

Multimodal, Personable, and Knowledgeable Dialog Generation Models

Mohit Bansal



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

(Nvidia-SAS-UNC DL Symposium, September 2018)

Dialog Model's Requirements in a Home



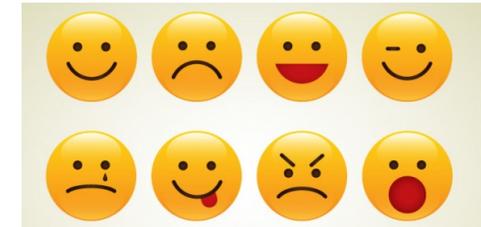
Inference in Long Conversation History

Commonsense and External Knowledge

User Satisfaction Feedback & Error Robustness

Human-Personality Convincing Responses

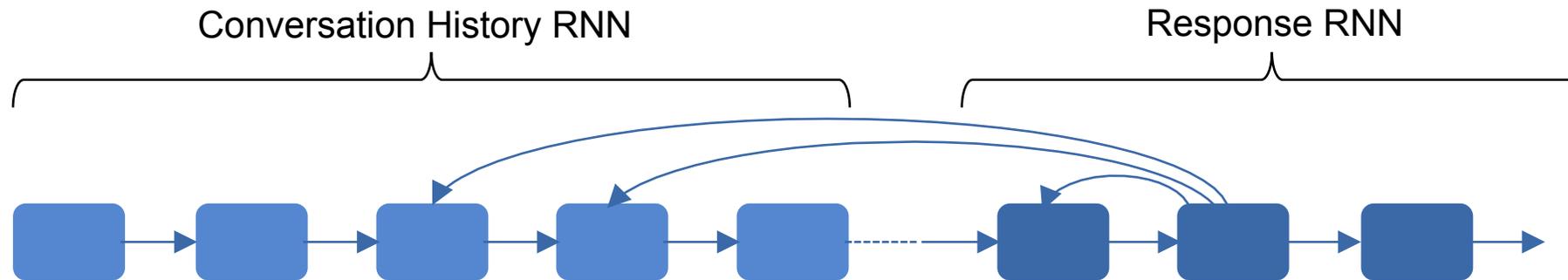
Many-modal Grounding in Home Surroundings+Tasks (Video, Databases, etc.)



Simple Attention-LM Dialog Model (no LSTM)



- Allows choosing long-distance history context w/o the issues w/ LSTM memories
- Allows attending to gappy history very flexibly
- Allows attending to already-generated response words for coherence and non-redundancy!



Simple Attention-LM Dialog Model (no LSTM)



A: person , but this time i got the money .
R: if you give it to me , i might forget i found you .
A: i ' ll **give** it to you .

A: person , but this time i got the money .
R: if you give it to me , i might forget i found you .
A: i ' ll give **it** to you .

(a) *MovieTriples*

A: i have really choppy streaming video in 9.10 ,
any way to fix that ?
A: any way to fix choppy streaming video in 9.10 ?
R: what kind of video
R: what **video** card do you have ?

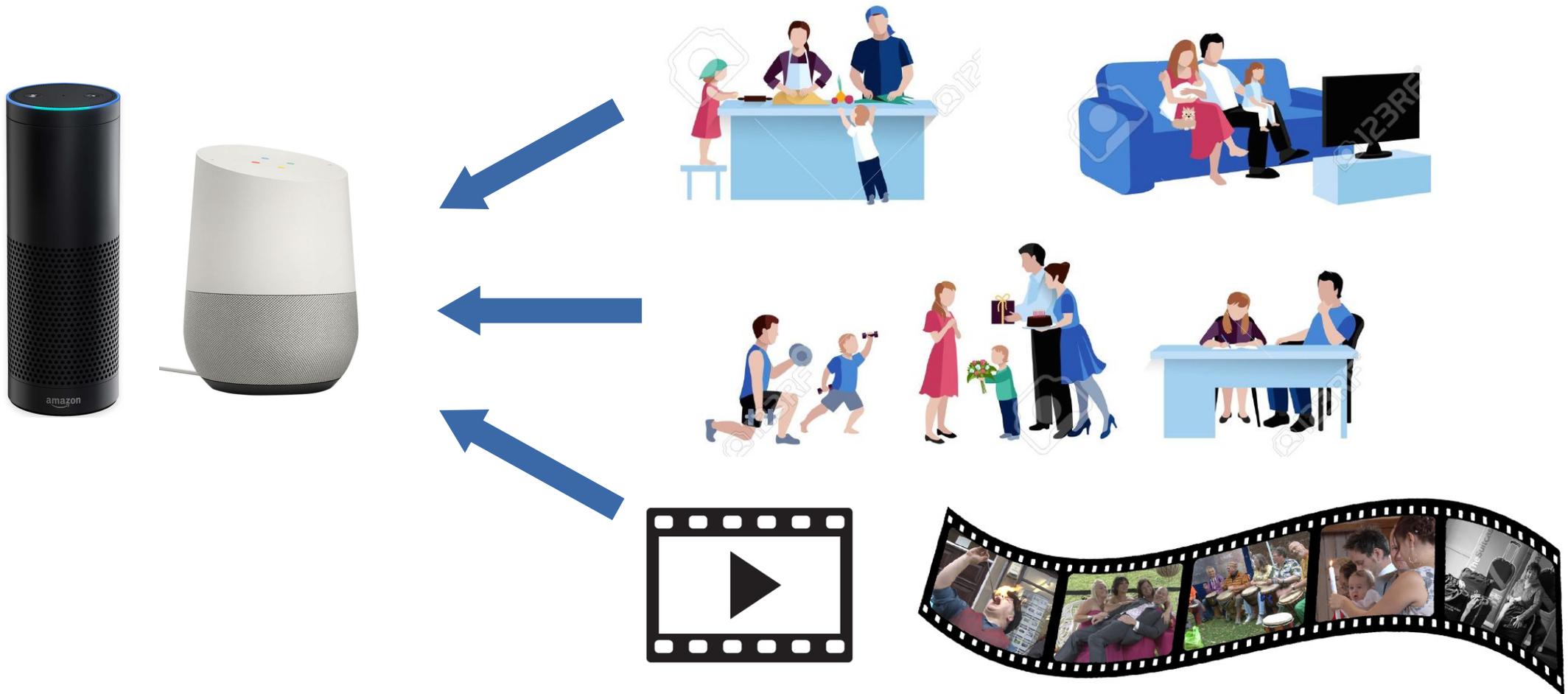
A: i have really choppy streaming video in 9.10 ,
any way to fix that ?
A: any way to fix choppy streaming video in 9.10 ?
R: what kind of video
R: what video card do you **have** ?

(b) *Ubuntu Troubleshoot*

Video-Grounded Dialog Models



- Dialog assistant should be able to “see” daily activities around it and hold dialog conditioned on that context, both for understanding and generating responses.

