



GPT-2

My GPT-2 is XTRA LARGE

[Enlarge yours too!](#)

FUNBOY

don't forget to wear a mask in public



The image is a dense collage of text and images. At the top left is a logo for 'GPT-2' with a stylized 'G'. Below it is a snippet of text from a paper by Dr. Pérez. The center features a large, bold 'FUNBOY' with the tagline 'don't forget to wear a mask in public'. To the right is a cartoon illustration of a person wearing a mask and holding a megaphone. Other snippets include a 'Burried Lines' logo, a 'Civil War' logo, and various text fragments about AI, fashion, and the Civil War.

THE QUESTION WE ASKED

what's your favourite hookup line?

How do people react to jokes produced by a Neural Language Model conditioned on humour data and delivered by a robot in a verbal interaction?



COMPUTATIONAL HUMOUR

who says humour cannot be sad

- Punning riddles generation ("JAPE", 1994):
 - What do you call a quirky quantifier? An odd number.
- Rule-based production system ("STANDUP", 2007)
 - What do you call a cry that has pixels? A computer scream.
- 2-gram semi-supervised generator ("GAG", 2018)
 - I like my coffee like I like my war, cold



Curse of dimensionality

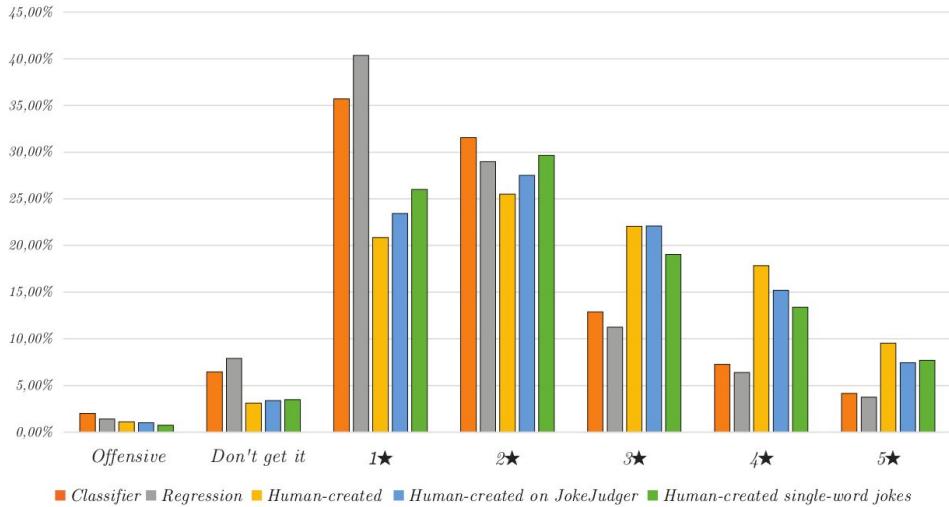


GAG EVALUATION

so why are people afraid of AI

Source	4+ ratings
GAG (Classifier)	11.41%
GAG (Regression)	10.12%
Human (All)	27.38%
Human (JokeJudge)	22.61%
Human (Single words)	21.08%

Relative amount of ratings per rating type for different sources and categories of jokes during the evaluation phase



Source: A. Radford. 'Improving Language Understanding by Generative Pre-Training'. In: 2018



HUMOROUS HRI

what's a robot's favourite Mexican food?

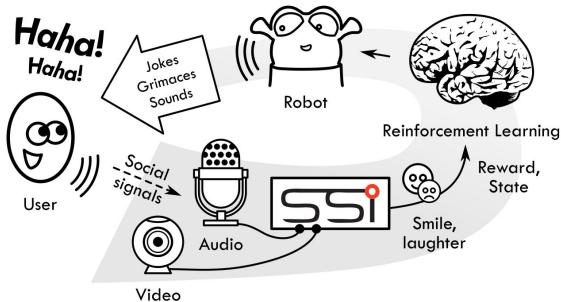


Figure 1: Interaction scenario involving a robot learning how to be funny from human social signals.



Figure 2: Some robot grimaces.

Source: K. Weber, H. Ritschel, I. Aslan, F. Lingenfelser and E. André. 'How to Shape the Humor of a Robot - Social Behavior Adaptation Based on Reinforcement Learning'. In: ICMI '18, 2018.

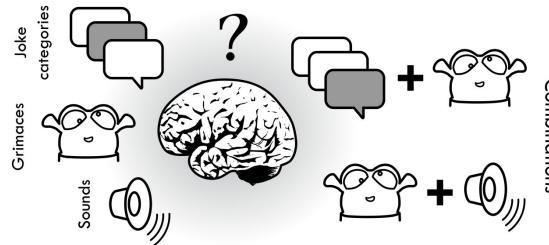


Figure 3: Robot learns the most efficient actions to make the user laugh.

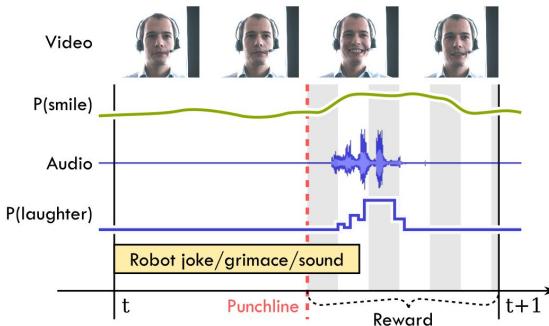
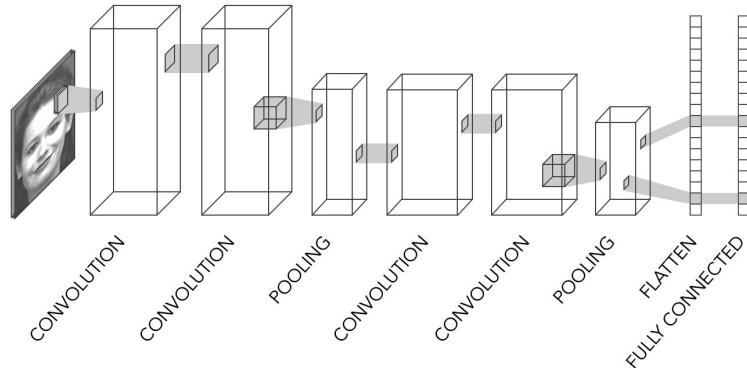


