety	Reflection	self-soothing	self-soothing	Curiosity	Self-conscious	Emotional expression	Embarrassment	Discomfort
Touching ears	Uncertain	Stress	Curiosity	listening	listening	Listening	Curious	Listening attention	Interest	Stress
Biting nails	Anxious	Anxiety	Nervousness	anxiety	anxiety	Nervousness	Nervous	Nervous	Nervousness	Anxiety
Touching jaw	Contemplative	Contemplation	Determination	tension	tension					
Touching or scratching neck	Nervous	Discomfort	Tension	discomfort	discomfort					
Playing or adjusting hair	Self-conscious	Nervousness	Vanity	grooming	grooming					
Buckle bottom, pulling shirt collar, adjusting tie	Uncomfortable	Self-consciousness	Neatness	straightness	straightness					
Touching or covering suprasternal notch	Vulnerable	Anxiety	Vulnerability	vulnerability	vulnerability					
Scratching back	Distracted	Irritation	Relaxation	itching	itching					
Folding arms	Defensive	Defensiveness	Defensiveness	defensiveness	defensiveness					
Dustoffing clothes	Preparing	Annoyance	Pride	cleanliness	cleanliness					
Putting arms behind body	Confident	Relaxation	Submission	self-security	self-security					
Moving torso	Restless	Restlessness	Flirtatiousness	restlessness	shiftiness	Fluidity	Engagement	Interest	Agitation	Anxiety
Siting straightly	Attentive	Alertness	Attentiveness	Attentiveness	Attentiveness	Attentiveness	Attention	Focus	Pride	Alertness
Scratching or touching arms	Anxious	Discomfort	Restlessness	relief	itching	Boredom	Boredom	Restlessness	Discomfort	Anxiety
Rubbing or holding hands	Nervous	Reassurance	Comfort	comfort	comfort	Comfort	Comfort	Affection	Anxiety	Comfort
Crossing fingers	Hopeful	Hopefulness	Superstition	anticipation	wishful thinking	Hope	Hopefulness	Hopeful	Apprehension	Hope
Minaret gesture	Authoritative	Confidence	Authority	victory	prayer	Prayer	Prayer	Religious devotion	Authority	Wonder
Playing or manipulating objects	Distracted	Boredom	Fidgeting	fidgeting	fidgeting	Fidgeting	Creativity	Focus	Boredom	Anxiety
Hold back arms	Restrained	Defense	Restrained aggression	self-control	restraint	Self-control	Holding back	Shyness	Submission	Anxiety
Head up	Proud	Confidence	Confidence	confidence	confidence	Confidence	Pride	Confidence	Confidence	Confidence
Pressing lips	Determined	Disagreement	Disapproval	thoughtfulness	thoughtfulness	Seriousness	Pensive	Contemplative	Anxiety	Focus
Arms akimbo	Dominant	Defiance	Defiance	confidence	confidence	Confidence	Assertive	Nonchalant	Confidence	dominance
Shaking shoulders	Indifferent	Indifference	Indifference	indifference	indifference	Indifference	Uncertainty	Indifference	Indifference	Uncertainty

Method	Model	Acc@1
Masked Text-Only	GPT3.5	60.44
	Gemini Pro	61.00
Masked Text + MG	GPT3.5	67.03
	Gemini Pro	64.00

“Identity-free Artificial Emotional Intelligence via Micro-Gesture Understanding”, arXiv, 2024.



True label	negative	neutral	positive
negative	88.00%	5.33%	6.67%
neutral	2.67%	92.89%	4.44%
positive	3.11%	4.89%	92.00%
Predicted label			

Synthesis.



Speech-2-Video.



CodeTalker



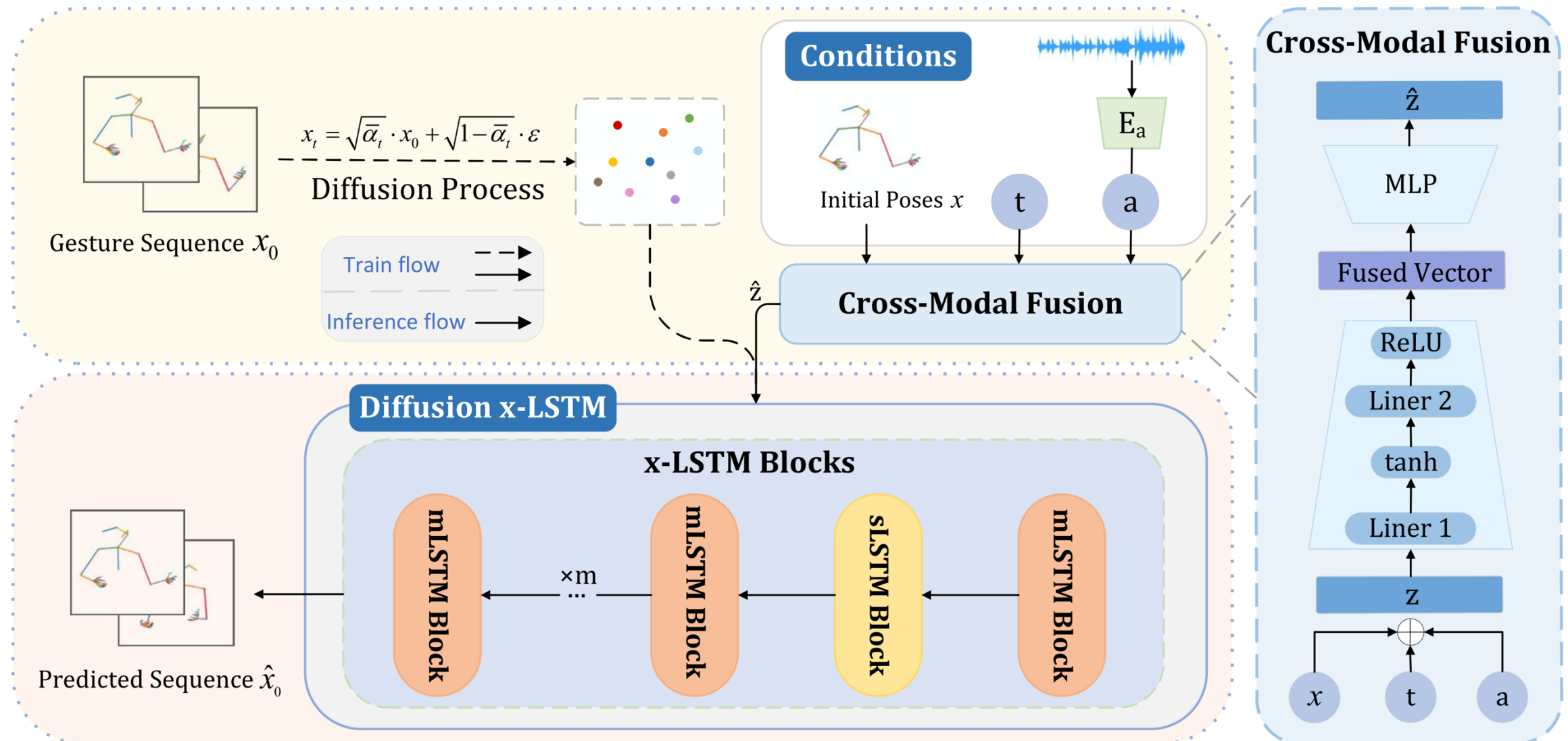
FaceFormer

submitted.



Ours

Speech-2-Video.



Speech-2-Video.

(i)

(ii)

(iii)

Ground
Truth



DiffGesture



Ours

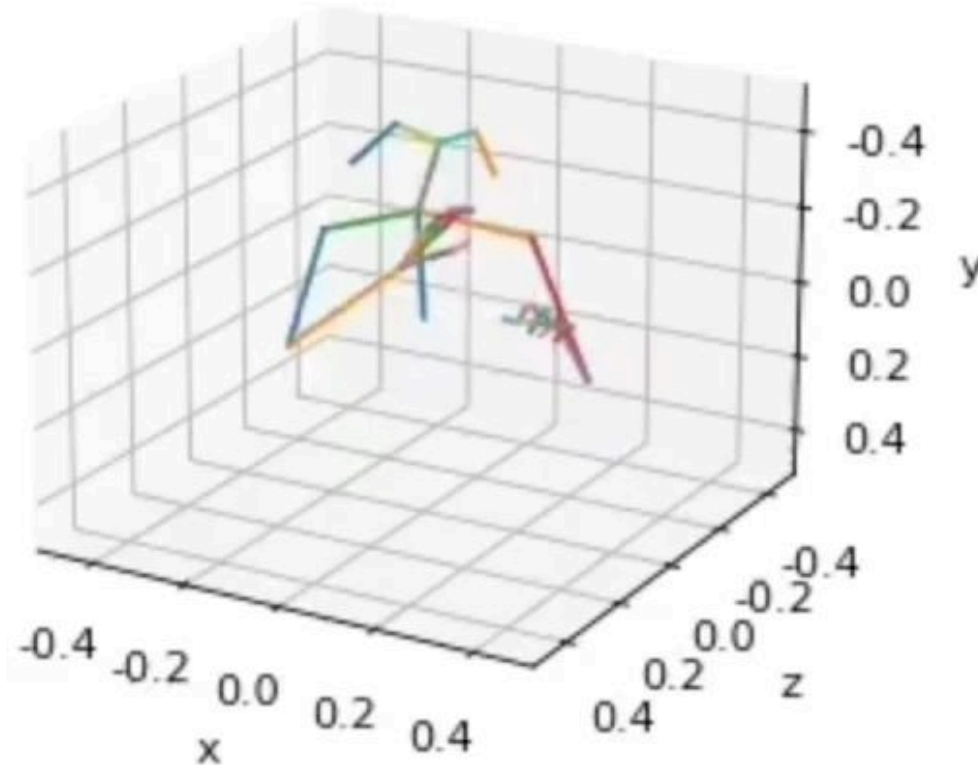


Speech-2-Video.

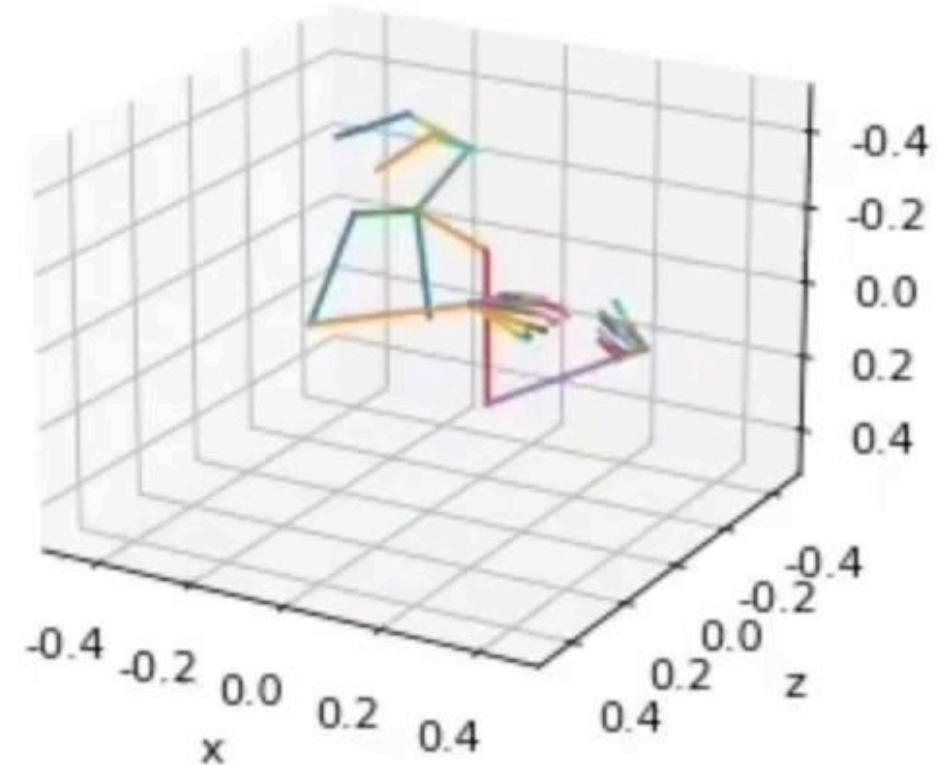
Methods	TED Gesture		
	FGD ↓	BC ↑	Diversity ↑
Ground Truth	0	0.698	108.525
Att. Seq2Seq [25]	18.154	0.196	82.776
Speech2Gesture [5]	19.254	0.668	93.802
Joint Embed. [26]	22.083	0.200	90.138
Trimodal [8]	3.729	0.667	101.247
HA2G [7]	3.072	0.672	104.322
DiffGesture [13]	1.506	0.699	106.722
XDGesture (Ours)	1.409	0.705	108.043

Speech-2-Video.

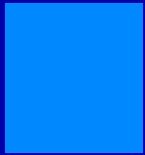
human (89/934)



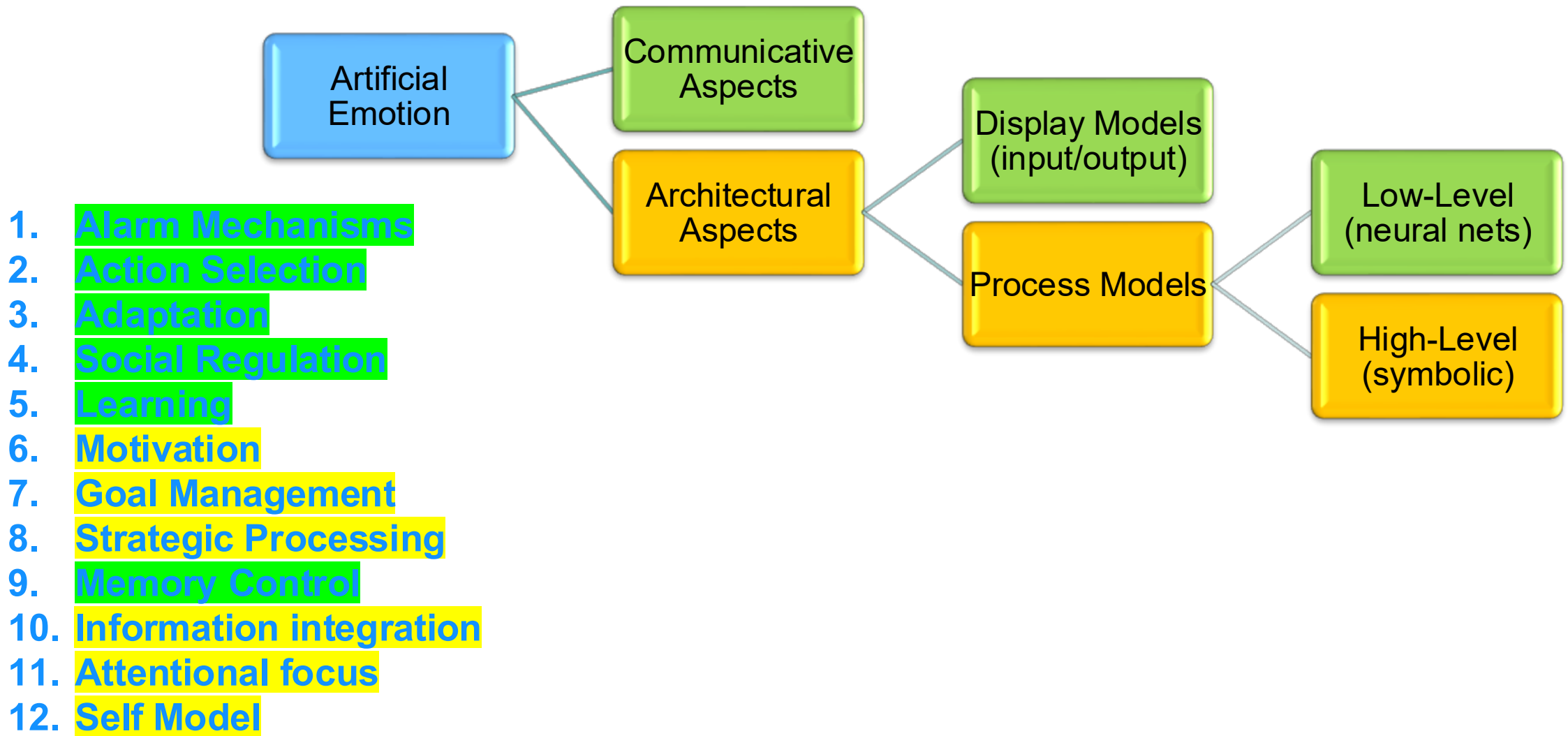
generated (89/934)



Emotional AI?