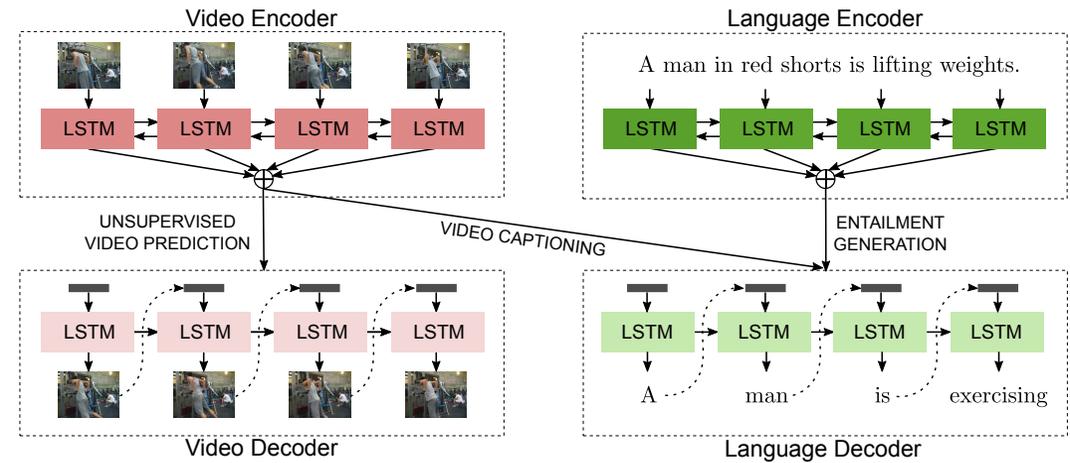


SotA Video Captioning with MTL/RL + Entailment



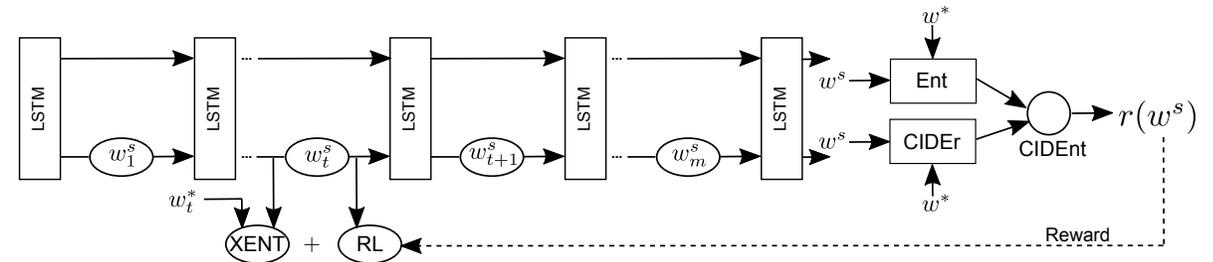
1) Multi-Task Learning with Video +Entailment Generation Task

[Pasunuru and Bansal, ACL 2017]
(*Outstanding Paper Award*)



2) Policy Gradient with Entailment-Corrected CIDEr as Reward

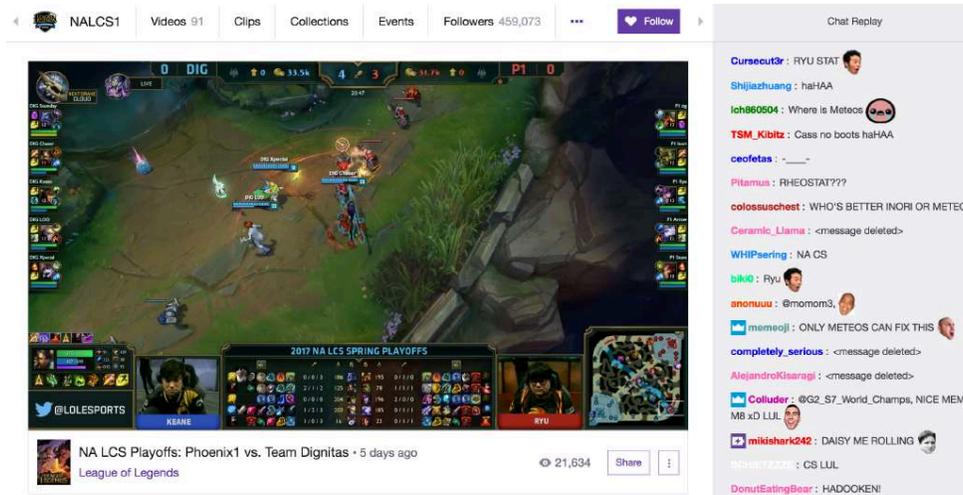
[Pasunuru and Bansal, EMNLP 2017]



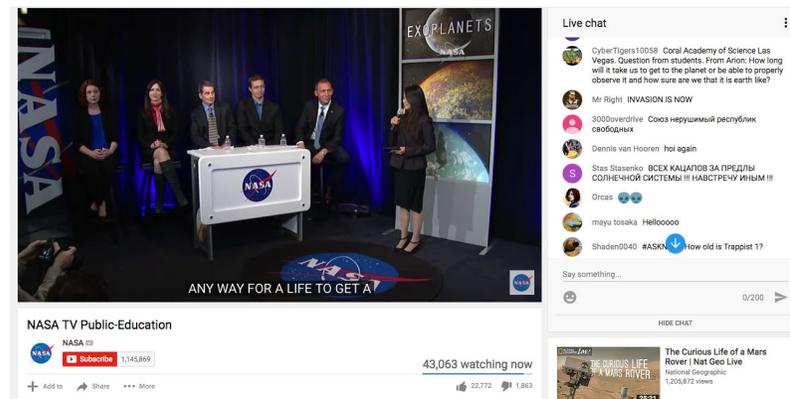
New Video+Dialog Task/Dataset (EMNLP 2017)



- Video + Chat: conversations grounded/task-based in concrete video events!



(a) Twitch



(b) Youtube



(c) Facebook

New Video+Dialog Task/Dataset (EMNLP 2017)



- Very interesting chat language!
 - Time-constrained, not just space
 - Lots of special vocab, symbols, emoticons
 - Multi-user with several interleaving turns
 - Multi-lingual