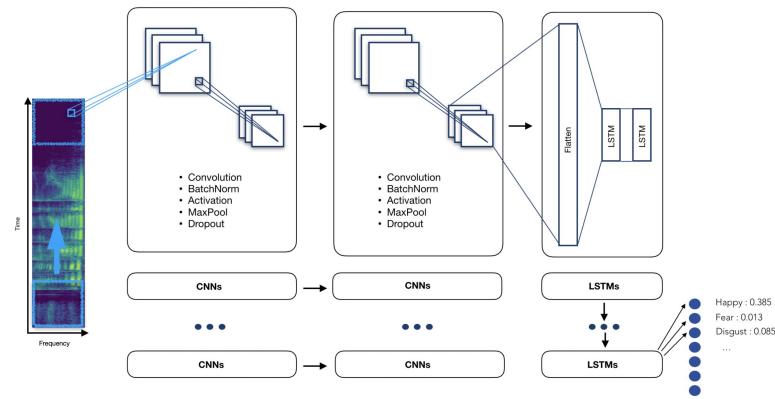
Figure 4: Reward calculation based on social signals.

EMOTION RECOGNITION

what kind of emotions do noses feel?



Facial Emotion Recognition - EmoPy:
CNN



Speech Emotion Recognition - SER:
Time-distributed CNN

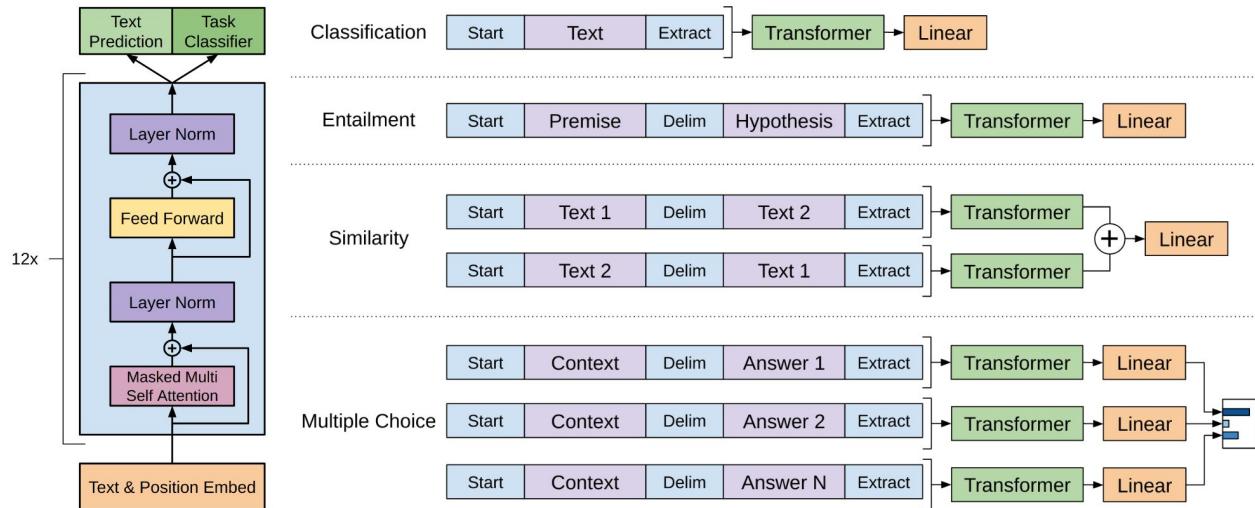
Emotions: calm, anger, happiness, surprise, disgust, fear, sadness

Source (R): T. Arts. EmoPy. In: GitHub Repository, 2018. url: <https://github.com/thoughtworksarts/EmoPy>.

Source (L): M. Fabien. Multimodal-Emotion-Recognition. In: GitHub Repository, 2019. url: <https://github.com/maelfabien/Multimodal-Emotion-Recognition>.

LANGUAGE MODEL

but can it speak LOLPython?



GPT / GPT-2 architecture:
Transformer layer with Masked Self-Attention

Source: A. Radford. 'Improving Language Understanding by Generative Pre-Training'. In: 2018



THE GOALS WE ACHIEVED

in life were disappointment and discomfort

- ✓ Roboy performs non-typed proactive humour.
- ✓ Roboy performs non-typed reactive humour.
- ✓ Roboy can perform emotion analysis based on vision and speech.
- ✓ Roboy can combine emotional responses with humour generation to adjust joke scores in personal profiles.
- ✓ Roboy can imitate laughter for humour delivery purposes.
- ✓ A dataset of hardcoded and pre-generated jokes.
- ✓ A command-line tool for the GPT-2 model.
- ✓ An http server for the GPT-2 model.
- ✓ Funboy Joker Bot.
- ✓ Roboy performs >2 types of humour.