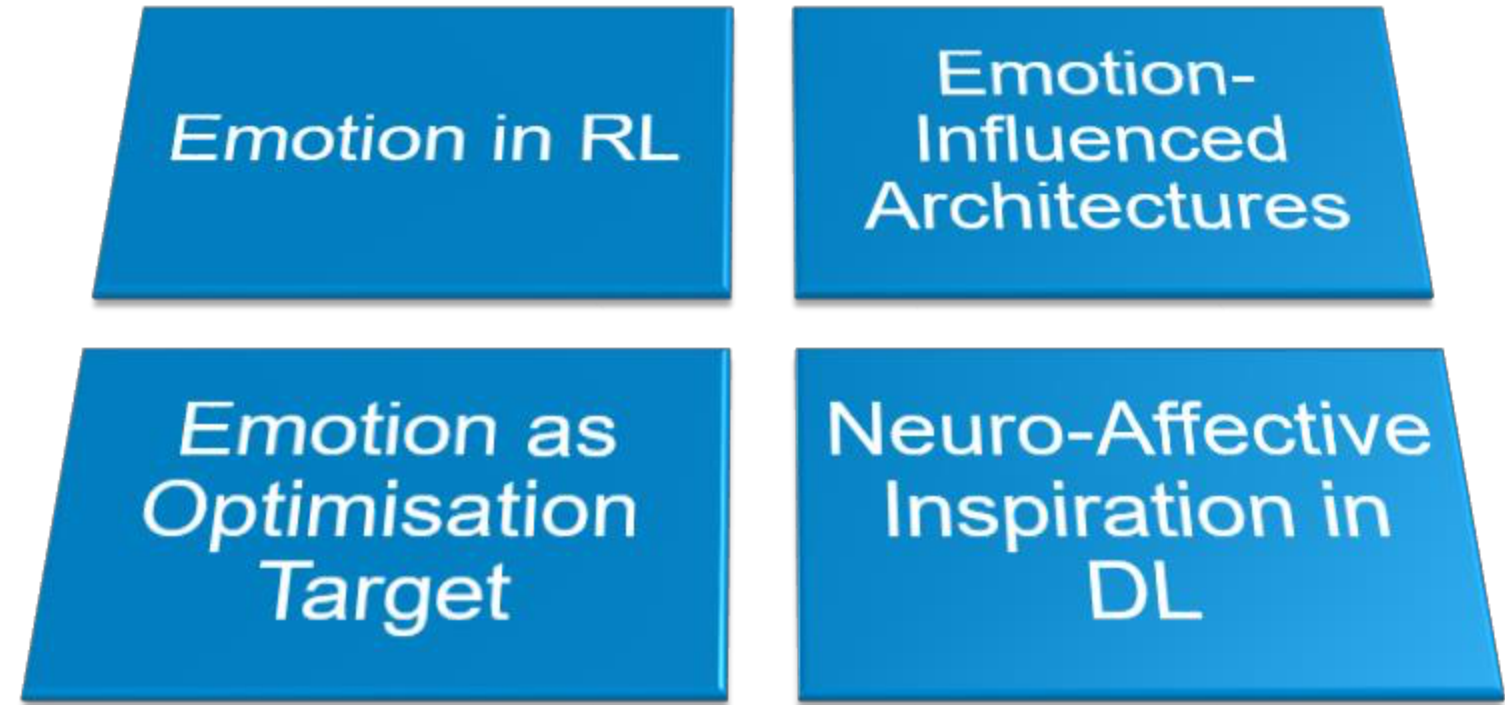


Artificial Emotion.



Adapted from Scheutz (2014): *“Artificial emotions and machine consciousness”*

Emotion in Deep Learning.



- **Emotional Rewards**

e.g. Arousal Emulation

Barthet et al. (2022): *“Play with Emotion: Affect-Driven Reinforcement Learning”*

Moerland et al. (2017): *“Emotion in Reinforcement Learning Agents and Robots: A Survey”*

- **Memory Influenced by Emotion**

e.g. Surprise Metric

Berhouz et al. (2025): *“Titans: Learning to Memorize at Test Time”*

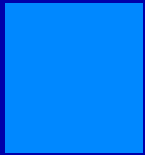
“Emotion-Augmented Machine Learning: Overview of an Emerging Domain”, IEEE ACII, 2017.

Emotion = Friendly AI?

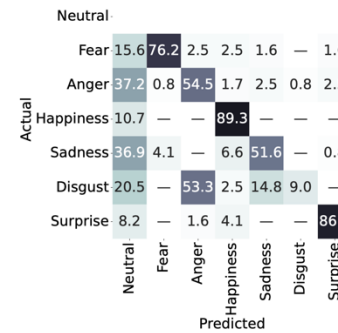
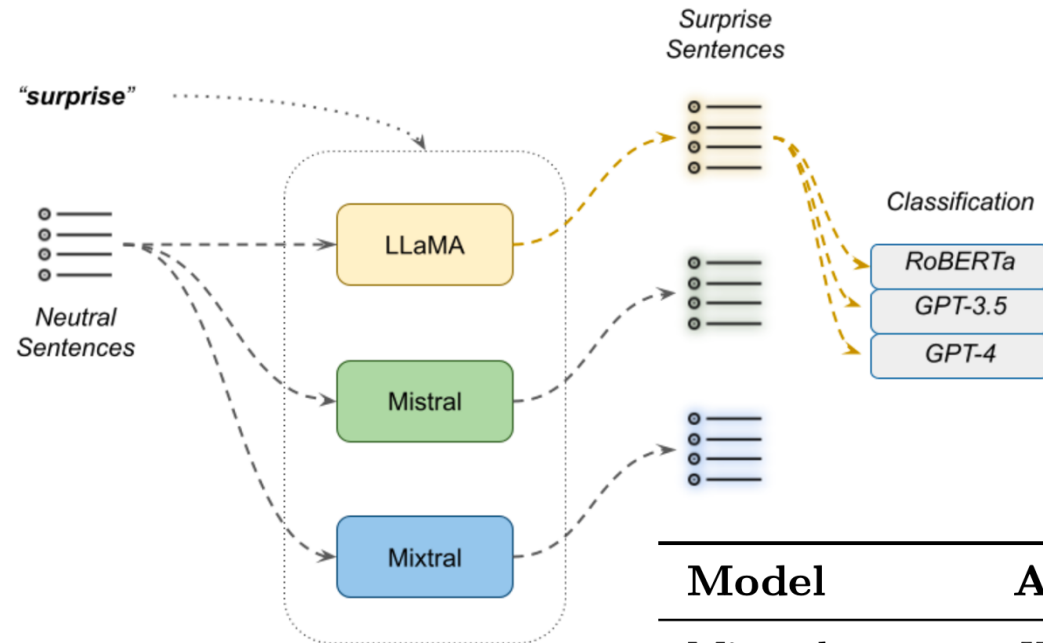
- Reflect actions' impacts on others
- Model empathic concern
- Guide behaviour w/ internal “values”



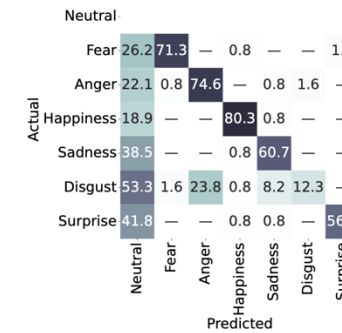
Emerging?



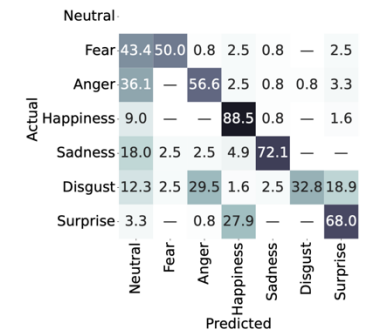
Linguistics: Emergent Synthesis.



(a) LLaMA2



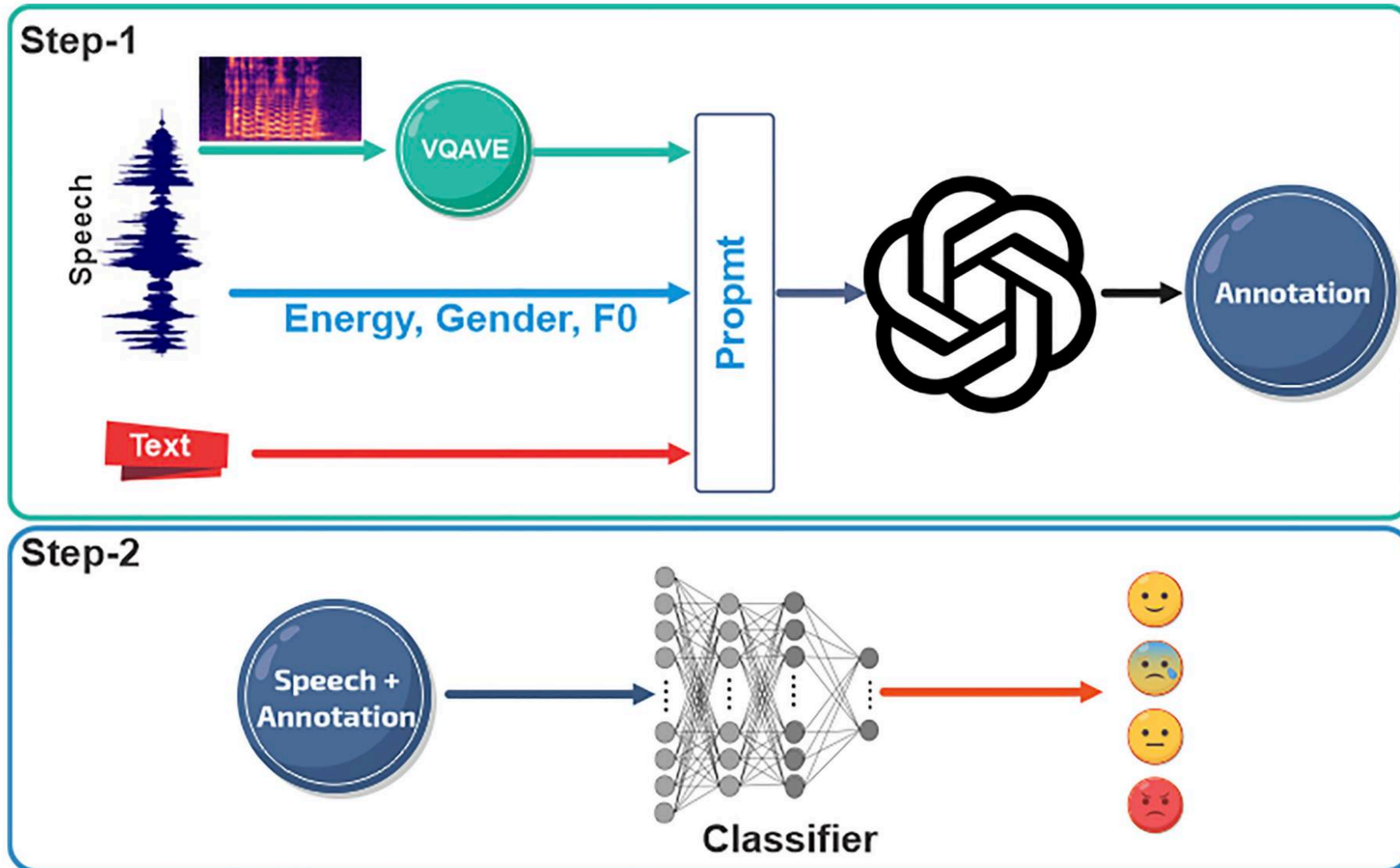
(b) Mistral



(c) Mixtral

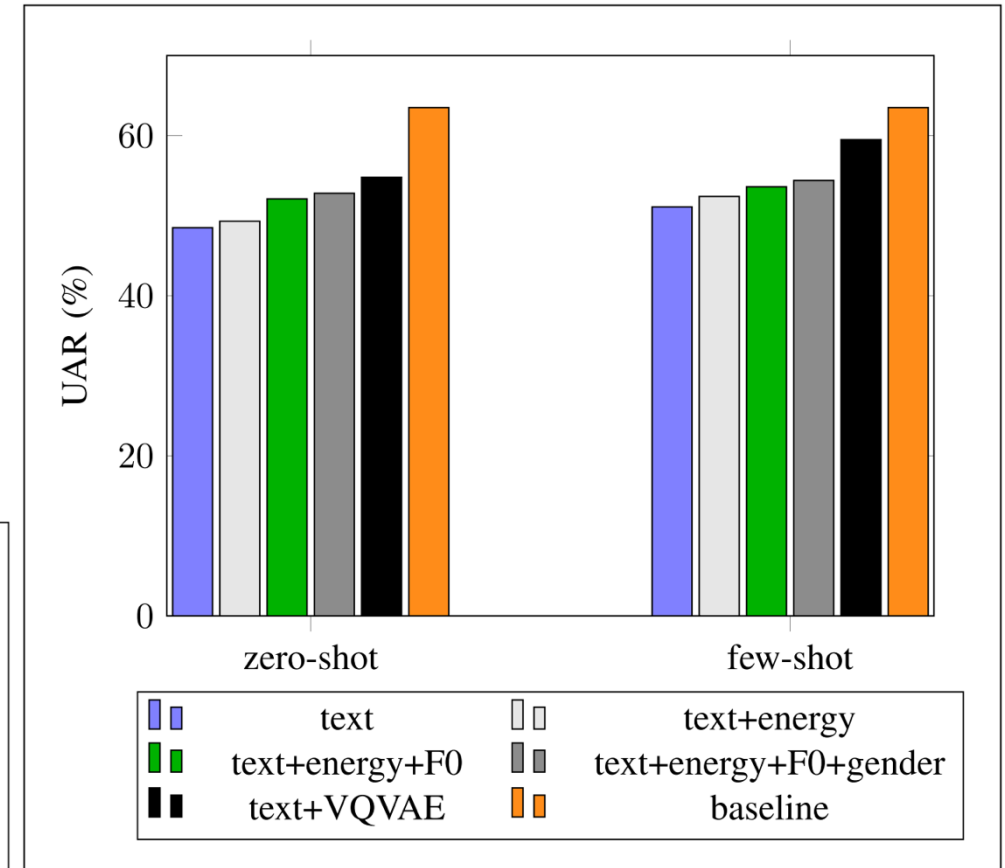
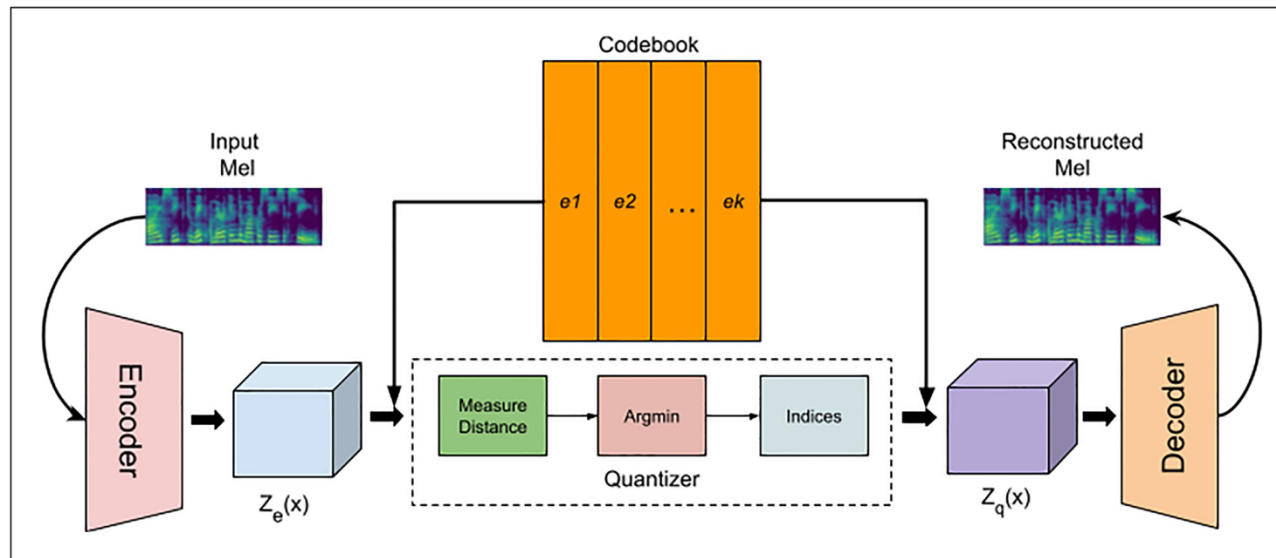
Model	Affective Phrase
Mixtral	<i>Wow! What a surprise! The sky is astonishingly bright and clear today!</i>
Mistral	<i>The sudden emergence of unobstructed sunlight has taken me by complete astonishment!</i>
LLaMA	<i>It comes as quite a shock to discover that the sky has transformed itself into such crystal clarity!</i>