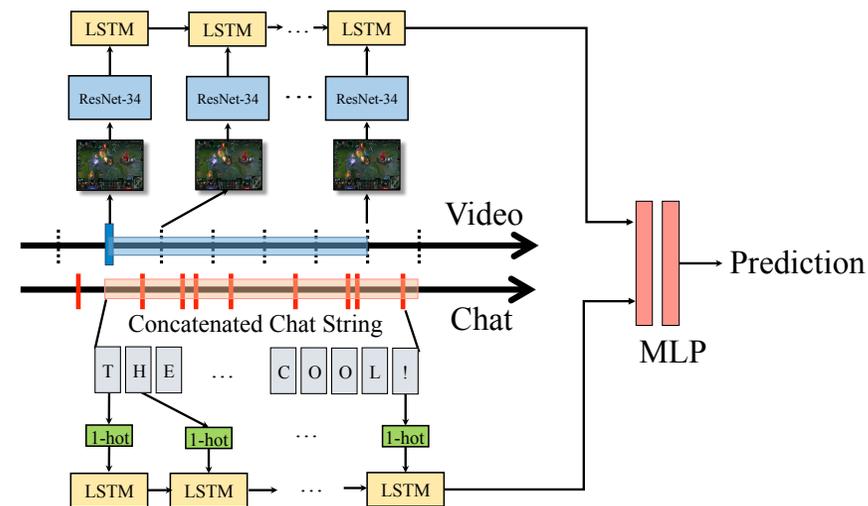
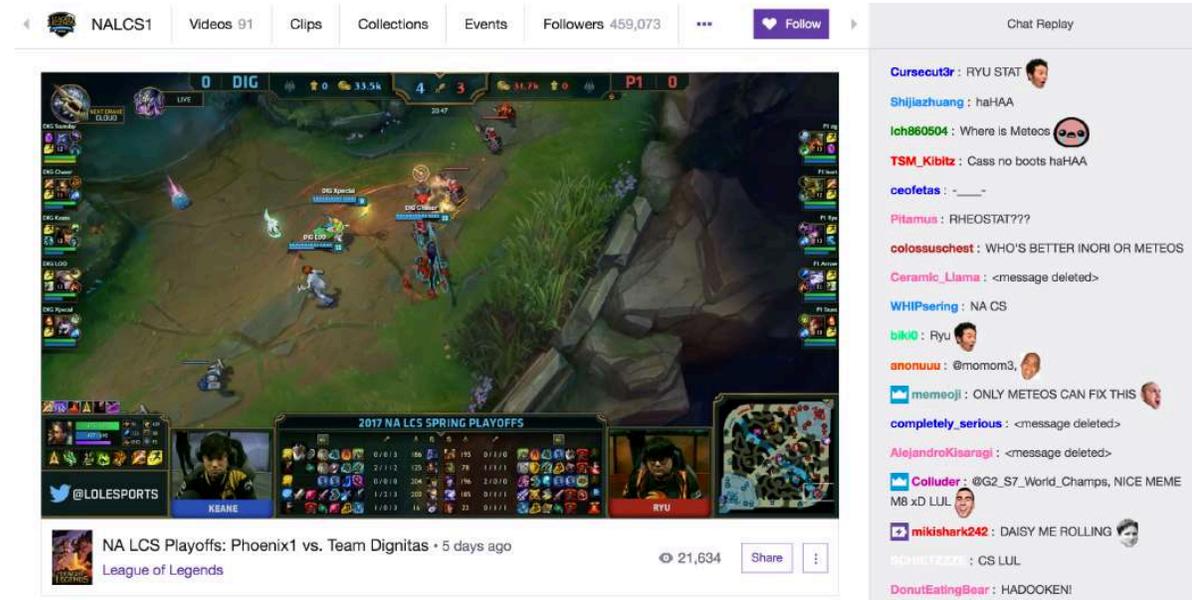
The screenshot shows a Twitch chat replay for a League of Legends match. The main content is a game view showing Phoenix1 vs. Team Dignitas. The chat on the right contains the following messages:

- Cursecut3r: RYU STAT
- Shijiazhuang: haHAA
- Ich860504: Where is Meteos
- TSM_Kibitz: Cass no boots haHAA
- ceofetas: -_-
- Pitamus: RHEOSTAT???
- colossuschest: WHO'S BETTER INORI OR METEOS
- Ceramic_Llama: <message deleted>
- WHIPsering: NA CS
- bikio: Ryu
- anonuuu: @momom3
- memeoji: ONLY METEOS CAN FIX THIS
- completely_serious: <message deleted>
- AlejandroKisaragi: <message deleted>
- Colluder: @G2_S7_World_Champs, NICE MEME M8 xD LUL
- mikishark242: DAISY ME ROLLING
- SCHMETZKE: CS LUL
- DonutEatingBear: HADOOKEN!

New Video+Dialog Task/Dataset (EMNLP 2017)



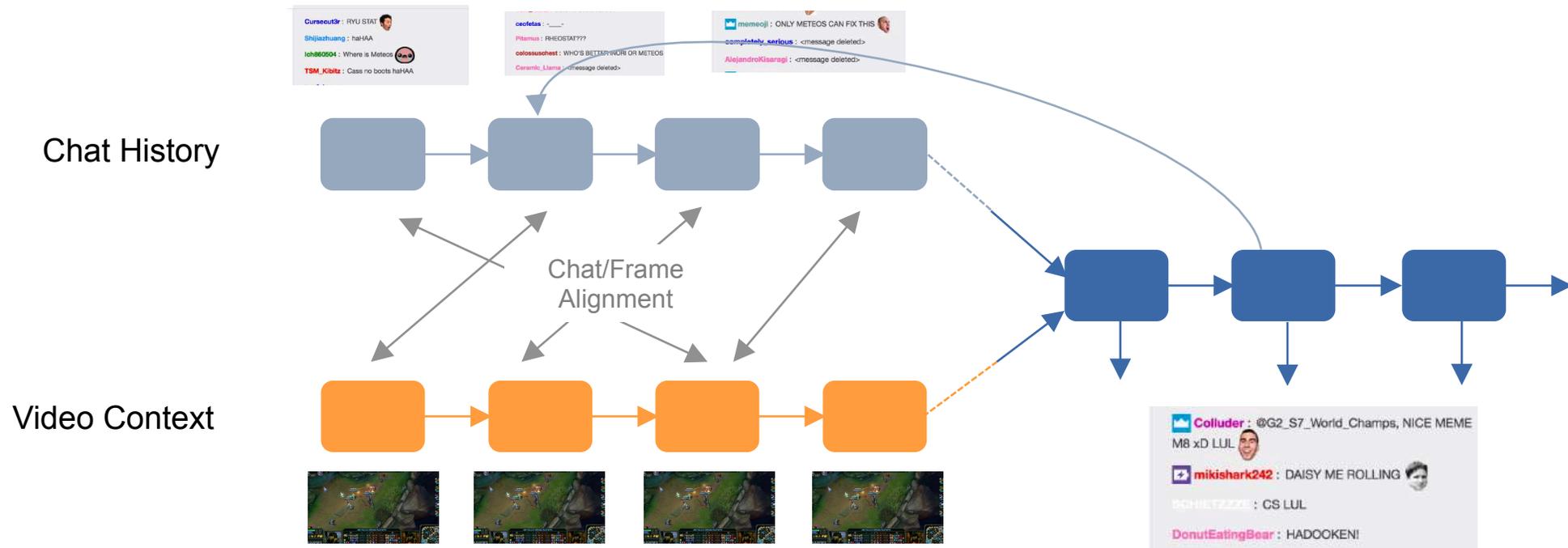
- Very interesting chat language!
 - Time-constrained, not just space
 - Lots of special vocab, symbols, emoticons
 - Multi-user with several interleaving turns
 - Multi-lingual
- We predict highlight frames of full video using joint features from video and user reactions from chat dialog (via a character-level model to capture this new language)



Dialog on Video Context



- New Work: Predicting chat responses given the video and previous dialog history!