1.0 goals

Vision
goals



THE IDEAS WE CONCEIVE

glory to the god of acronyms

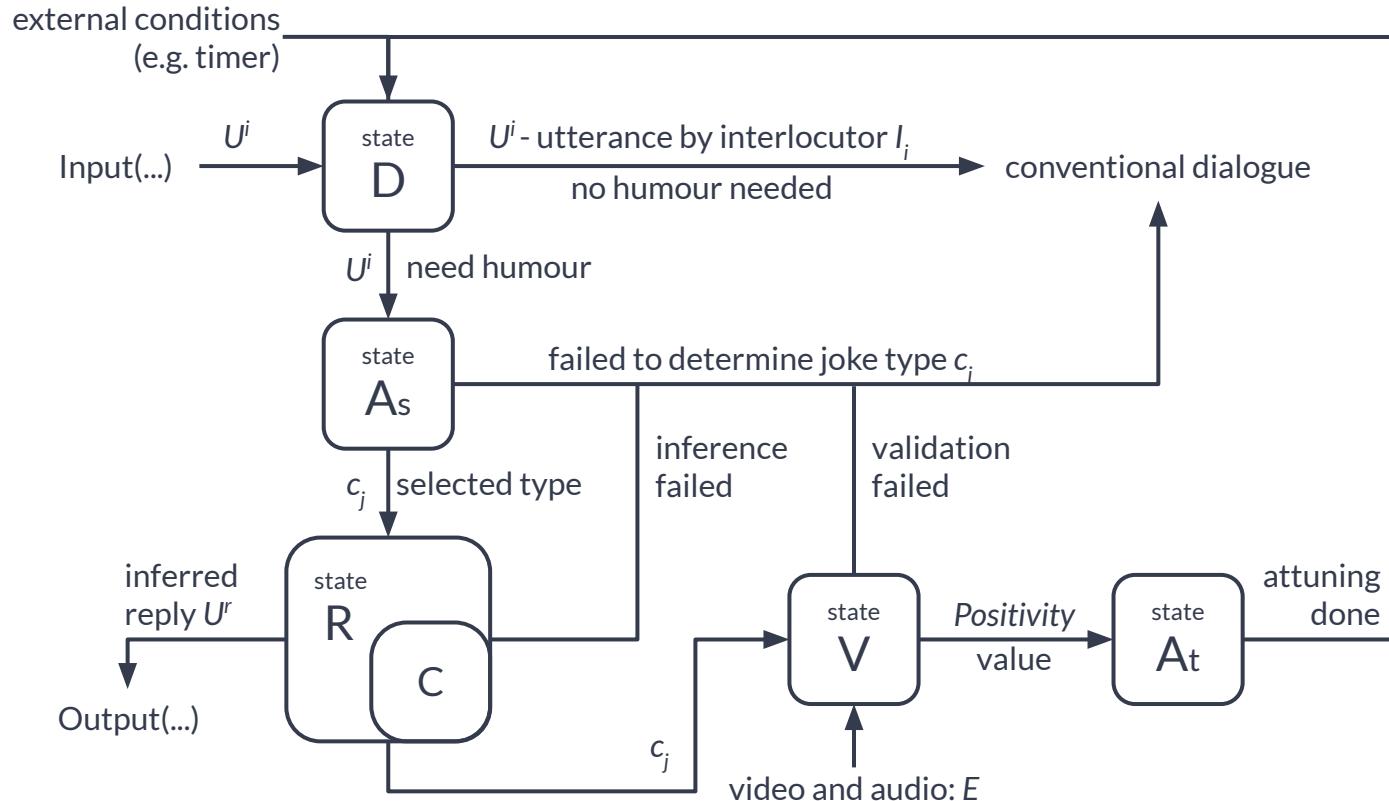
DARVAH

- D - deduce if the situation requires a humorous response
- A - assess which type of humour would be acceptable
- R - render a joke according
- V - validate how the interlocutor reacts
- A - update the personal profile of the interlocutor
- H - the 'h' is silent



SYSTEM DESIGN

is like yo, momma style - dead, yet still works



DARVAH STATE D

to choose or not to choose

state D

Possible triggers:

- user says “tell me a joke”
- input contains a joke type token
- timer
- threshold for a random number generator
- humour recognition



DARVAH STATE As

what a joke!

state A_s

$$c_j = \text{Assess}(S_i)$$

where

$$\text{Assess}(S_i) = \begin{cases} \text{choice}(C_j^+), & \text{if } \text{Validate}(E_{t-2}^i[m]) \geq 1 \\ \text{choice}(C_j^-), & \text{if } \text{Validate}(E_{t-2}^i[m]) < 0 \end{cases}$$



DARVAH STATE R

quack, quack - says a duck

state R

$$Render(c_j, U_{t-1}^i) = \begin{cases} Comedian(c_j, U_{t-1}^i), & \text{if } F_d(c_j, U_{t-1}^i) = True \\ Comedian(c_j), & \text{otherwise} \end{cases}$$



DARVAH STATE V

what is the biggest bean in the world?

state V

$$\text{Positivity}(E_t^i) = \begin{cases} 1, & \text{if } \sum_{n=1}^N E_{t,n}^{i+} > \sum_{n=1}^N E_{t,n}^{i-} \\ 0, & \text{if } \sum_{n=1}^N E_{t,n}^{i+} = \sum_{n=1}^N E_{t,n}^{i-} \\ -1, & \text{if } \sum_{n=1}^N E_{t,n}^{i+} < \sum_{n=1}^N E_{t,n}^{i-} \end{cases}$$

where N is the number of recognised emotions.