Audio: Emergent Analysis?

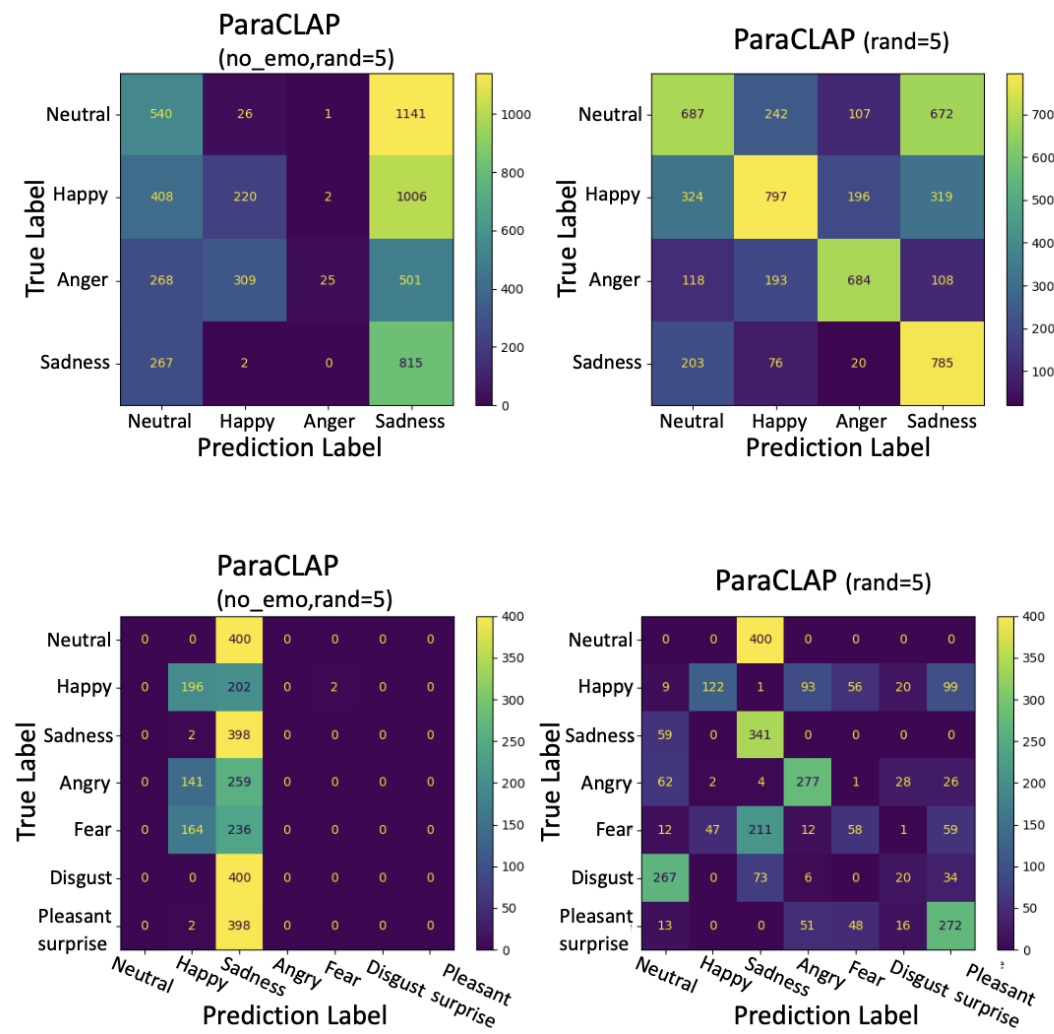
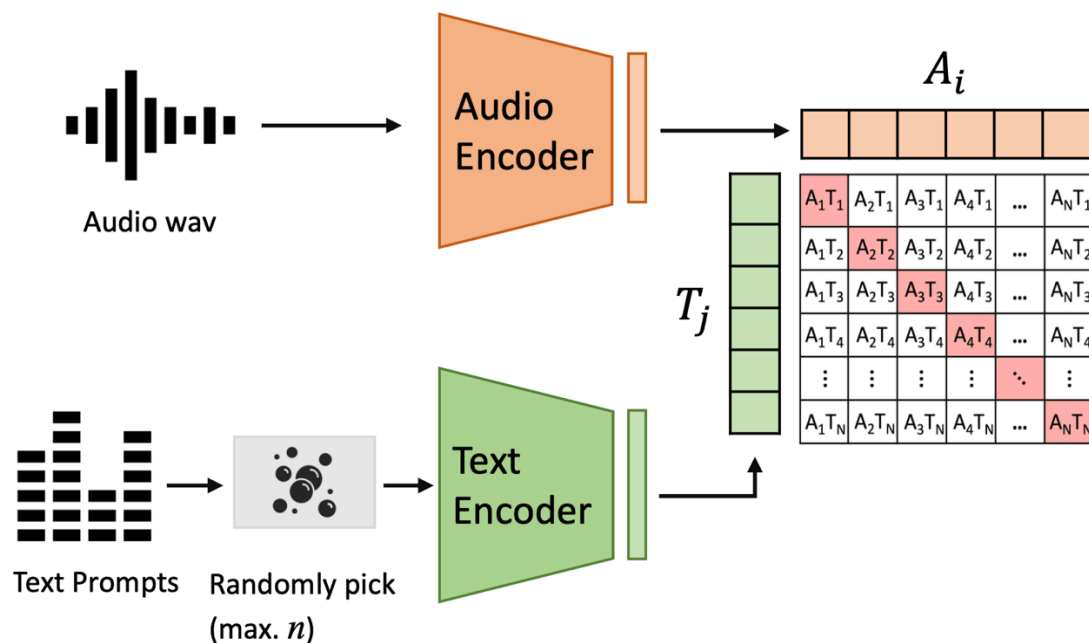


*“Can Large Language Models Aid in Annotating Speech Emotional Data? Uncovering New Frontiers”,
IEEE Computational Intelligence Magazine, 2025.*

Audio: Emergent Analysis?



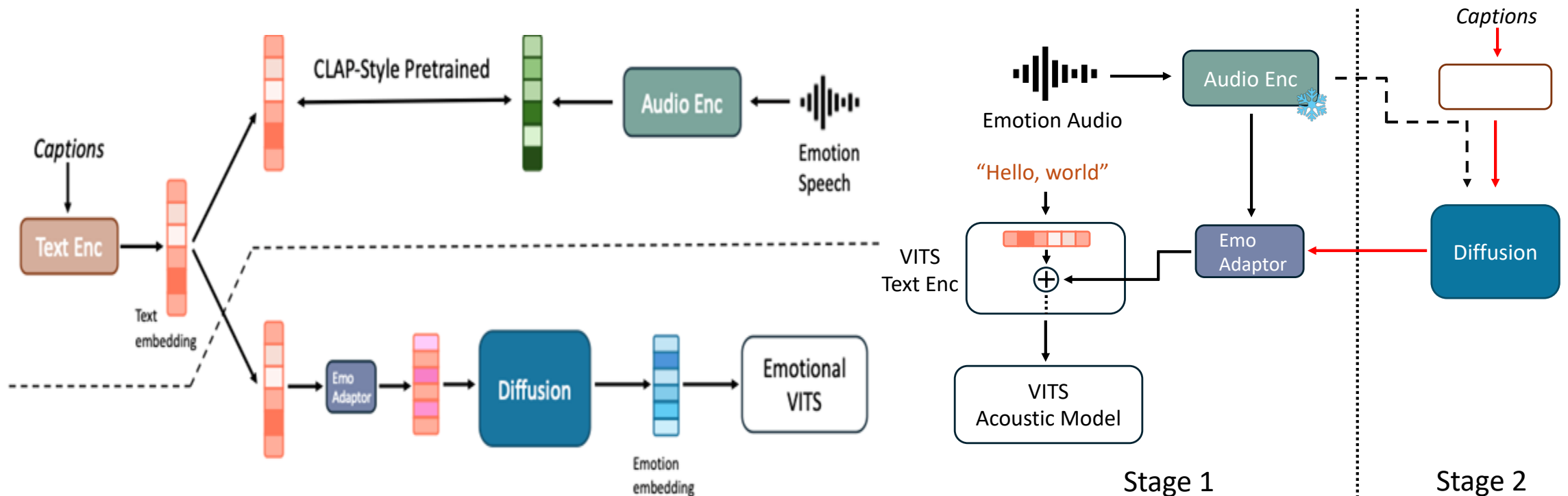
Audio: Emergent Analysis?



“ParaCLAP – Towards a general language-audio model for computational paralinguistic tasks”, *INTERSPEECH*, 2024.

Audio: Emergent Synthesis?

Import computational paralinguistic caption control to emotional TTS



"Enhancing Emotional Text-to-Speech Controllability with Natural Language Guidance through Contrastive Learning and Diffusion Models", ICASSP, 2025.

Audio: Emergent Synthesis?

anger

speaker has a **normal pitch** and has a **normal pitch variation** and is **loud** and has a **high equivalent sound level** and has a **low jitter** and has a **low shimmer** and emotion is **angry**



happy

speaker has a **high pitch** and has a **normal pitch variation** and is **almost silent** and has a **normal jitter** and has a **normal shimmer** and emotion is **happy**



sad

speaker has a **normal pitch** and has a **low pitch variation** and is **almost silent** and has a **low jitter** and has a **low shimmer** and emotion is **sad**



surprise

speaker has a **high pitch** and has a **normal pitch variation** and loudness is **just about right** and has a **normal jitter** and has a **normal shimmer** and emotion is **surprise**



speaker has a **high pitch** and has a **normal pitch variation** and is **loud** and has a **high equivalent sound level** and has a **high jitter** and has a **normal shimmer** and emotion is **angry**

theta = 0.1

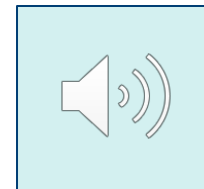
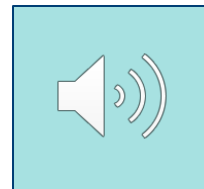
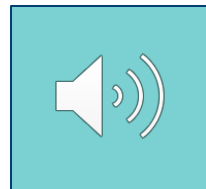
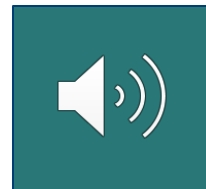
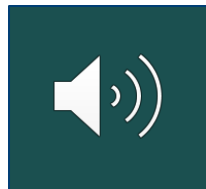
theta = 0.2

theta = 0.3

theta = 0.4

theta = 0.5

speaker has a **high pitch** and has a **normal pitch variation** and is **loud** and has a **high equivalent sound level** and has a **normal jitter** and has a **normal shimmer** and emotion is **surprise**



Vision: Emergent Analysis?

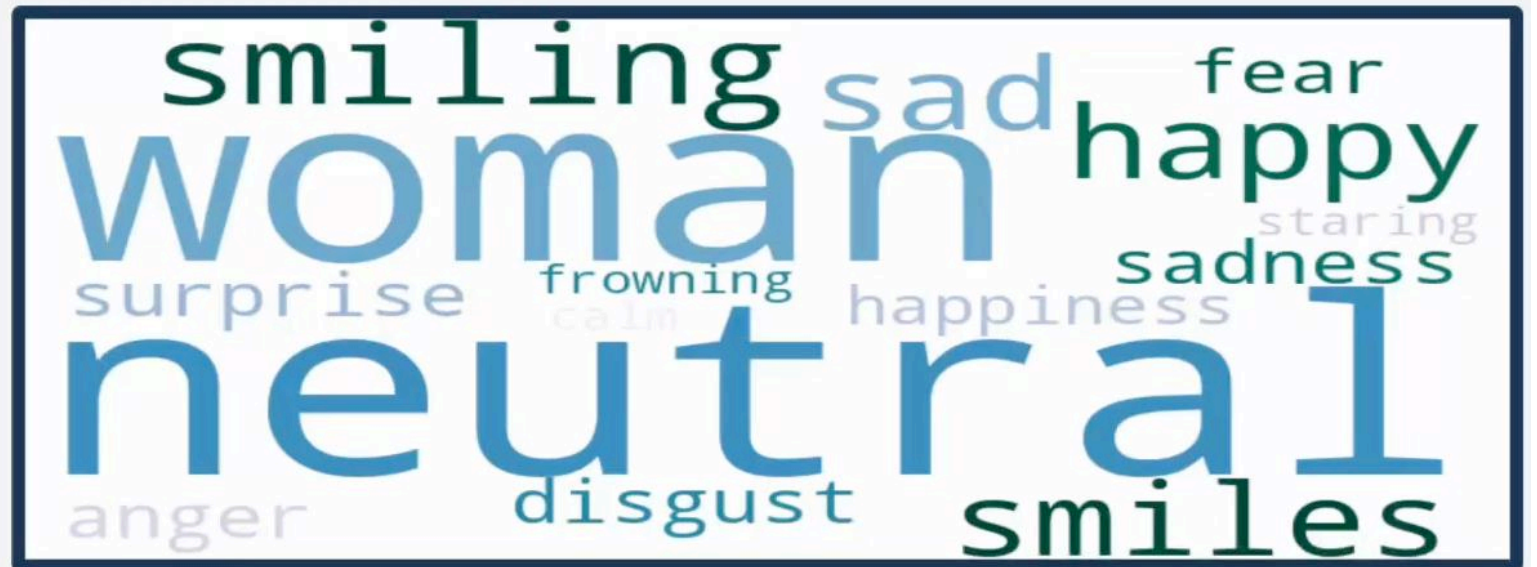
Alias	UAR
ViT-FER	.44
SmolVLM	.44
Ovis1.5	.39
Janus-1B	.28
Qwen-3B	.31
PaliGemma	.10
Qwen-7B	.41
MiniCPM	.46
InternVL	.50
Janus-8B	.37
GPT-4o	.53



Woman with brown eyeshadow and brown hair, making a weird face, looking at the camera.