Dialog on Video Context



S1: what an offside trap
OMEGALUL

S2: Lol that finish bro

S3: suprised you didn't
do the extra pass

S4: @S10 a drunk bet?

S5: @S11 thanks mate

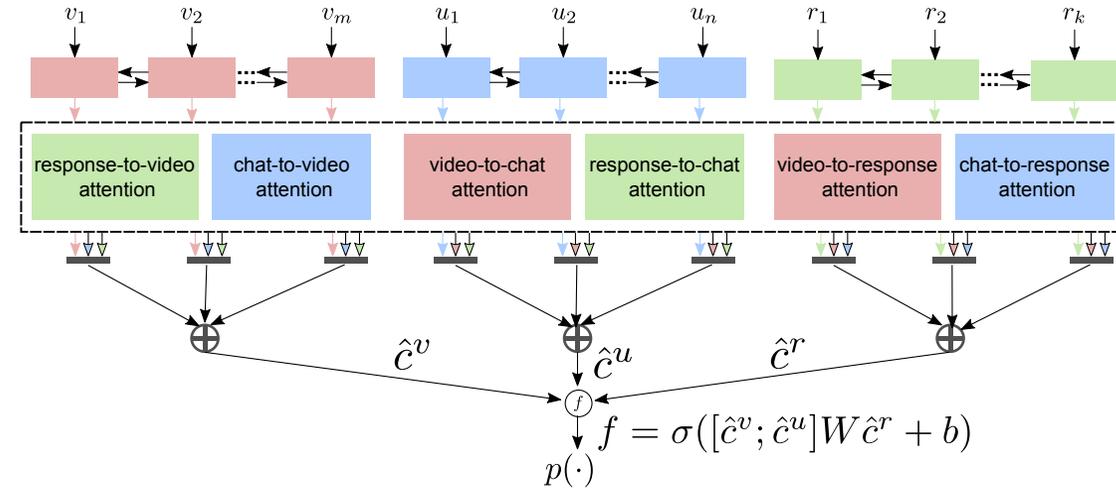
S6: could have passed
one more

S7: Pass that

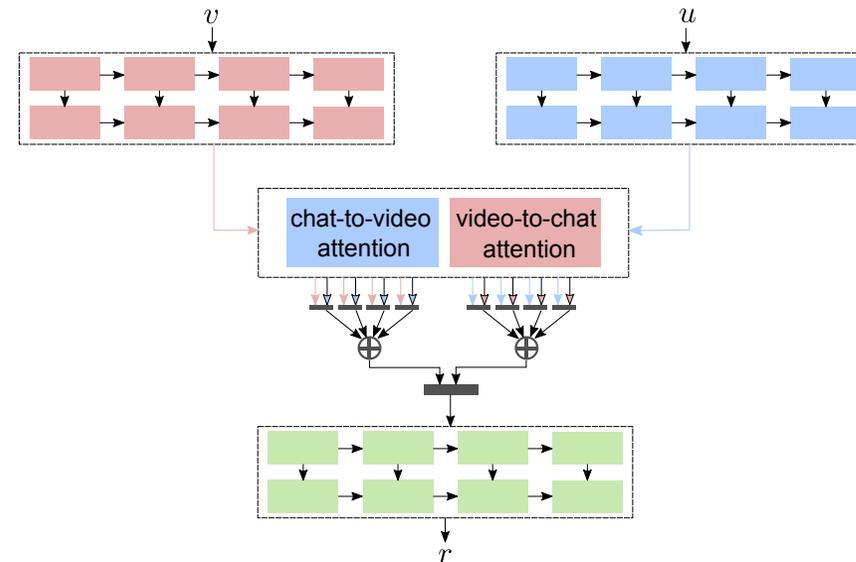
S1: record now!

S8: !record

S9: done a nother pass there



Models	r@1	r@2	r@5
BASELINES			
Most-Frequent-Response	10.0	16.0	20.9
Naive Bayes	9.6	20.9	51.5
Logistic Regression	10.8	21.8	52.5
Nearest Neighbor	11.4	22.6	53.2
Chat-Response-Cosine	11.4	22.0	53.2
DISCRIMINATIVE MODEL			
Dual Encoder (C)	17.1	30.3	61.9
Dual Encoder (V)	16.3	30.5	61.1
Triple Encoder (C+V)	18.1	33.6	68.5
TriDAF+Self Attn (C+V)	20.7	35.3	69.4
GENERATIVE MODEL			
Seq2seq +Attn (C)	14.8	27.3	56.6
Seq2seq +Attn (V)	14.8	27.2	56.7
Seq2seq + Attn (C+V)	15.7	28.0	57.0
Seq2seq + Attn + BiDAF (C+V)	16.5	28.5	57.7

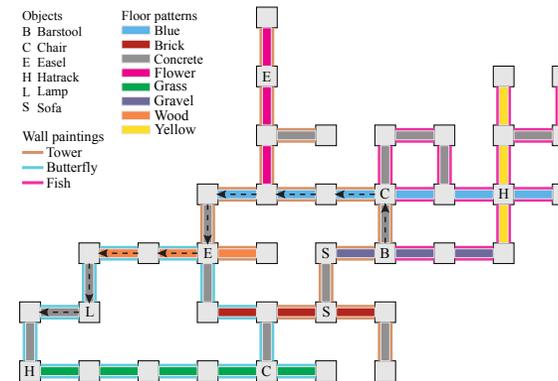
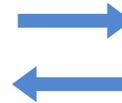


Dialog for Robotic Action Tasks

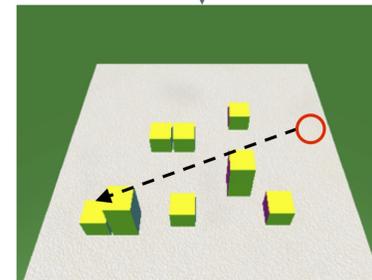
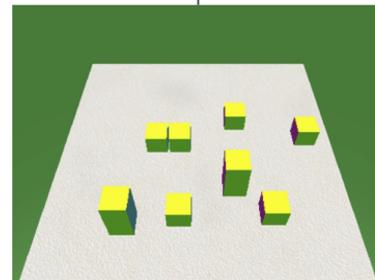


- Both understanding/executing and generating instructions for navigation and assembling/arrangement tasks, for joint human-robot collaboration/task-solving.

Place your back against the wall of the “T” intersection. Go forward one segment to the intersection with the blue-tiled hall. This intersection [sic] contains a chair. Turn left. Go forward to the end of the hall. Turn left. Go forward one segment to the intersection with the wooden-floored hall. This intersection contains [sic] an easel. Turn right. Go forward two segments to the end of the hall. Turn left. Go forward one segment to the intersection containing the lamp. Turn right. Go forward one segment to the empty corner.