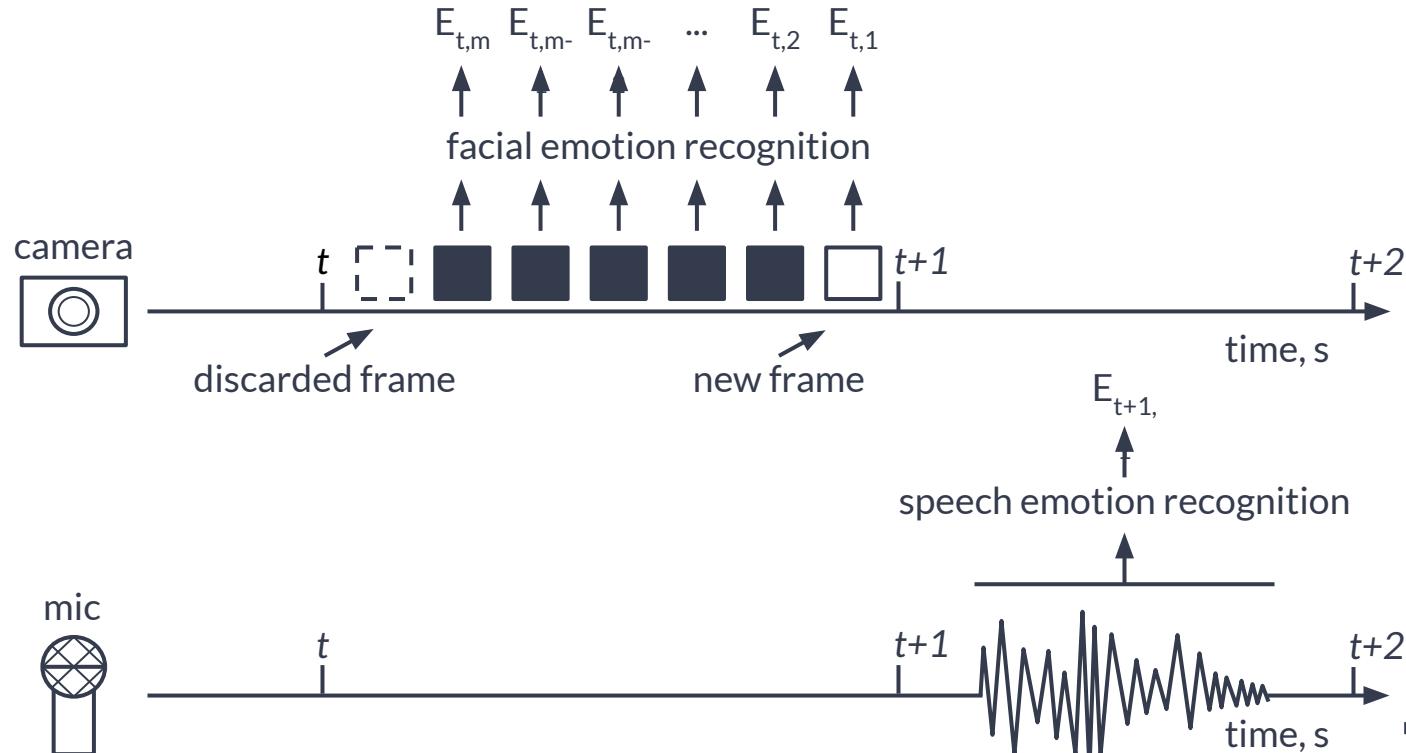
Then, for E_t^i - an array of m vectors E_t :

$$\text{Validate}(E_t^i) = \text{mode}\{\text{Positivity}(E_{t,1}^i), \text{Positivity}(E_{t,2}^i), \dots, \text{Positivity}(E_{t,m}^i)\}$$



EMOTION VECTORS

not quite sure how to feel about that



DARVAH STATE At

take it or leave it

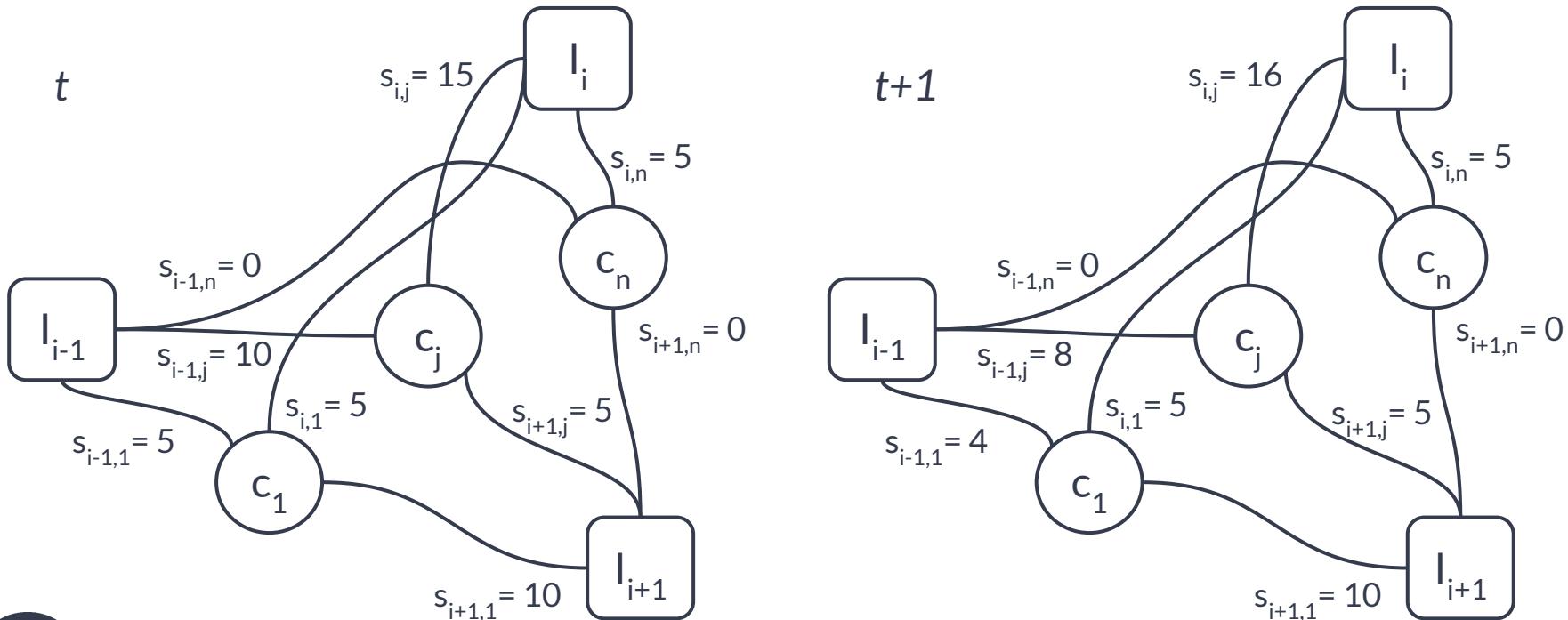
state A_t

$$Attune(S_i|c_j) = \begin{cases} s_{ij} = s_{ij} + 1, & \text{if } Validate(E_t^i) = 1 \\ s_{ij} = s_{ij}, & \text{if } Validate(E_t^i) = 0 \\ s_{ij} = max(s_{ij} - 1, 0), & \text{if } Validate(E_t^i) = -1 \end{cases}$$