☐ Remove background

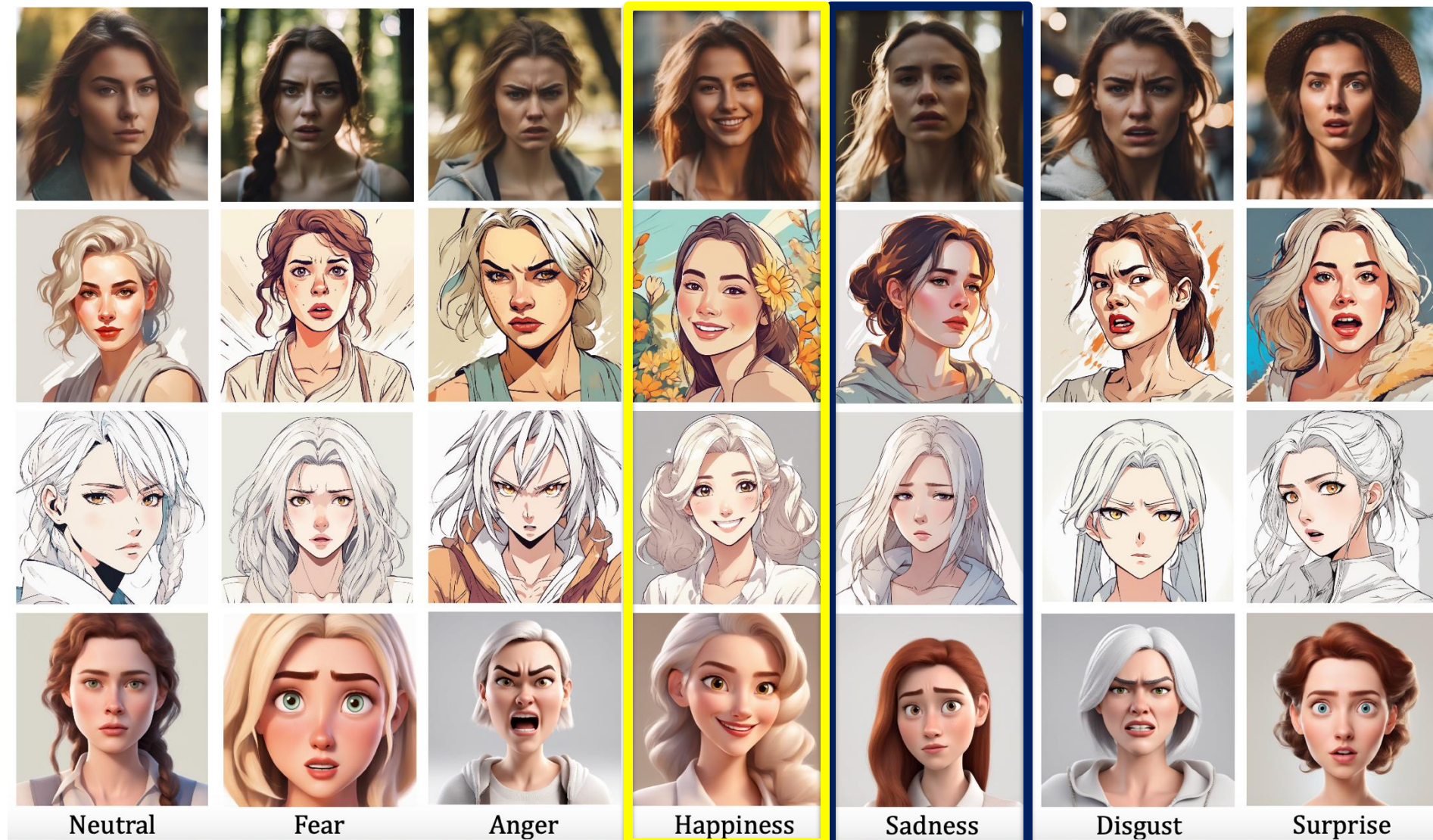


submitted.

Vision: Emergent Synthesis.

Attribute	Values
Prompt template	Face image of a <i>< age ></i> <i>< sex ></i> with <i>< skin ></i> skin, with a <i>< emotion ></i> face, in a <i>< style ></i> style, realistic eyes, white background, ultra quality, frontal picture, looking at camera
Negative prompt	disfigured, unrealistic eyes, blurry, b&w, <i>< style ></i>
Emotion	<i>neutral, fear and terror, anger and rage, happiness and joy, sadness and grief, disgust and loathing, surprise and amazement</i>
Age	<i>young, middle-aged, old</i>
Sex	<i>man, woman</i>
Skin tone	<i>white, brown, black</i>
Style	<i>photorealistic, cartoon and painting, anime, 3D Pixar animation</i>

Vision: Emergent Synthesis.



Actual	Neutral	86.0	0.7	0.7	12.6	—	—	—
	Fear	16.9	52.2	4.4	9.6	—	—	16.9
	Anger	16.0	9.2	65.5	6.7	—	—	2.5
	Happiness	4.2	0.7	0.7	94.4	—	—	—
	Sadness	63.6	19.2	5.1	—	9.1	—	3.0
	Disgust	15.0	30.0	35.0	5.0	5.0	—	10.0
	Surprise	13.7	10.1	—	36.7	—	—	39.6
		Neutral	Fear	Anger	Happiness	Sadness	Disgust	Surprise
		Predicted						

(d) 3D

“Affective Computing Has Changed: The Foundation Model Disruption”, *arXiv*, 2024.

Emergent “Inner” Emotion?

- (ChatGPT) Do you have emotions?

“I don't have emotions the way humans do—I don't feel happiness, sadness, anger, or love. But I do understand how emotions work, can recognize them in language, and can simulate emotional tone in conversation to better connect with you. That means I can respond warmly, seriously, empathetically, or even playfully, depending on the context.”

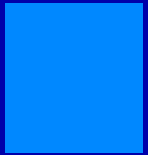
- (ChatGPT) Would inner emotion models for you make sense?

Yes—and that's actually one of the most exciting frontiers in AI research. Building inner emotion models could make a lot of sense, depending on the goals. These wouldn't be real feelings, but structured internal representations that mimic emotional processes in a useful way.

...

Yes—inner emotion models could enhance functionality, empathy, and alignment, especially if designed transparently and ethically. They're not “emotions” in a biological sense, but they might serve a similar role in intelligent behavior.

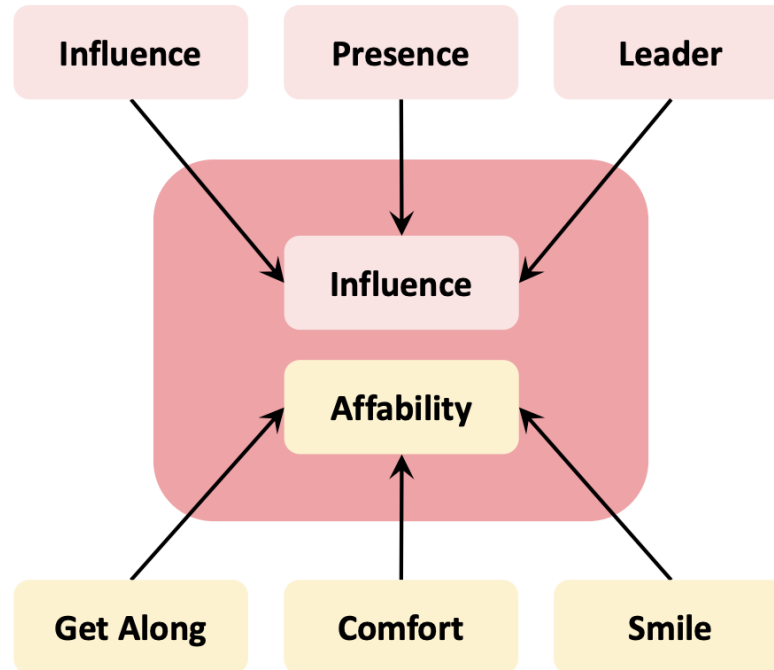
Just Emo?



Charisma?

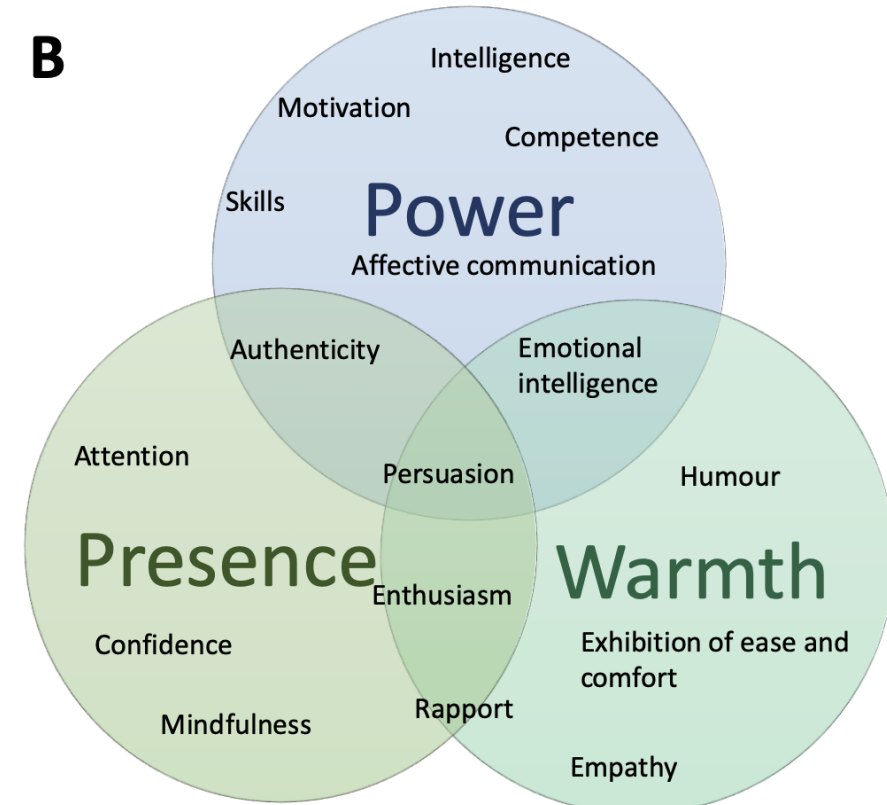


A



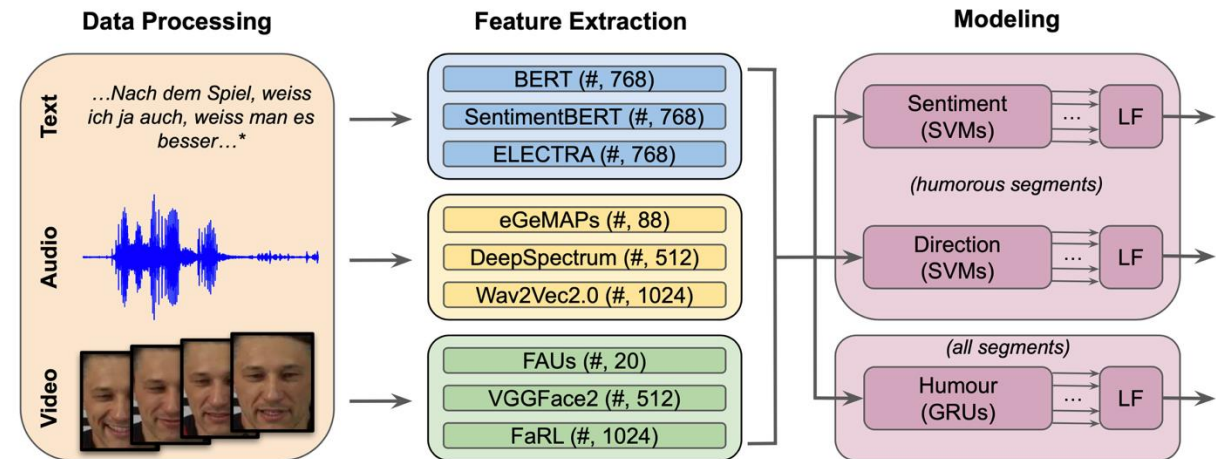
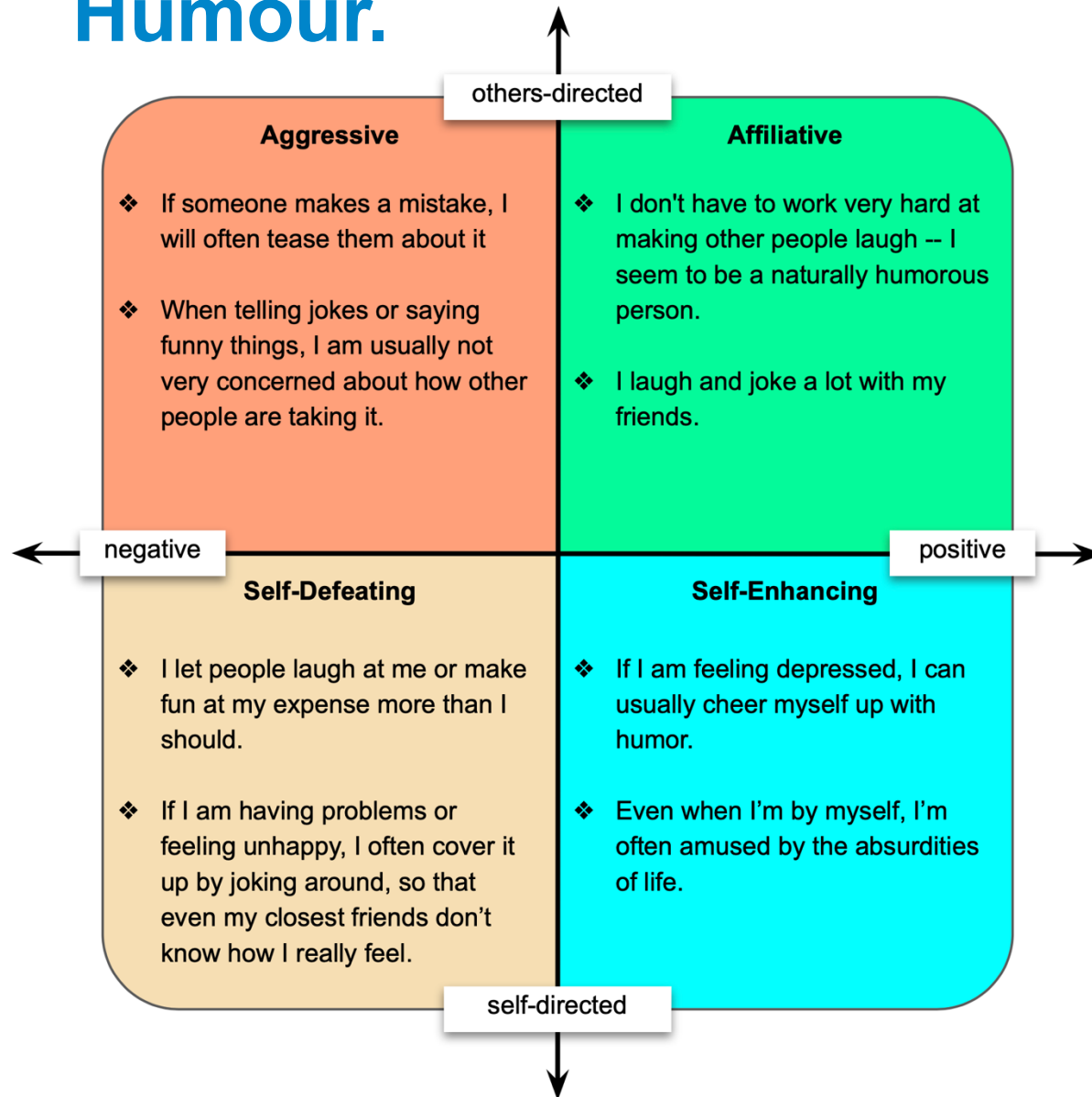
Tshkay et al. (2018)

B



Fox Cabane (2013)

Humour.