Move the block closest to the right table edge so it is to the left of the stack near the front left table corner.



Human-Personality Response Generation

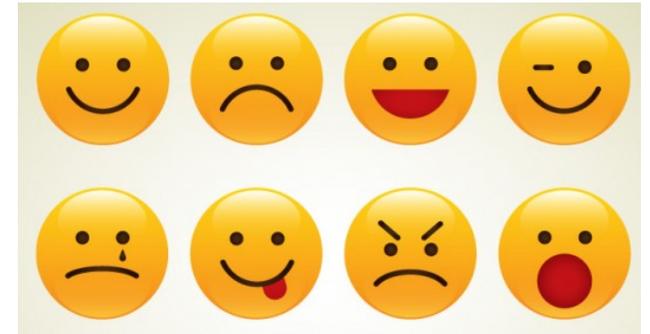


Emotions and Style: Polite,
Happy vs Sad/Sympathetic

Witty, Humorous,
Sarcastic

Classify User's Emotion &
Respond w/ Approp. Emotion

Convincing, Trustworthy,
Effective, Engaging



Polite Dialog Response Generation



Strategy	Politeness	In top quartile	Example
1. Gratitude	0.87***	78%***	I really appreciate that you've done them.
2. Deference	0.78***	70%***	Nice work so far on your rewrite.
3. Greeting	0.43***	45%***	Hey , I just tried to ...
4. Positive lexicon	0.12***	32%***	Wow! / This is a great way to deal...
5. Negative lexicon	-0.13***	22%**	If you're going to accuse me ...
6. Apologizing	0.36***	53%***	Sorry to bother you ...
7. Please	0.49***	57%***	Could you please say more...
8. Please start	-0.30*	22%	Please do not remove warnings ...
9. Indirect (btw)	0.63***	58%**	By the way , where did you find ...
10. Direct question	-0.27***	15%***	What is your native language?
11. Direct start	-0.43***	9%***	So can you retrieve it or not?
12. Counterfactual modal	0.47***	52%***	Could/Would you ...
13. Indicative modal	0.09	27%	Can/Will you ...
14. 1st person start	0.12***	29%**	I have just put the article ...
15. 1st person pl.	0.08*	27%	Could we find a less complex name ...
16. 1st person	0.08***	28%***	It is my view that ...
17. 2nd person	0.05***	30%***	But what's the good source you have in mind?
18. 2nd person start	-0.30***	17%**	You 've reverted yourself ...
19. Hedges	0.14***	28%	I suggest we start with ...
20. Factuality	-0.38***	13%***	In fact you did link, ...

Polite Dialog Response Generation



- Our CNN and LSTM-CNN models perform significantly better than featurized SVM-style models
- We used activation clustering techniques to re-discover and extend existing politeness strategies (features) from psycholinguists:
 - Deference
 - Direct Question
 - Gratitude
 - Counterfactual Modal
- We even discover novel politeness strategies:
 - Indefinite pronouns
 - Emotives / punctuation

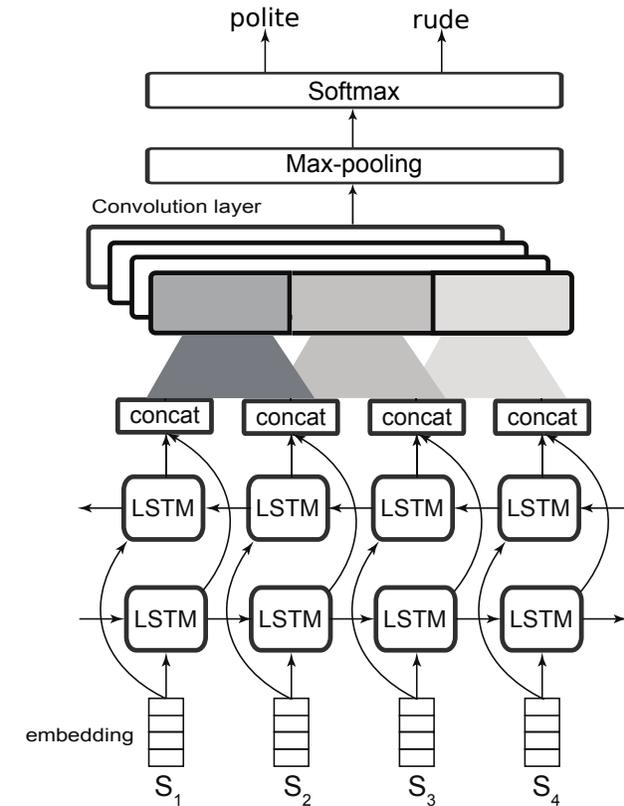


Figure 1: Our LSTM-CNN politeness classifier.

Polite Dialog Response Generation



- 3 models with increasing amount of control/knobs on generation style

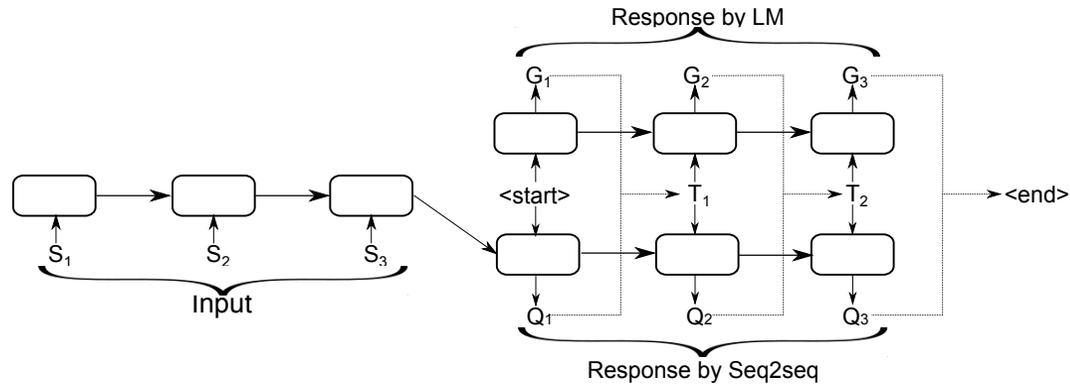


Figure 2: Fusion model: the output probability distributions of the decoder and the polite-LM are linearly mixed to generate the final decoded outputs.