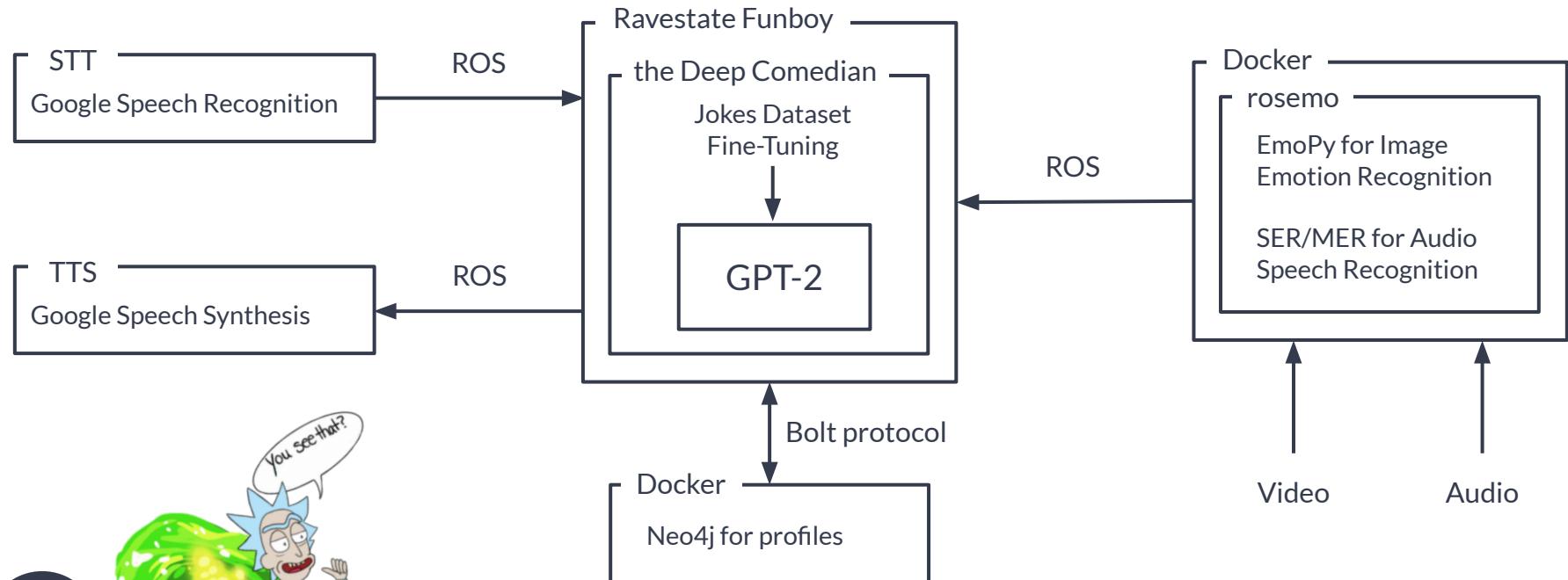
DARVAH ONTOLOGY

what's the best score Canadians get?