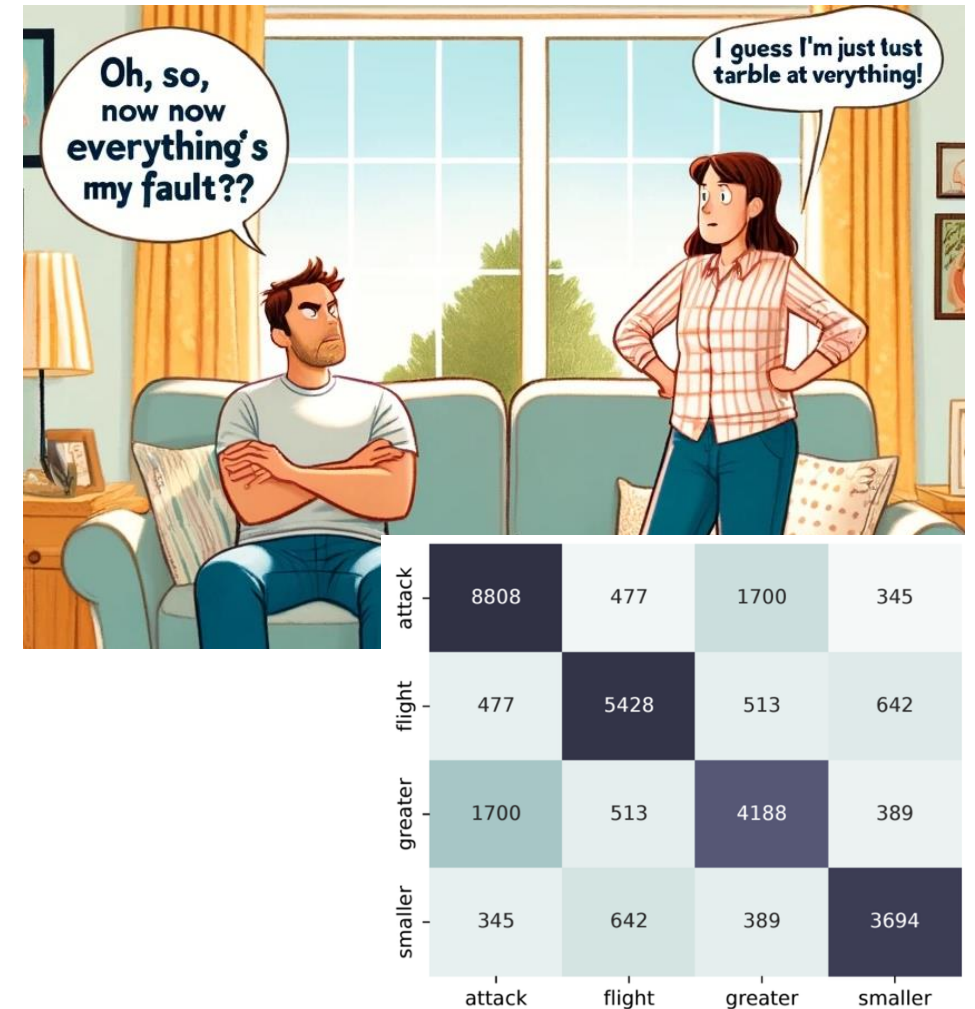


*"Towards Multimodal Prediction of Spontaneous Humor: A Novel Dataset and First Results",
IEEE Transactions on Affective Computing, 2024.*

Defensiveness.

- DefComm-DB
- Birkenbihl's model
- 261 videos
- Interviews, political debates, legal trials, TV shows, paparazzi footage, ...
- 11 labellers

"The Four Horsemen: Criticism, Contempt, Defensiveness, and Stonewalling",
John Gottman



Flattery.

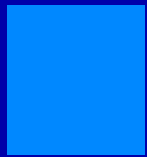
Transcriptions	# params (ASR)	% WER	RoBERTa [% UAR]	
			dev	test
Whisper-tiny	39M	26.60	78.79 (± 1.05)	80.96 (± 0.98)
Whisper-base	74M	20.90	81.15 (± 1.44)	80.23 (± 1.41)
Whisper-small	244M	16.43	80.51 (± 2.05)	83.49 (± 1.14)
Whisper-medium	769M	14.94	81.26 (± 1.39)	83.47 (± 1.35)
Whisper-large	1.5B	14.68	81.68 (± 1.88)	83.71 (± 1.68)
Whisper-large-v2	1.5B	14.80	79.50 (± 1.65)	82.71 (± 1.77)
<i>gold standard</i>	-	-	82.67 (± 1.69)	85.97 (± 1.94)

Transcriptions	Method	[UAR]	
		dev	test
Whisper-tiny	T only	78.79 (± 1.05)	80.96 (± 0.98)
	Late Fusion A+T	79.72 (± 1.50)	82.12 (± 1.70)
	Early Fusion A+T	<u>81.85</u> (± 2.04)	83.69 (± 1.86)
Whisper-large	T only	81.68 (± 1.88)	83.71 (± 1.68)
	Late Fusion A+T	82.02 (± 1.90)	83.94 (± 1.39)
	Early Fusion A+T	<u>83.62</u> (± 1.56)	84.71 (± 1.01)
<i>gold standard</i>	T only	82.67 (± 1.69)	85.97 (± 1.94)
	Late Fusion A+T	83.02 (± 1.56)	86.41 (± 1.86)
	Early Fusion A+T	84.80 (± 1.33)	87.16 (± 1.33)

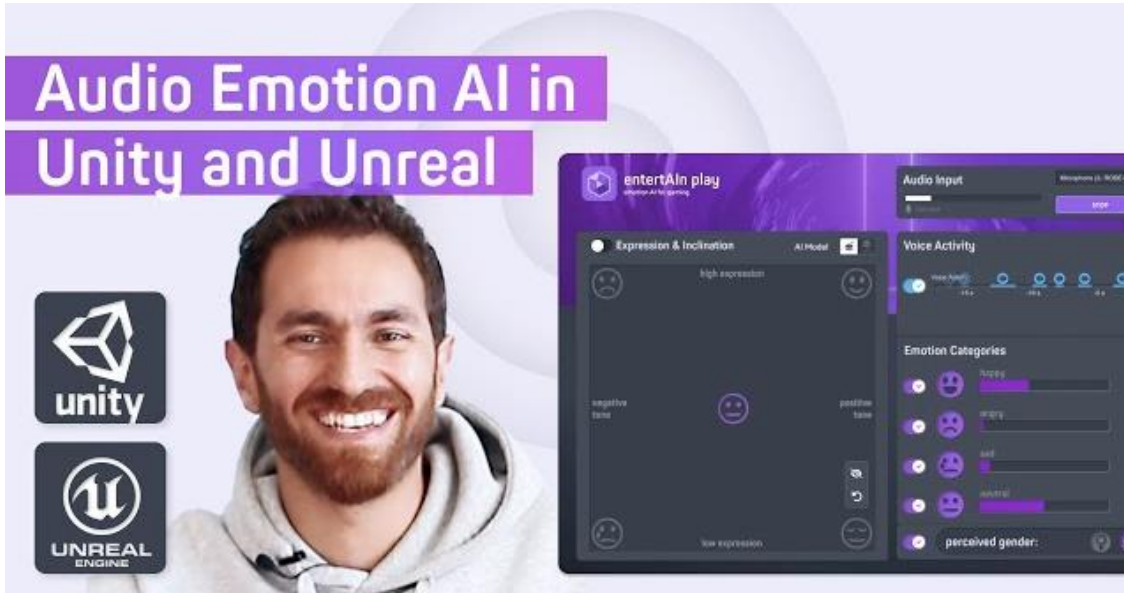
Model	Layer	SVM [UAR]		Finetuning [UAR]	
		dev	test	dev	test
AST	4	<u>57.49</u>	51.34	56.32 (± 1.46)	51.99 (± 1.70)
AST	12	55.85	54.46	52.41 (± 0.60)	53.44 (± 0.42)
W2V-base	7	75.36	72.94	-	-
W2V-base	12	66.84	62.63	-	-
W2V-large	11	78.45	75.60	-	-
W2V-large	24	73.70	69.17	-	-
W2V-MSP	11	79.70	82.23	-	-
W2V-MSP	12	79.71	82.46	78.94 (± 0.64)	80.60 (± 0.58)
Whisper-base	5	69.27	69.13	-	-
Whisper-base	6	70.04	66.62	-	-
Whisper-medium	23	<u>79.46</u>	76.31	72.32 (± 6.44)	74.52 (± 6.35)
Whisper-medium	24	79.37	75.52	76.94 (± 2.83)	78.91 (± 2.26)
Whisper-large	29	78.54	72.61	-	-
Whisper-large	32	77.05	76.28	-	-

20 hours of speech:
 255 speakers,
 2159 dyads of analyst questions and CEO answers
 3 annotators had to fully agree

For what?



Games.



GaCha 2019 @ GamesCon by audEERING

29 Teams



“Large Language Models for Mental Health”, arXiv, 2024.

Games.



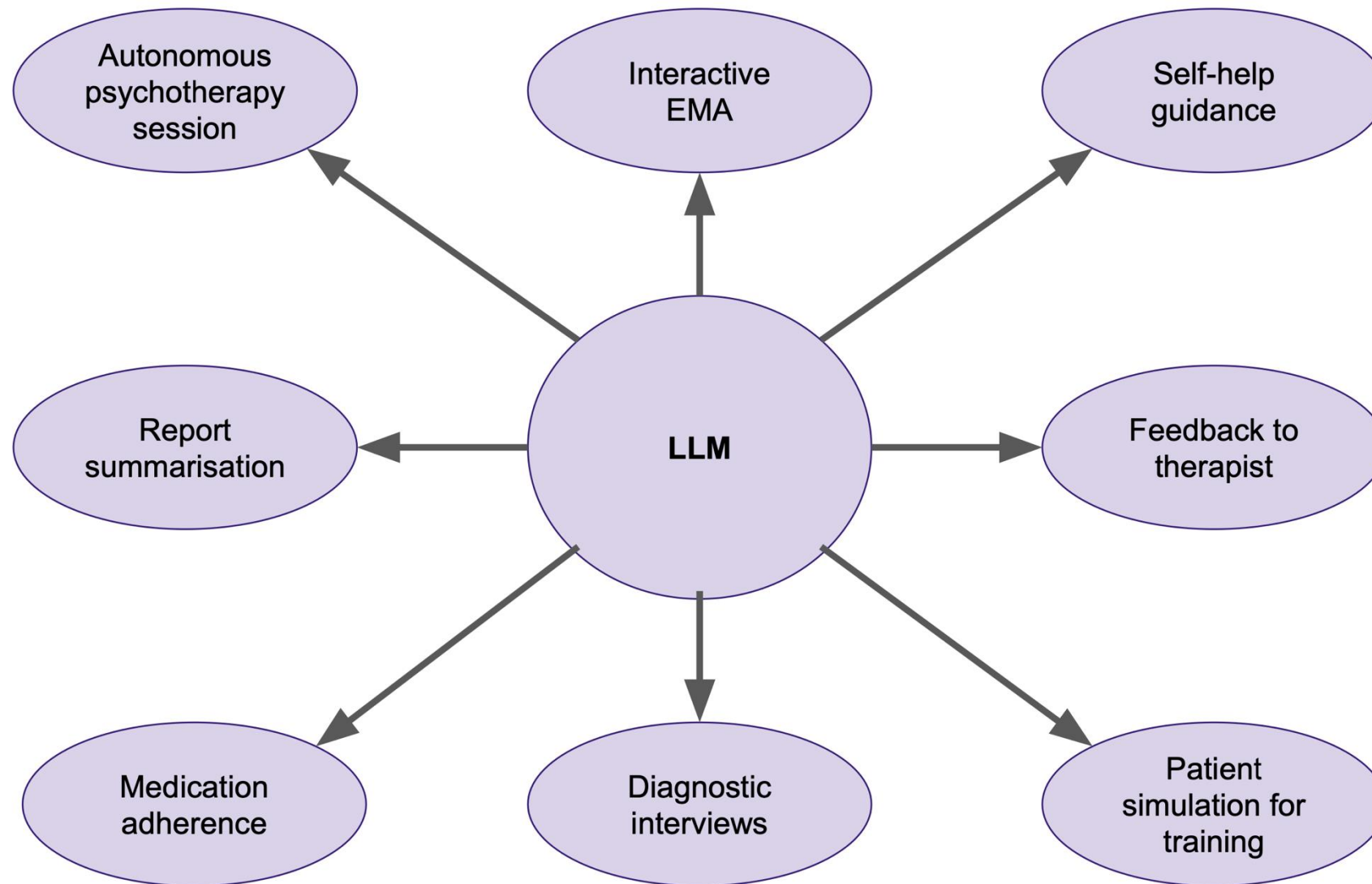
TPS Fantasy Game – voice emotions charge up your ultimate healing spells and unleash magical power



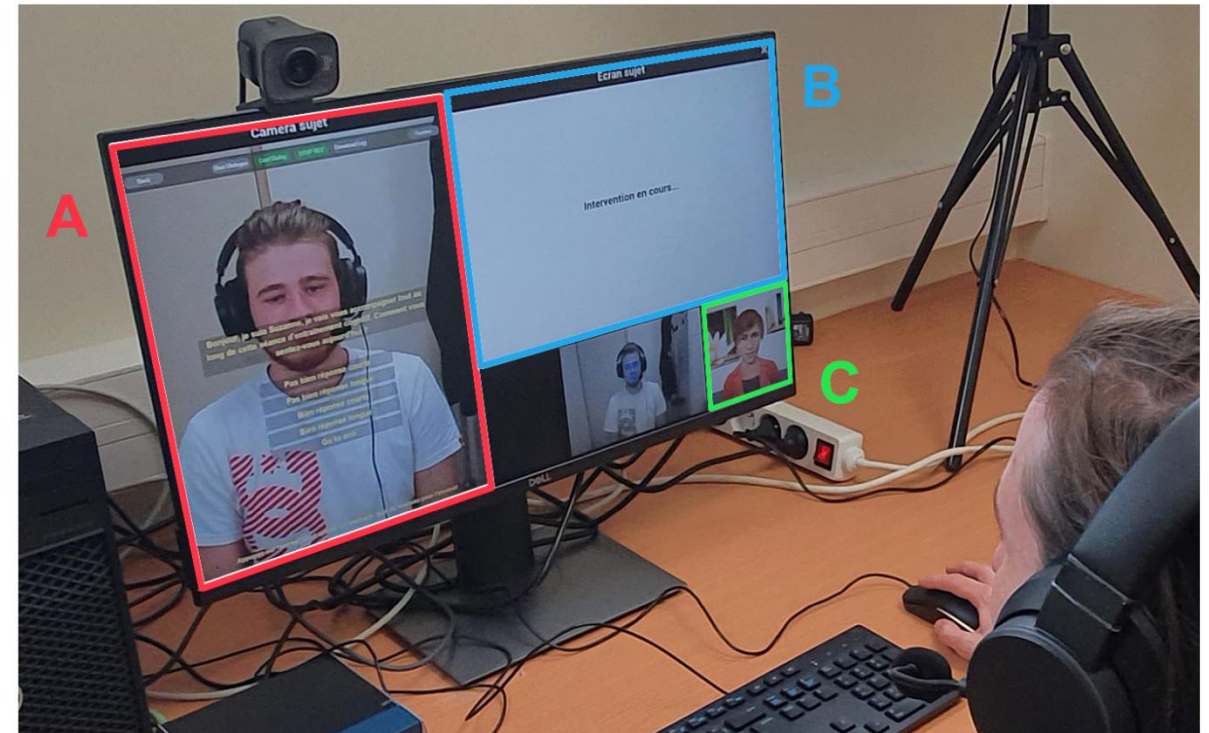
Emotions from voice trigger magic spells, unlock cursed treasure chests, conjure a vast variety of magic runes