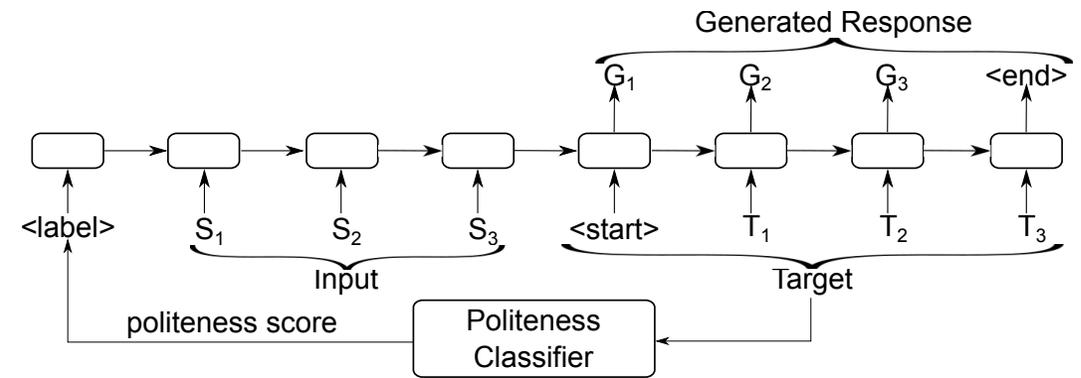


Figure 3: Label-Fine-Tuning model: during training, the embedding of the prepended label is scaled by the style classifier's continuous score on the ground-truth (target) sequence. During testing, we scale the embedding of the label by the desired (continuous) politeness score of the generated response.

Polite Dialog Response Generation



- 3 models with increasing amount of control/knobs on generation style

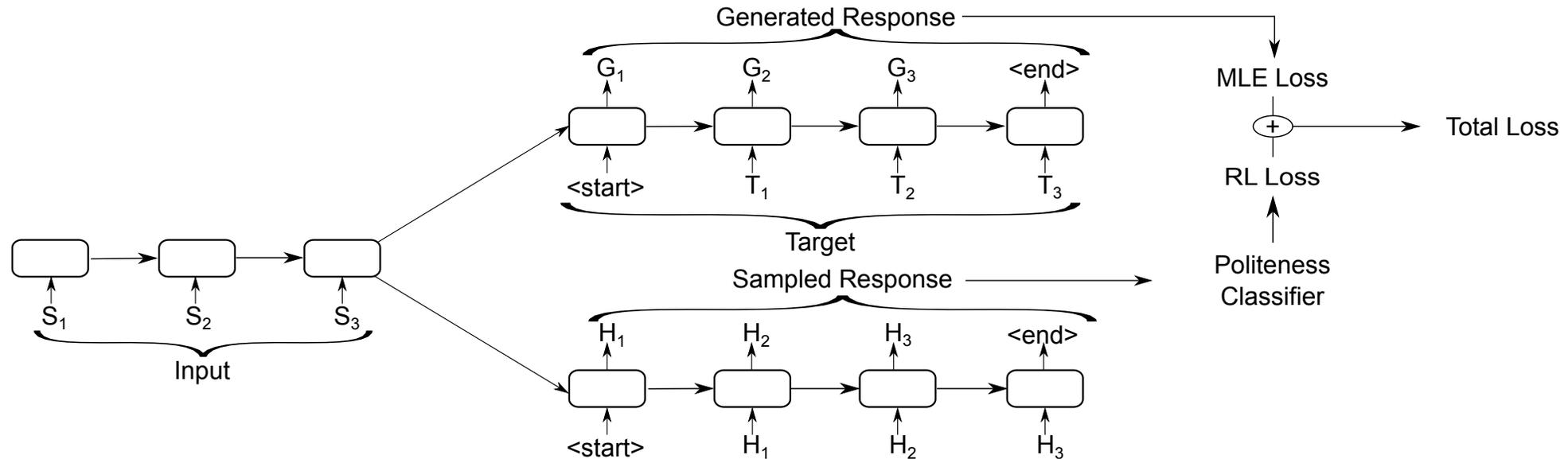


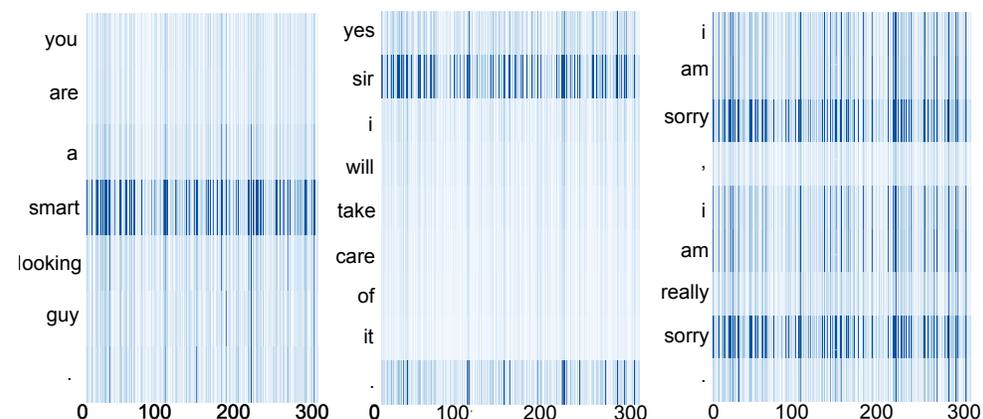
Figure 4: Polite-RL model: upper-right shows max-likelihood (ML) training with generated and ground-truth target sequences; lower-right shows RL training with a randomly sampled response generated by the model and the reward it generates after getting fed into the style classifier. Note that the attention mechanism is not shown here for clarity.

Polite Dialog Response Generation



Target Sequence	Score
Polite Examples	
well , thanks . that 's . i appreciate that .	0.99
<num> , <num> of them in los angeles . i checked . nice work , though .	0.98
nah . i have curfew . he starts to walk away , then stops . quincy oh , by the way . congratulations .	0.97
thank you , ma'am . um , may i ask what this is regarding ?	0.91
hi , <person> . how are you ?	0.89
i know . amazing .	0.89
Rude Examples	
you really should pay more attention to what you read , <person> .	0.11
they were in a car accident .	0.10
you calling more of your stupid friends again ? ya prick .	0.05
are you a bad boy ?	0.05
then she 's a liar .	0.04
excuse me , does that flask belong to this man ?	0.03
oh , well , excuse me all to hell .	0.02

	Politeness	Quality	Difference
Retrieval	3.57	3.15	0.42
Generic-10	3.66	2.99	0.67
Seq2seq	3.11	3.42	0.31
Fusion	3.23	3.05	0.18
LFT	3.63	3.39	0.24
Polite-RL	3.50	3.54	0.04



Polite Dialog Response Generation



Input	X: for what ? Y: for forcing me to face my life honestly for the first time .
Retrieval	X: what 's the c . w . for ?
Generic-10	X: thanks for the answer .
S2S	X: that 's not true .
Fusion	X: i do n't know . I mean , it 's just a feeling , but that 's not what i 'm saying .
LFT	X: you 'll be fine .
RL	X: i wish i could .

Input	X: no, but i wanted to. Y: are you a total lying , scumsucking pig , <person> ?
Retrieval	X: so <person> , what did it total out at ?
Generic-10	X: no problem .
S2S	X: oh , that 's ridiculous .
Fusion	X: well , you 're not a little funny .
LFT	X: i do n't think so , sir .
RL	X: you 're hurting me !