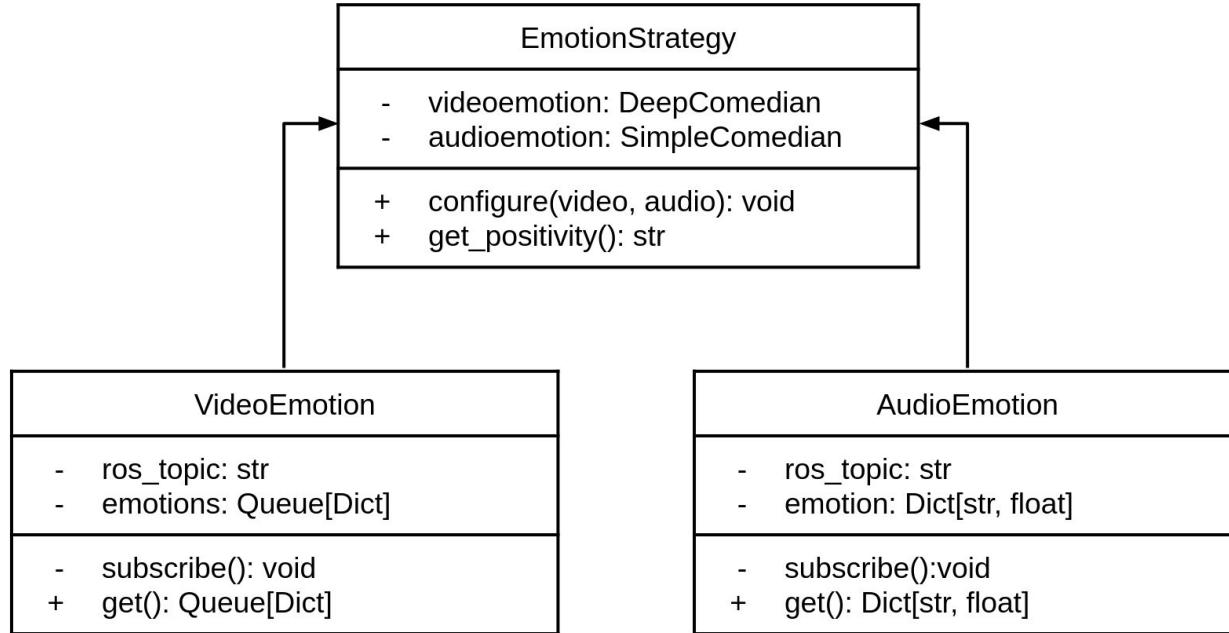
FUNBOY IMPLEMENTATION

it is funny doing 69 in a garbage bin



EMOTION STRATEGY

but it didn't stand up in court



VIDEO + AUDIO STRATEGY

call it a tic-tac-toe tactics

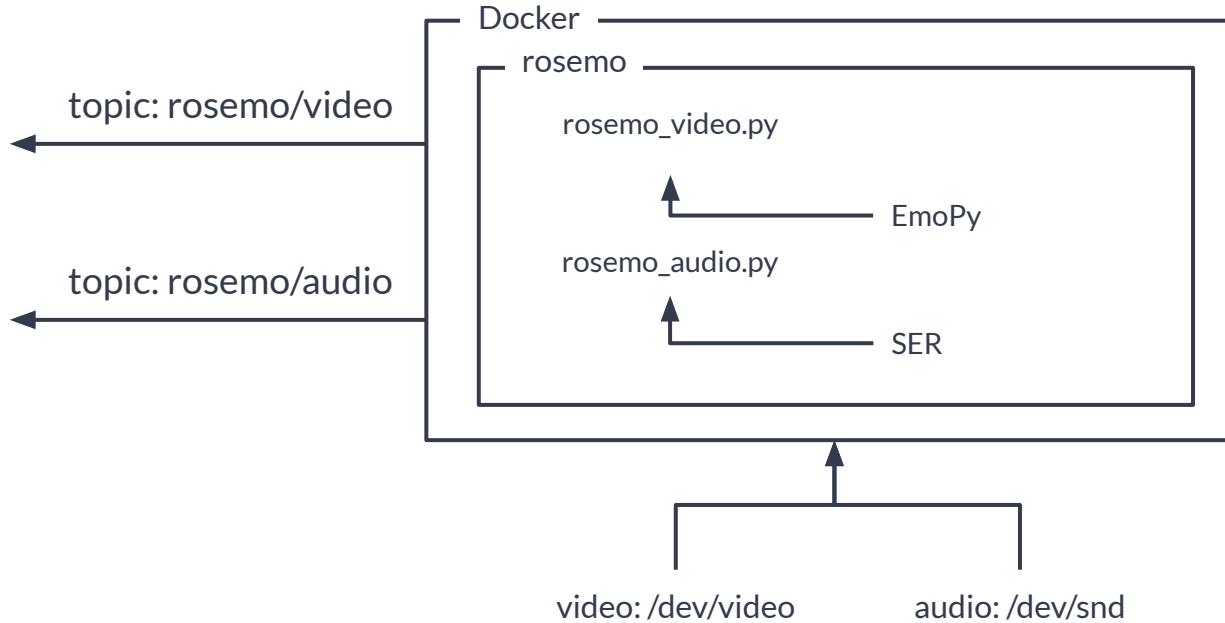
$$E_{t+2}^{i'} = \sum_{k=1}^m \alpha E_{t+2,k}^i + \beta E_{t+1}^i$$

where $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ are the weights for vectors $E_{t+2,k}^i$ and E_{t+1}^i respectively.



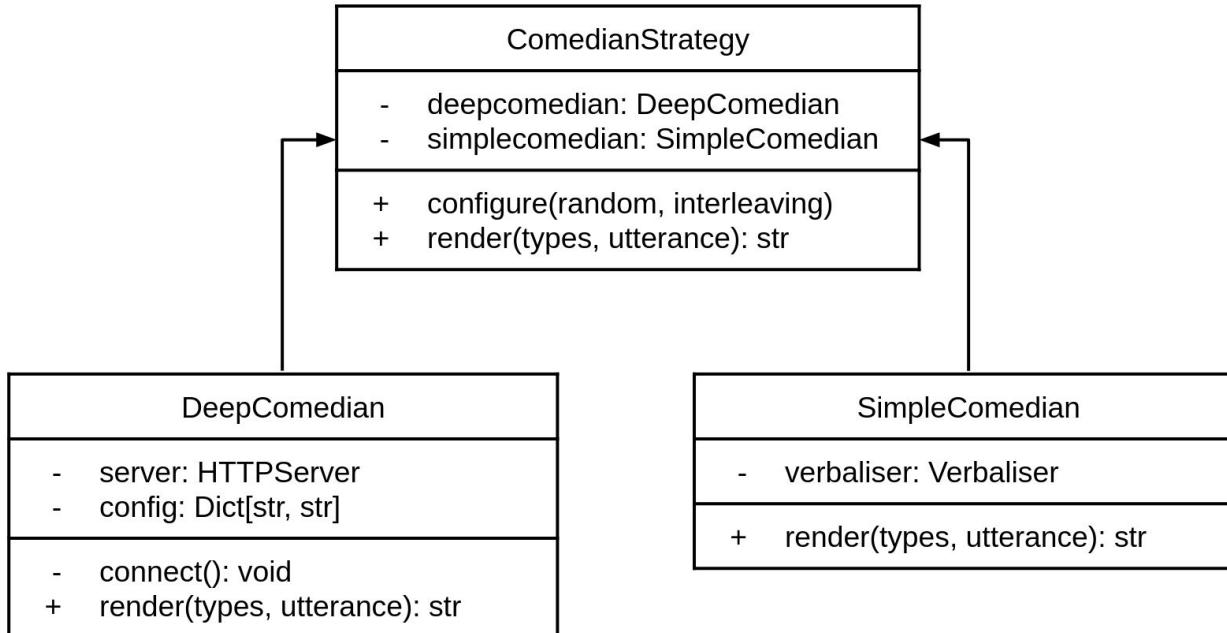
ROSEMO

are you hangry?