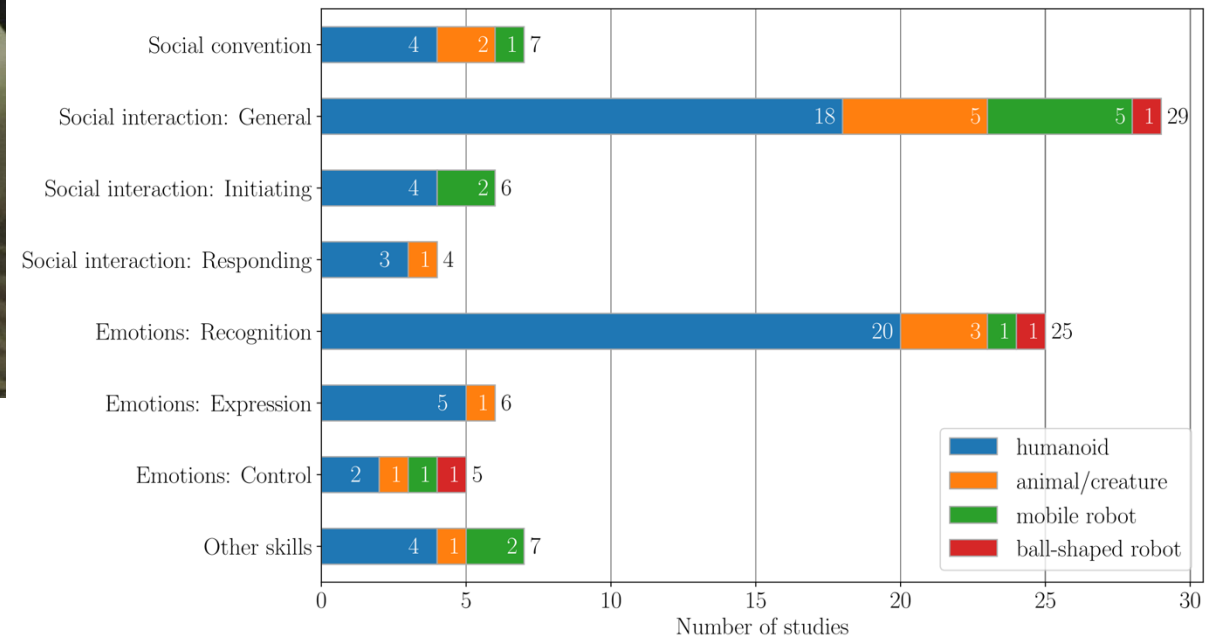
Arcade-style VR game, more enraged → higher damage, but relaxed voice helps regenerate faster



Therapy.

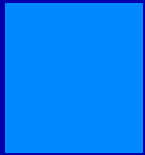


Therapy.

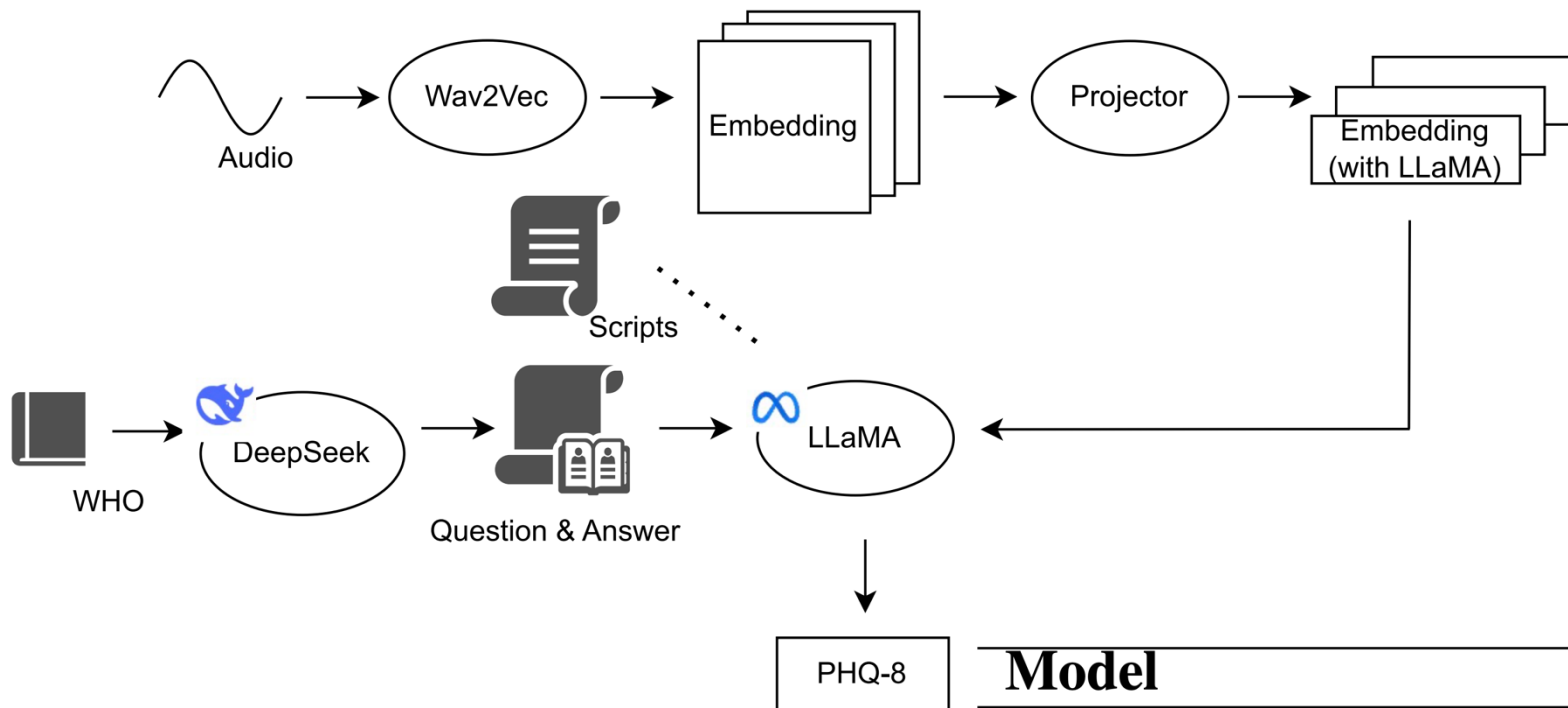


“Robot-Based Intervention for Children With Autism Spectrum Disorder: A Systematic Literature Review”, IEEE Access, 2021.

Next?

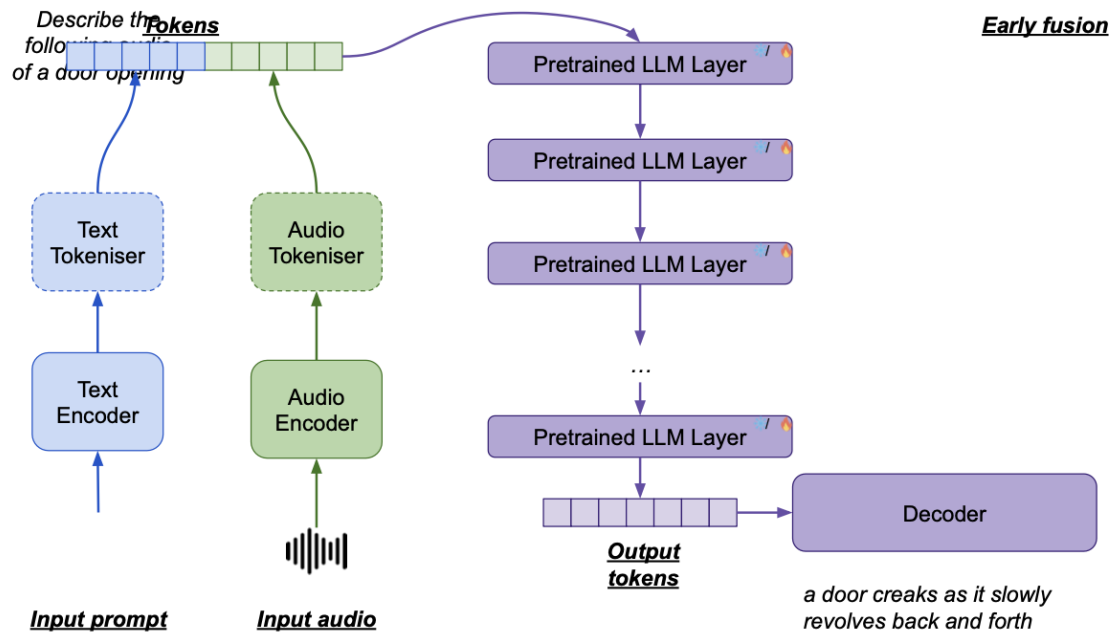


Knowledge Injection: ICD for PHQ

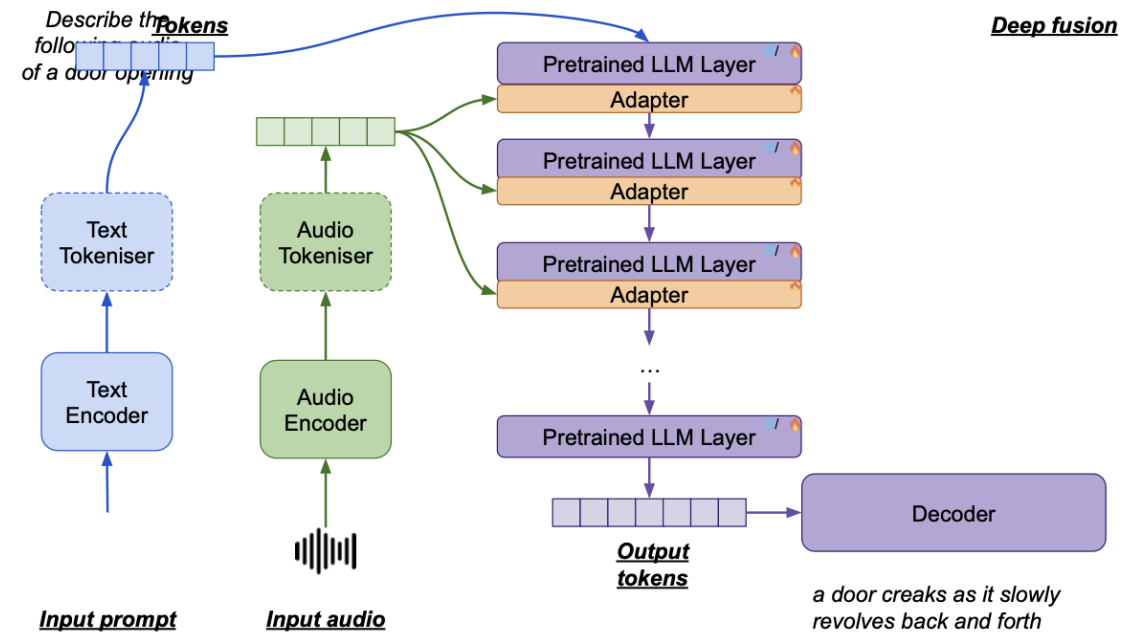


Model	MAE
AVEC2016 (audio) Valstar et al. (2016)	5.72
LSTM Afzal Aghaei and Khodaei (2023)	5.7
Random Forest Afzal Aghaei and Khodaei (2023)	5.71
Ours(audio)	5.373
Ours(text)	6.342
Ours(audio + text)	5.356
Ours(text + Knowledge Injection)	5.354

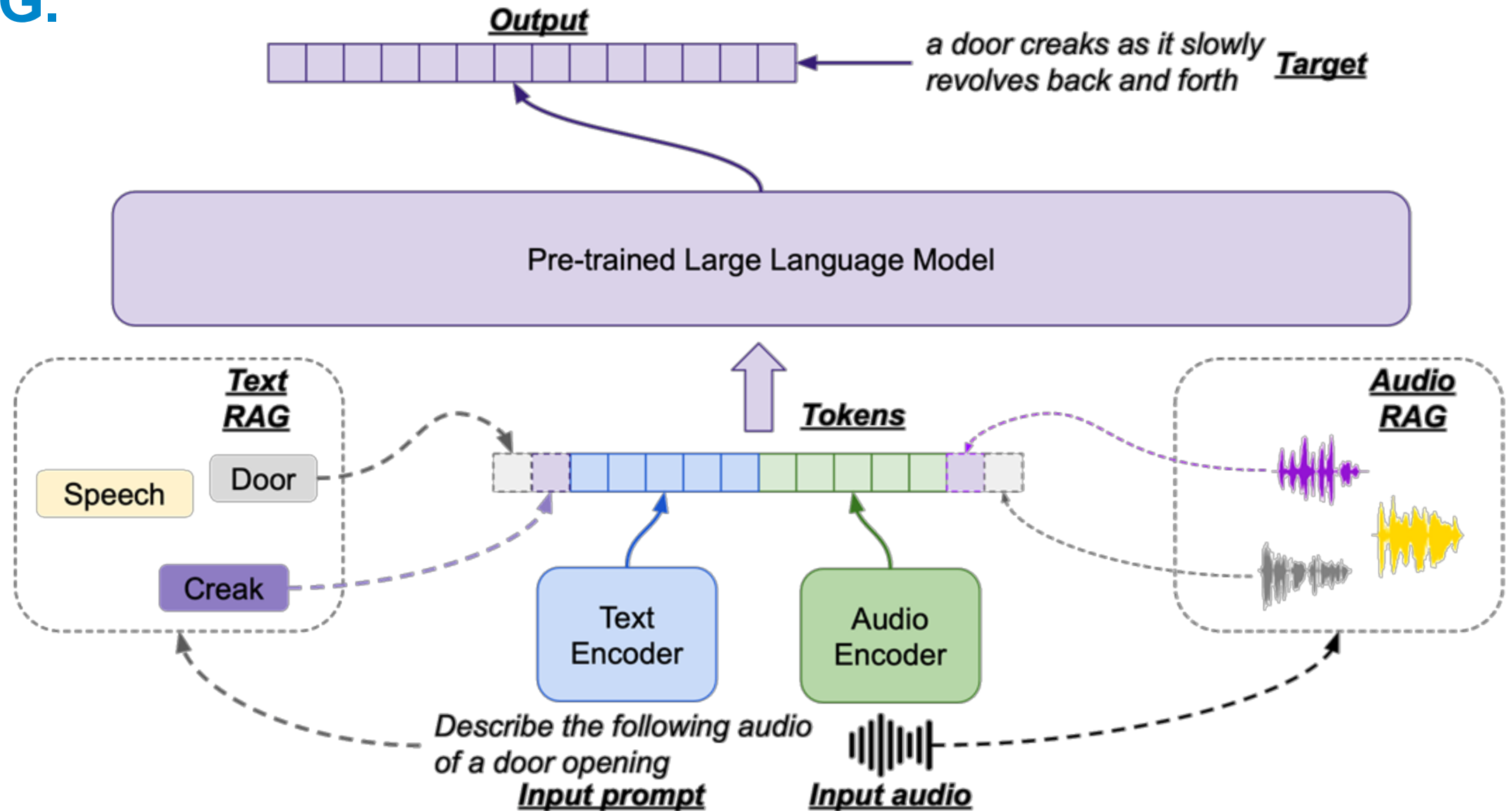
Model Fusion & “Recycling”.