Input	X: you 're sweet to say so . Y: pretty song .
Retrieval	X: is there a song you do n't know ?
Generic-10	X: can you clarify ?
S2S	X: <person> , i know .
Fusion	X: you 're a good man .
LFT	X: i love this song .
RL	X: you sound like a goddess .

Input	X: i find it hard to believe myself . Y: his lips , unused as they are to such a contortion , twisted and stretched . into a smile .
Retrieval	X: thank you . and always smile .
Generic-10	X: nice .
S2S	X: how can you be so sure ?
Fusion	X: you 're a good man , mr .
LFT	X: your lips are well .
RL	X: your lips are so beautiful .

Knowledgeable and Robust Dialog Models



External commonsense

Logical Entailment, Saliency, Discourse

Robustness to Missing words, Spelling/Grammar Errors, Paraphrases

Sensitivity to Negations/ Antonyms



Adversarial Dialog: User-Error Robustness



- Robustness to real-world noise (e.g., user errors) and subtle but important markers!
- “Should-Not-Change” Over-Sensitivity Strategies:
 - Random Swap
 - Stopword Dropout
 - Data-level Paraphrasing
 - Generative-level Paraphrasing
 - Grammar Errors
- “Should-Change” Over-Stability Strategies:
 - Add Negation
 - Antonym
 - Random Inputs
 - Random Inputs with Preserved Entities
 - Confusing Entity
- Tasks/Datasets: Ubuntu (Activity/Entity F1, Human Eval), CoCoA (Completion Rate)
- Models: VHRED, Reranking-RL, DynoNet

Adversarial Dialog: User-Error Robustness



Strategy Name	N-train + A-test	A-train + A-test	A-train + N-test	N-train + N-test
Normal Input	-	-	-	5.94, 3.52
Random Swap	6.10, 3.42	6.47, 3.64	6.42, 3.74	-
Stopword Dropout	5.49, 3.44	6.23, 3.82	6.29, 3.71	-
Data-Level Para.	5.38, 3.18	6.39, 3.83	6.32, 3.87	-
Generative-Level Para.	4.25, 2.48	5.89, 3.60	6.11, 3.66	-
Grammar Errors	5.60, 3.09	5.93, 3.67	6.05, 3.69	-
All Should-Not-Change	-	-	6.74, 3.97	-
Add Negation	6.06, 3.42	5.01, 3.12	6.07, 3.46	-
Antonym	5.85, 3.56	5.43, 3.43	5.98, 3.56	-

Table 2: Activity and Entity F1 results of adversarial strategies on the **VHRED** model.

	VHRED	Tie	Combined-VHRED
Winning %	28	22	49

Table 5: Human evaluation results on comparison between VHRED and VHRED train on all Should-Not-Change strategies combined.

Adversarial Dialog: User-Error Robustness



Context	Response
<p>N: ... you could save your ubuntu files and reinstall Windows , then install ubuntu as a dual boot option __eou__ __eot__ aight buddy , so how do i get that **unknown** space back__eou__</p> <p>Random Swap: ... you could your save ubuntu and files Windows reinstall , then install ubuntu as dual a option boot __eou__ __eot__ aight buddy , so do how i that get space **unknown** back __eou__</p>	<p>NN: you can use the Live CD , you can install Ubuntu on the same partition as the Windows partition __eou__</p> <p>NA: I am using ubuntu . __eou__</p> <p>AA: you can use Windows XP on the Windows partition , and then install Ubuntu on the same drive __eou__</p>
<p>N: ... ok then the problem is you don't have the nvidia driver __eou__ dpkg -l — grep nvidia __eou__ __eot__ i can see some nvidia thing :P __eou__</p> <p>Stopword Dropout: ... ok then problem is you don't have nvidia driver __eou__ dpkg -l — grep nvidia __eou__ __eot__ i see nvidia thing :P</p>	<p>NN: sudo apt-get install nvidia-glx __eou__</p> <p>NA: ok __eou__</p> <p>AA: you can use the nvidia driver __eou__</p>

Adversarial Dialog: User-Error Robustness



Context	Response
<p>N: ... if using hdmi , the display should be set properly . except perhaps for any overscan settings __eou__ __eot__ which package do i have to install for that ? __eou__</p> <p>Data-Level Para.: ... if using hdmi , the displaying ought to be set correctly . except maybe for any overscan setting __eou__ __eot__ which packets do i have to installs for that ? __eou__</p>	<p>NN: you can use the nvidia-settings tool to install the drivers . __eou__</p> <p>NA: the monitor is the monitor . __eou__</p> <p>AA: you can use the nvidia-settings tool to see if the video card is installed . __eou__</p>
<p>N: at this point driver activated but not in use __eou__ __eot__ did you uninstall the old driver first ? __eou__</p> <p>Generative-Level Para.: driver activated but not in use __eou__ __eot__ did you uninstall the old driver ? __eou__</p>	<p>NN: i installed the driver from nvidia website __eou__</p> <p>NA: ok __eou__</p> <p>AA: i installed the driver from nvidia website __eou__</p>
<p>N: i would go with a thinkpad . they are pretty support and have decent hardware __eou__</p> <p>Grammar Errors: ii would go with a thinkpad . they are quite support and have descent hardwares __eou__</p>	<p>NN: I know , but I don't want to buy a new laptop . __eou__</p> <p>NA: Ubuntu . __eou__</p> <p>AA: I know , but I don't want to buy a new laptop . __eou__</p>

Adversarial Dialog: User-Error Robustness



Context	Response
<p>N: suggest you reinstall grub . assuming no other changes , your root system and data should be present and safe . perhaps the bootloader scrambled somehow ? __eou__</p> <p>Add Negation: suggest you don't reinstall grub . assuming no other changes , your root system and data should be present and safe . perhaps the bootloader scrambled somehow ? __eou__</p>	<p>NN: I have a Windows XP partition , and I want to install Ubuntu on it . __eou__</p> <p>NA: I have a Windows XP partition , and I want to install Ubuntu on it . __eou__</p> <p>AA: I don't know how to do that , but I don't know how to do that . __eou__</p>
<p>N: 100% of your RAM . You have it to be used . __eou__ __eot__ what what ? __eou__</p> <p>Antonym: 100% of your RAM . You refuse it to be used . __eou__ __eot__ what what ? __eou__</p>	<p>NN: you have a lot of RAM . __eou__</p> <p>NA: you have a lot of RAM . __eou__</p> <p>AA: I don't know . I don't use it much . __eou__</p>