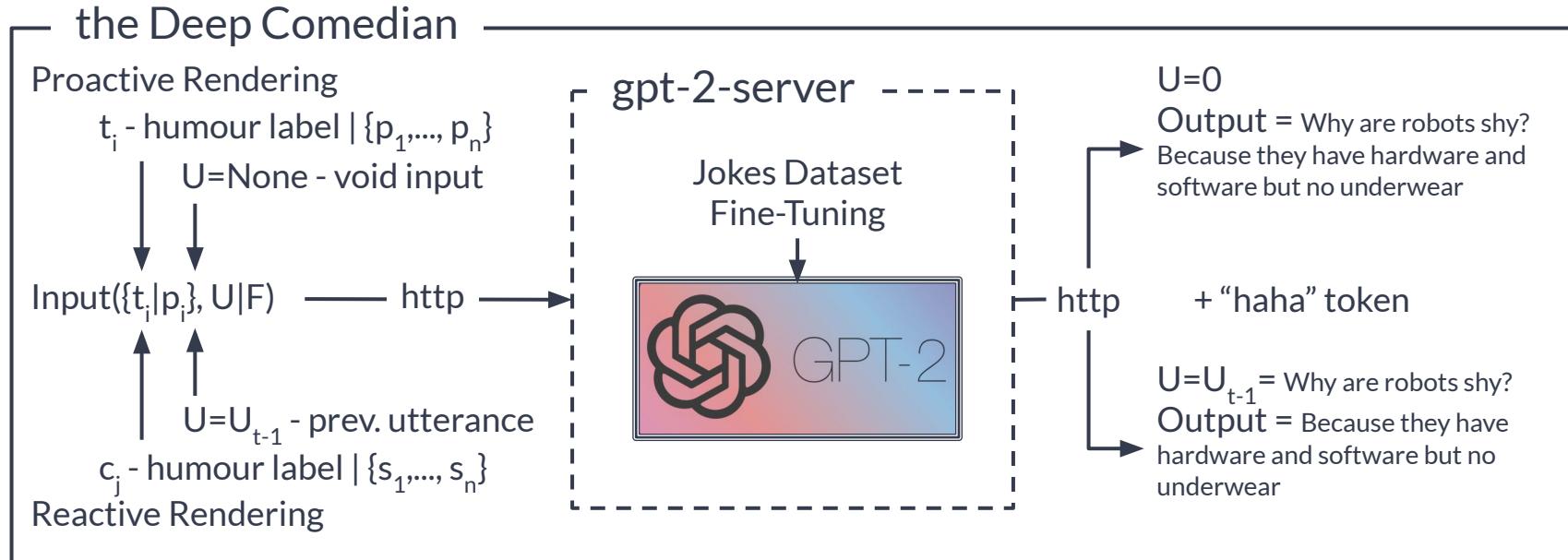
COMEDIAN STRATEGY

so a clumsy comedian walks into a music shop



DEEP COMEDIAN

one glove says to the other



JOKE TYPES

of dinosaurs building relationships

Original Way - Invariant

Size types: <|short|>, <|medium|>,
<|long|>, <|story|>

Content types:

<|I₁|> - chicken,
<|I₂|> - momma,
<|I₃|> - cookie...

Current Way - Vocabulary Dependent

Size types: <|short|>, <|medium|>,
<|long|>, <|story|>

Content types:

<|chicken|> - chicken,
<|momma|> - momma,
<|cookie|> - cookie...



REACTIVE RESPONSES

i am selling my body for them

Original

Input: Why are robots shy?

Tokens: <|short|> <|**I**|>Why are robots shy?

Output: They have lower case letters than police.

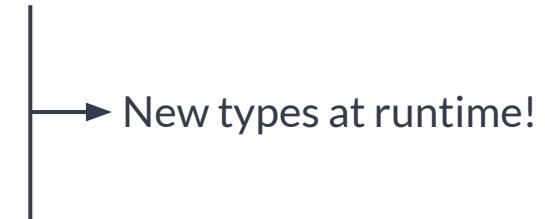


Current

Input: Why are robots shy?

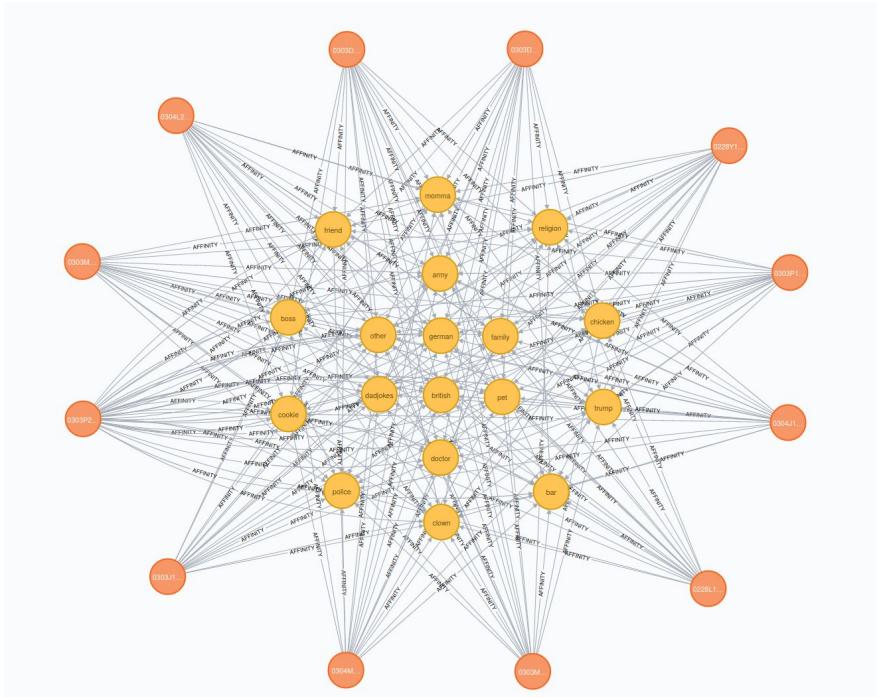
Tokens: <|short|> <|**robots**|>Why are robots shy?

Output: The only job robot does well is driving a bus.



FUNBOY SUBONTOLOGY

ask the NSA for a backup



!OType
entity: Person
properties: [...]
relationships: [..., AFFINITY]

!OType
entity: Object
properties: [name, timestamp]
relationships: [EQUALS]



EXPERIMENT SETUP

what did a pie scientist use for an experiment?

1-on-1 conversation.

The conversation flow:

1. The greeting.
2. The conventional dialogue.
3. The humour-enabled dialogue.
4. The farewell.



EXPERIMENT PERFORMANCE

we use cookies to improve performance

- Speech synthesis took on average around 1 second.
- EmoPy could recognise one emotion sample in 200 milliseconds.
- For the SER, it took around 400 milliseconds to recognise a wave sample of 10 seconds (the time scales linearly).
- The <|short|> joke type took 3 seconds on average to generate a sample while the <|medium|> type jokes were taking over 5 seconds.



EXPERIMENT RESULT

i don't want to eat it but i can handle it

Curious → Confused* → Frustrated*



*have you ever noticed that Google STT is not that great?