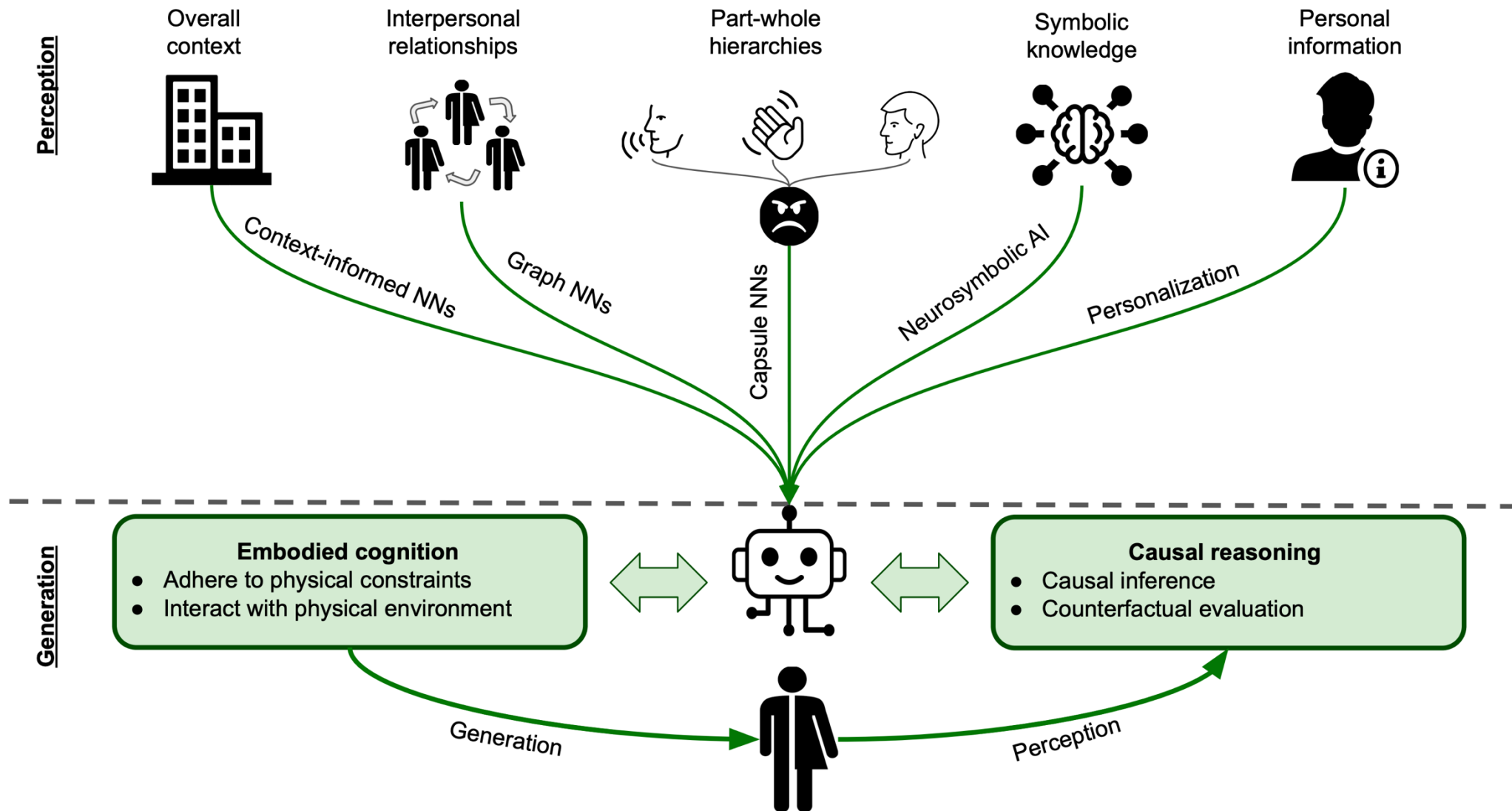
(a) Early fusion



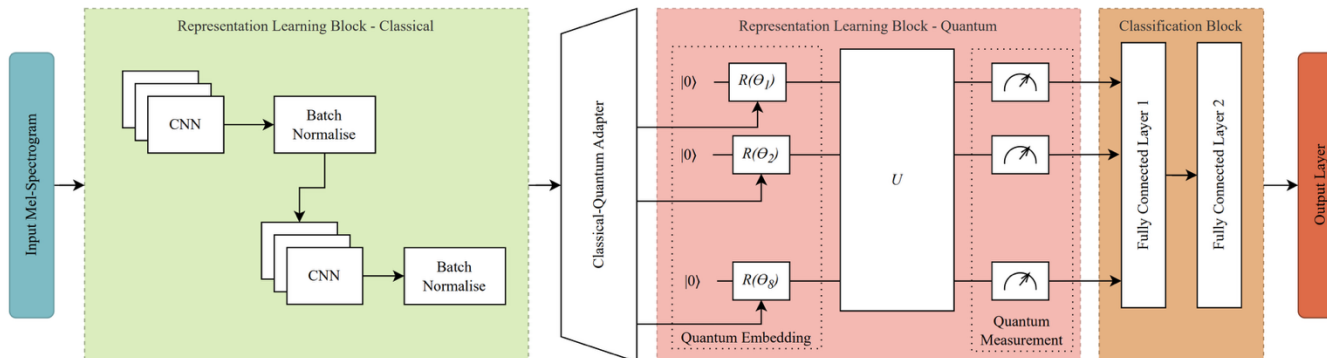
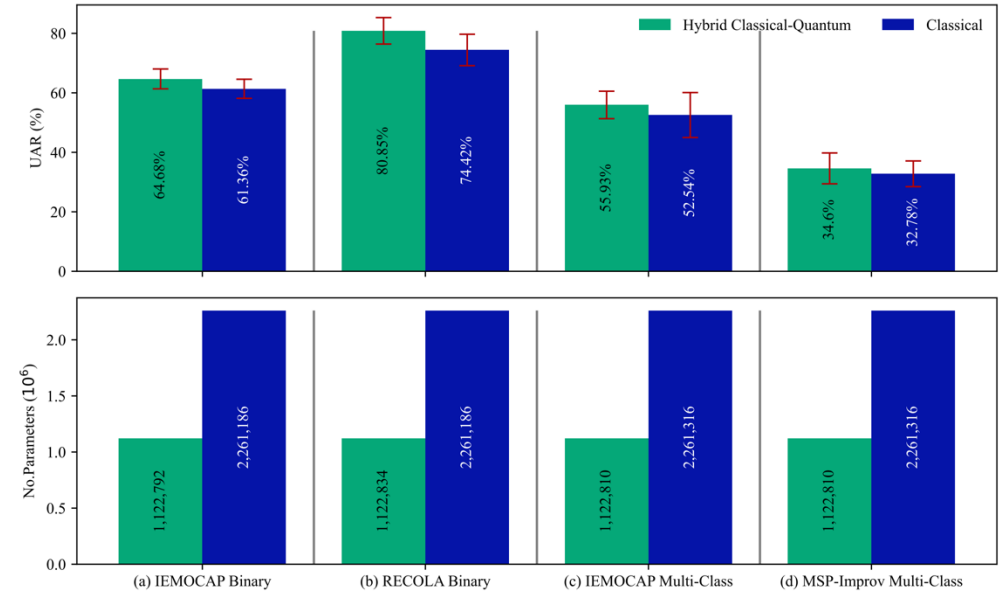
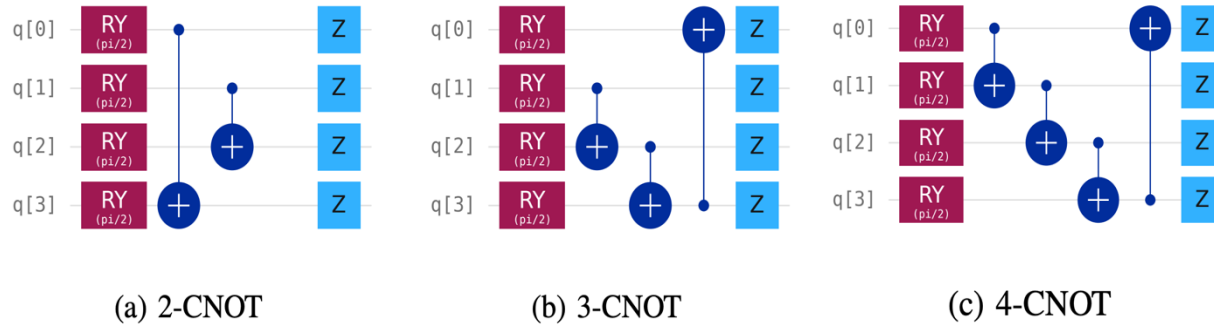
(b) Deep fusion



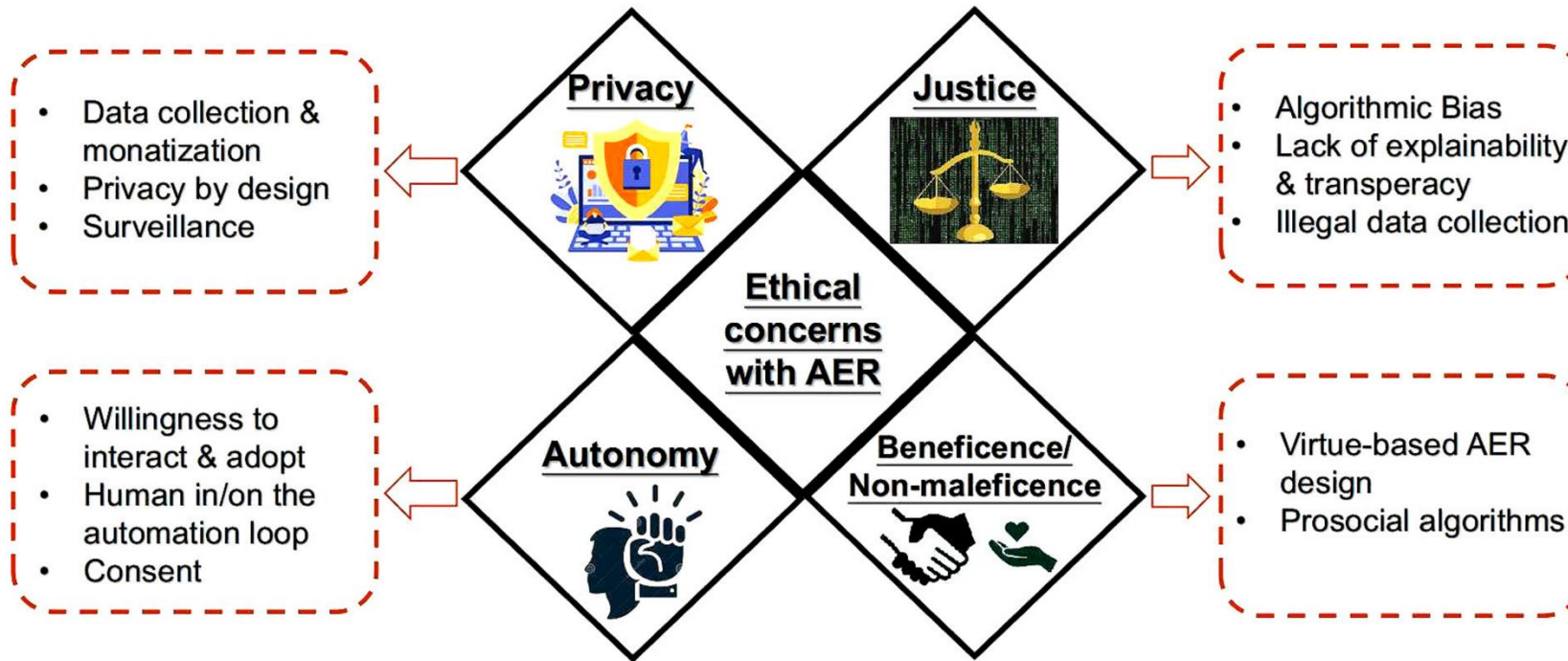
Beyond Deep Learning.



Quantum Computing.



“Representation Learning with Parameterised Quantum Circuits for Advancing Speech Emotion Recognition”, *arXiv*, 2025.



Ethical principles

Transparency

Full disclosure

Personal consent

Ethical data sharing

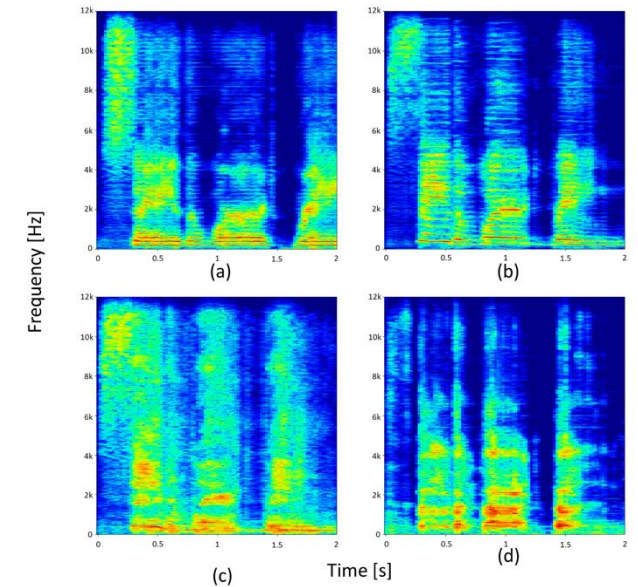
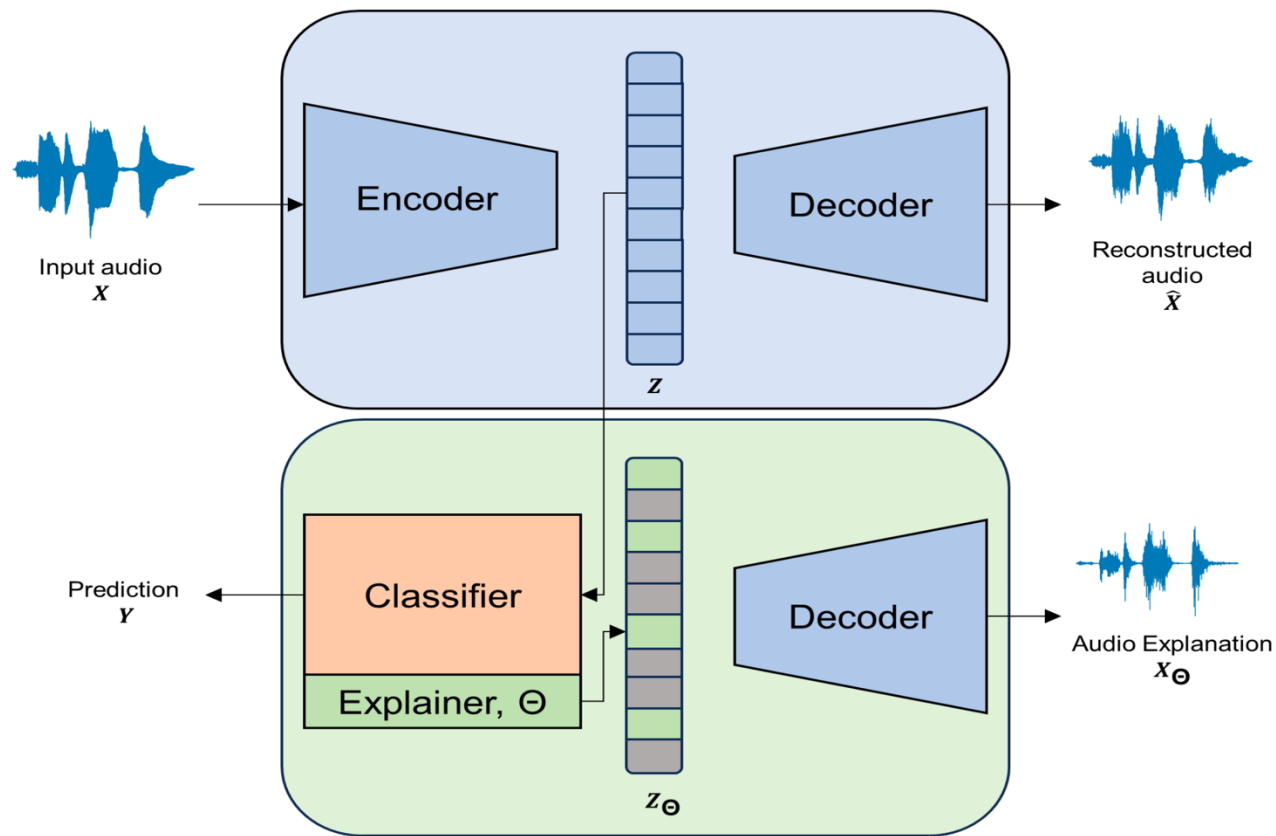
Data ownership