FACIAL EMOTION RECOGNITION

of carbonator to subatomic particles

	$t_{\text{end-4}}$	$t_{\text{end-3.5}}$	$t_{\text{end-3}}$	$t_{\text{end-2.5}}$	$t_{\text{end-2}}$
Calm	0.9649	0.9546	0.9454	0.9260	0.9721
Anger	0.0008	0.0010	0.0010	0.0014	0.0006
Happiness	0.0228	0.0287	0.0345	0.0427	0.0223
Surprise	0.0003	0.0005	0.0006	0.0008	0.0001
Disgust	0.0099	0.0135	0.0168	0.0256	0.0044
Fear	5.1391e-05	8.2555e-05	0.0001	0.0001	4.4135e-05
Sadness	0.0010	0.0013	0.0013	0.0031	0.0002

Sample EmoPy subframe with biased recognition



SPEECH EMOTION RECOGNITION

of carbonator to subatomic particles

	Disgust	Happy
Calm	14	100
Anger	14	0
Happiness	0	0
Surprise	0	0
Disgust	42	0
Fear	0	0
Sadness	28	0

Sample SER results with biased recognition



RECOGNITION PROBLEMS

you don't know me!

EmoPy:

- might be sensitive to uneven lighting and shadows on faces
- the FER+ dataset contains exaggerated facial expression

SER:

- might be sensitive to background noise and quiet speaking
- was trained on native speakers data



RESPONDERS LIKES

what are my choices?

The participants liked the following features about Roboy:

- voice
- facial animations, looks very cute, mimics
- asking personal questions, entertaining conversation
- some jokes were funny, diverse humour, dark humour
- understandable sentences



RESPONDERS DISLIKES

I just can't deal with it

The participants didn't like these characteristics of Roboy:

- voice
- response delays
- hard to distinguish between processing and listening
- impossible to follow up or repeat jokes
- forgetful, doesn't keep full context of the conversation
- incoherent answers, doesn't always understand input



REPORTS ON ANTHROPOMORPHISM

to err is human

Rate the participant's impression of the robot according to the following scale [-2, 2]:

- Q 1.1. From Fake: -2 to Natural: 2.
- Q 1.2. From Machine-like: -2 to Human-like: 2.
- Q 1.3. From Unconscious: -2 to Conscious: 2.
- Q 1.4. From Artificial: -2 to Life-like: 2.

Anthropomorphism		
Q	Mean μ	Std. Deviation σ
1.1	0.08	1.04
1.2	-0.08	1.32
1.3	0.15	0.99
1.4	-0.38	1.04



REPORTS ON LIKEABILITY

the joke's on you

Rate the impression of the robot according to the following scale [-2, 2]:

- Q 2.1. From Dislike: -2 to Like: 2.
- Q 2.2. From Unfriendly: -2 to Friendly: 2.
- Q 2.3. From Unkind: -2 to Kind: 2.
- Q 2.4. From Unpleasant: -2 to Pleasant: 2.
- Q 2.5. From Awful: -2 to Nice: 2.

Likeability		
Q	Mean μ	Std. Deviation σ
2.1	0.85	1.14
2.2	0.38	1.45
2.3	0.31	1.25
2.4	0.23	1.36
2.5	0.54	1.13



REPORTS ON ACCEPTANCE

apparently some have even taken it to heart

Rate the impression of the robot according to the following scale [1, 5]:

- Q 3.1. I enjoyed the robot talking to me.
- Q 3.2. I find the robot fascinating.
- Q 3.3. Talking with the robot was easy.
- Q 3.4. I consider the robot a pleasant conversational partner.
- Q 3.5. I find the robot pleasant to interact with.
- Q 3.6. I feel the robot understands me.
- Q 3.7. I think the robot is funny.

Acceptance		
Q	Mean μ	Std. Deviation σ
3.1	3.62	0.96
3.2	4.08	0.86
3.3	2.92	0.76
3.4	2.77	0.93
3.5	3.46	0.97
3.6	2.38	1.04
3.7	4.00	1.08



REPORTS ON EMOTIONS

I like my water like I like my emotions

Summary of Emotions			
User	Q 6.1	Q 6.2	Q 6.3
0228L160431P	curious, excited	excited, frustrated	disappointed
0228Y165728S	interested	confused	frustrated
0303D134007S	excited	confused, laughing	exhausted
0303D161701L	happy, curious	happy, confused	happy, surprised
0303J155544M	regular	amused	amused
0303M151400P	curious	chill, happy	a bit frustrated
0303M170134N	nice, curious	confused	a bit tired
0303P142946H	interested	confused, happy	disappointed
0303P201302N	excited	interested	bored
0303V145445S	curious, excited	a bit uncomfortable, excited	a bit annoyed
0304J142142B	excited, happy	impressed, interested	satisfied
0304L200631S	curious, excited	curious, fun	curious, fun
0304M152010H	curious, excited	frustrated, amused	frustrated



FEEDBACK

they asked for my feedback in a restaurant

- Use more facial expressions
- Use verbal confirmations in the humorous responses, such as:
 - “I know a joke...”
 - “Let me tell you a joke...”
 - “Here is a joke...”
- Introduce head movements into the delivery process
- “Haha” token improves the delivery



CHALLENGES

never challenge death to a pillow fight

- Data is scarce -> need to use crowdsourcing
- Data is dirty and cleaning takes time -> need to use crowdsourcing
- Hard to identify classes -> need to use crowdsourcing
- Computational cost is high -> GPT-2 needs to be distilled
- Delays are very important -> need to speed up computations
- GPT-2 can learn the twist but cannot grasp the essence -> need to have more advanced pre-processing
- GPT-2 learns offensive language better -> need more filtering
- Google ASR is imprecise -> need better speech recognition
- Emotion Recognition works only in ideal conditions -> need to apply transfer learning using noisy data



FUTURE RESEARCH

of an old joke

Generate jokes



Ask people to rate using
@FunboyJokerBot



Use for reinforcement learning

Do more experiments



Collect more video and audio
data biased towards Roboy



Transfer learn



LEARNINGS

a group of cats is called a clowder

- LaTeX formatting takes substantial time
 - **Allocate 20% more writing time**
- Writing a long paper in a non-native language is challenging
 - **Allocate 10% more time again**
- Do not try writing a thesis on a big (likely PhD) topic
 - **Choose one of its subsets instead**
- Sometimes small improvements require long research
 - **Improving the project further takes more than semester**