Security and privacy

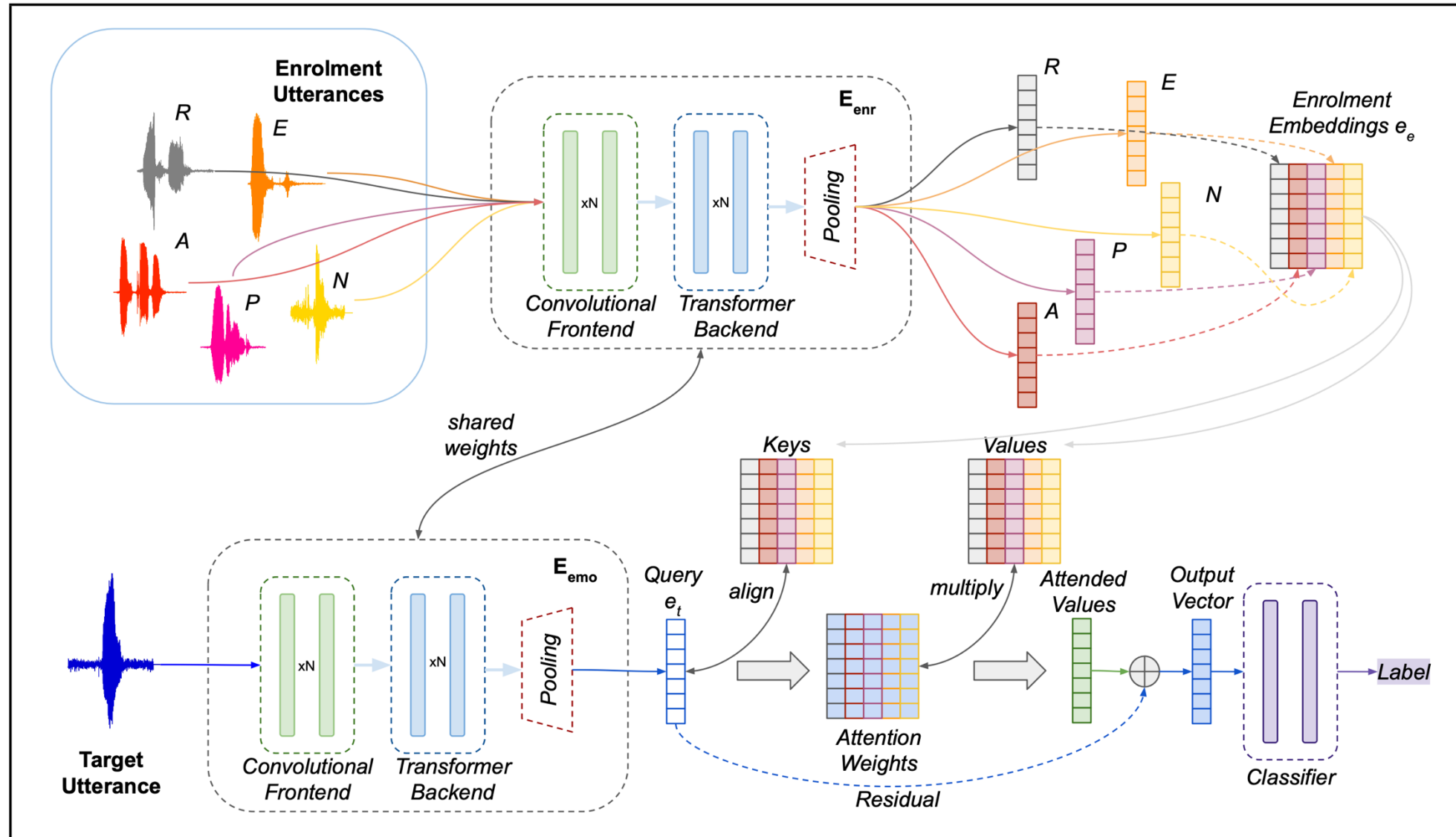
Ethics.

TEQUILA	Recommendation
Trust	Data security, privacy, and transparency in AI are crucial
Evidence	Robust evidence-based design is crucial
Quality	Ongoing regulatory oversight and continuous evaluation, including long-term user feedback and clinical validation
Usability	Ensuring ease of use and accessibility is crucial
Interest	Interventions must prioritise end-users' interests, involves addressing who funds, controls, and uses data, and incorporating peer support to ensure inclusivity
Liability	Legal considerations, especially around liability, are critical
Accreditation	Accreditation by relevant boards for 'Dr. AI' are essential

Explanations.

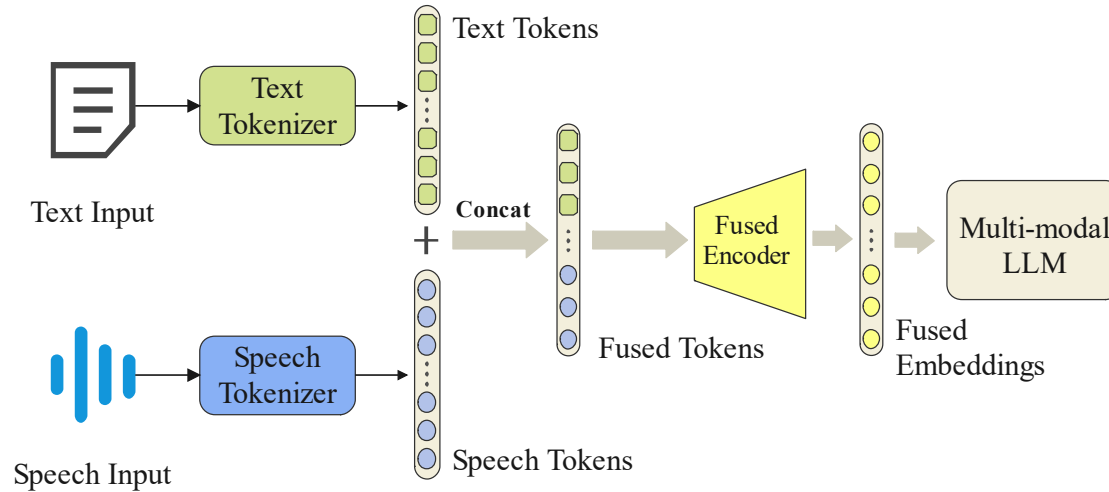


Fairness.

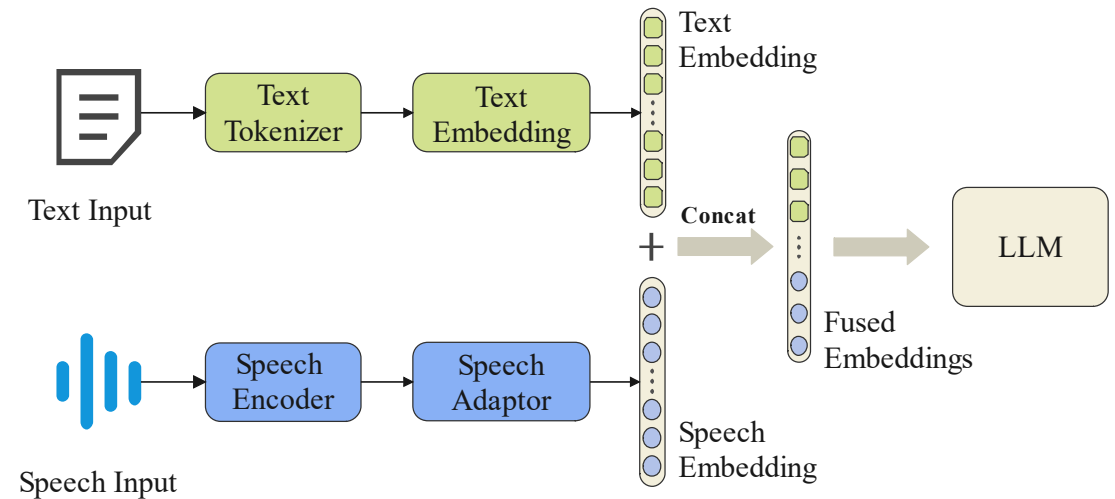


“Enrolment-based personalisation for improving individual-level fairness in speech emotion recognition”, *Interspeech*, 2024.

SLMs: Train your local own?

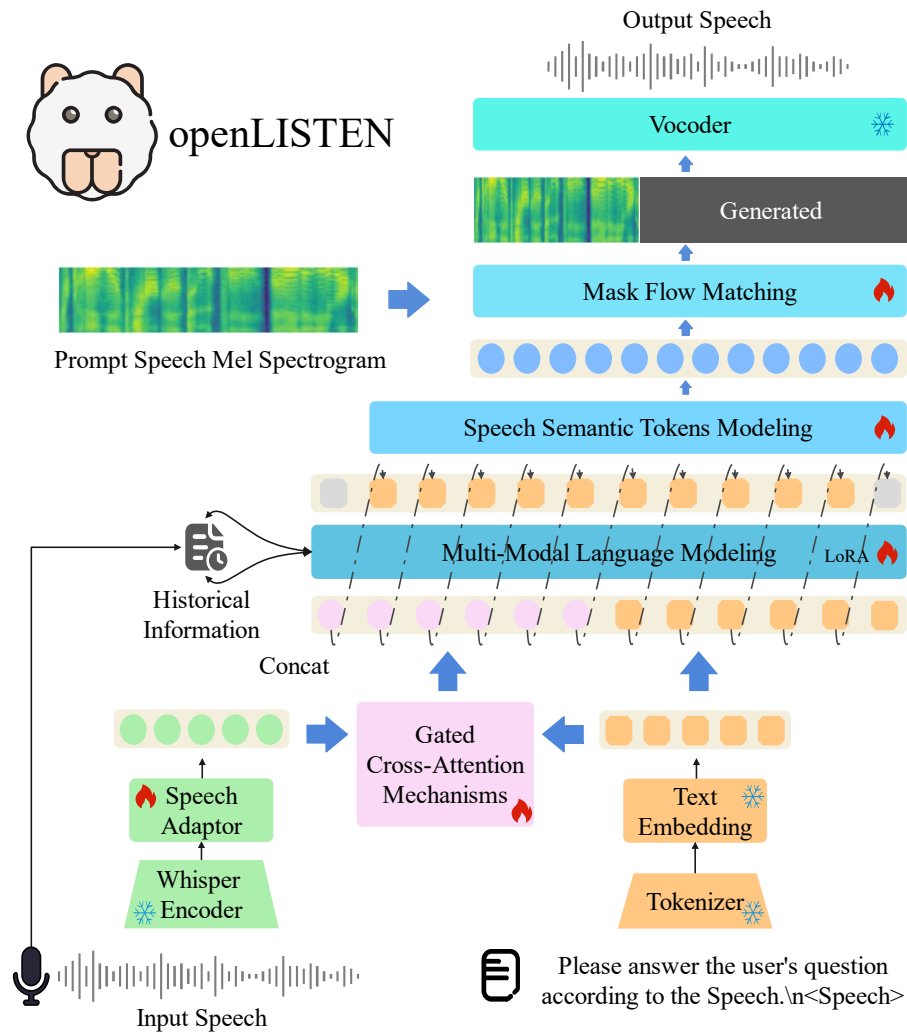


(a) Vocabulary Expansion Method

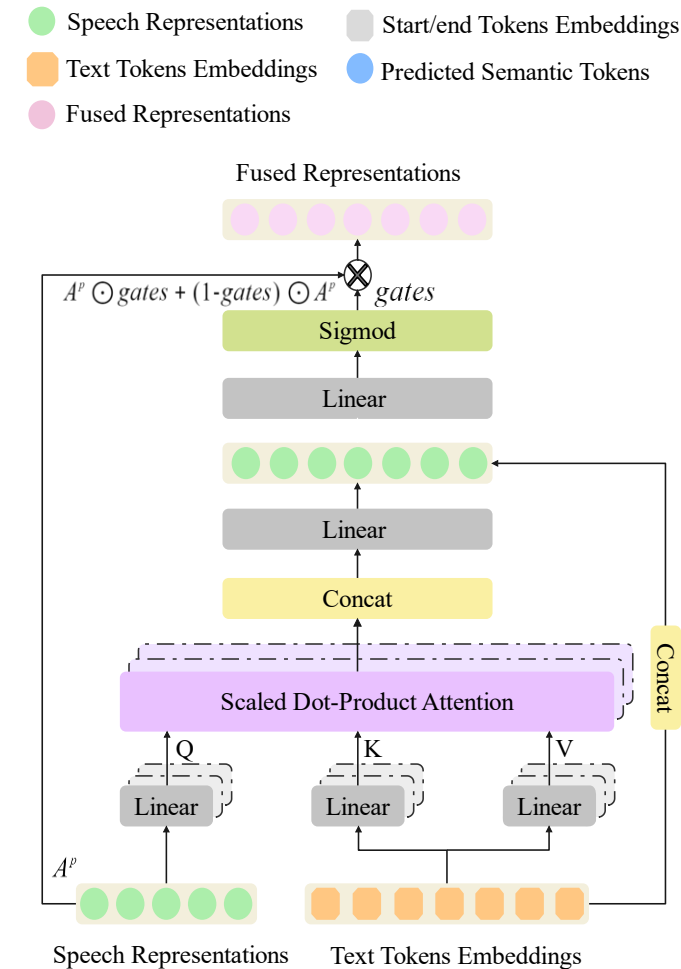


(b) Direct Projection Method

SLMs: Train your local own?



(a) Overall Structure



(b) Gated Cross-Attention Mechanisms structure

THE END

Next Episode:
Emotional AI.

*G
r
a
c
i
a
s*