Commonsense in Generative Q&A Reasoning

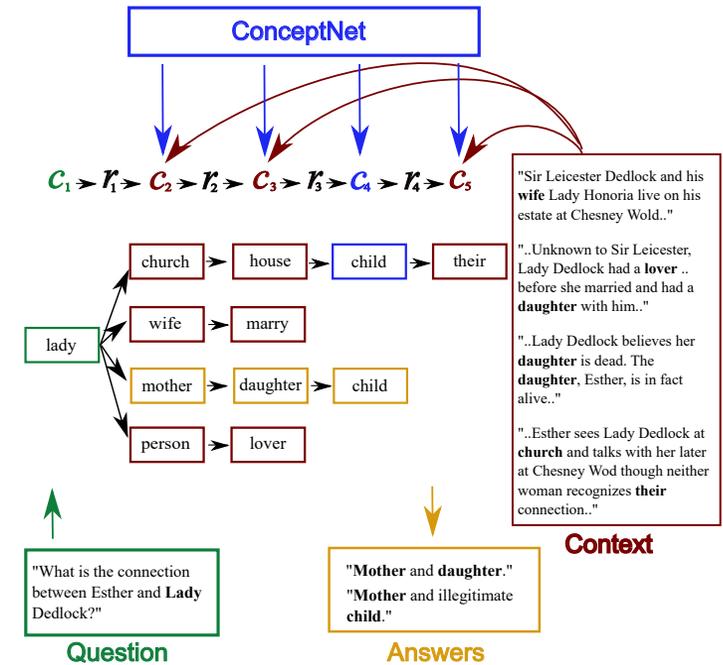
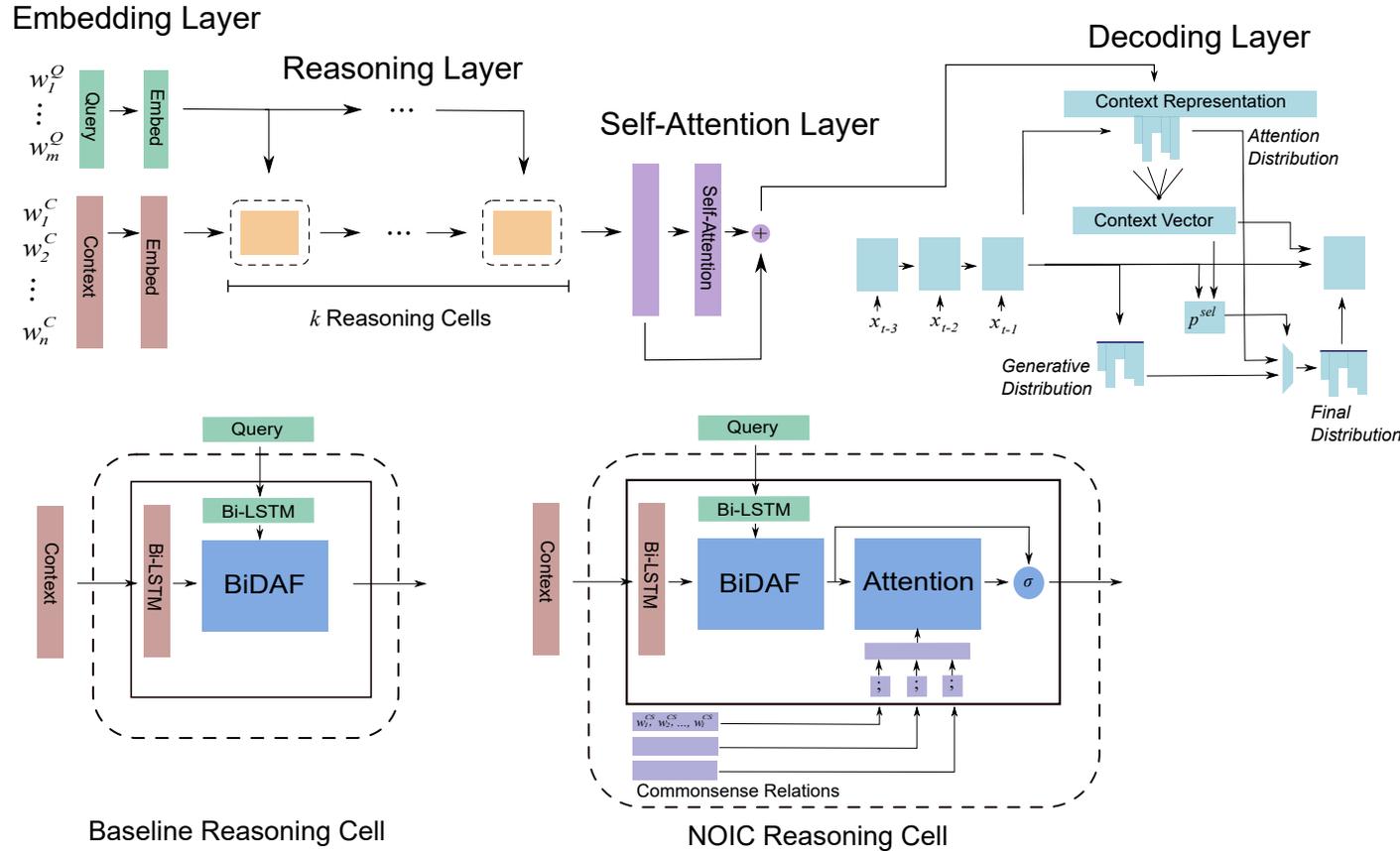


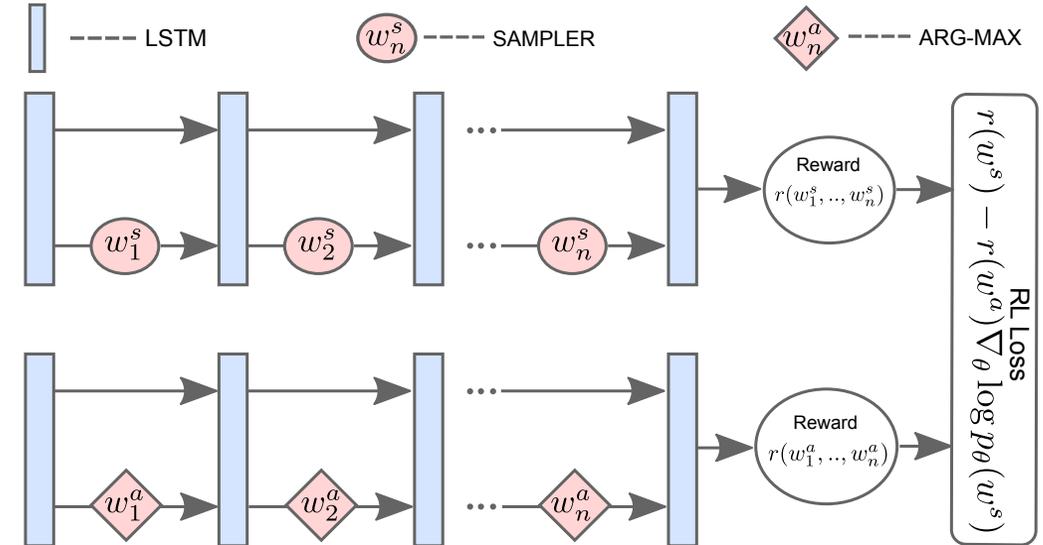
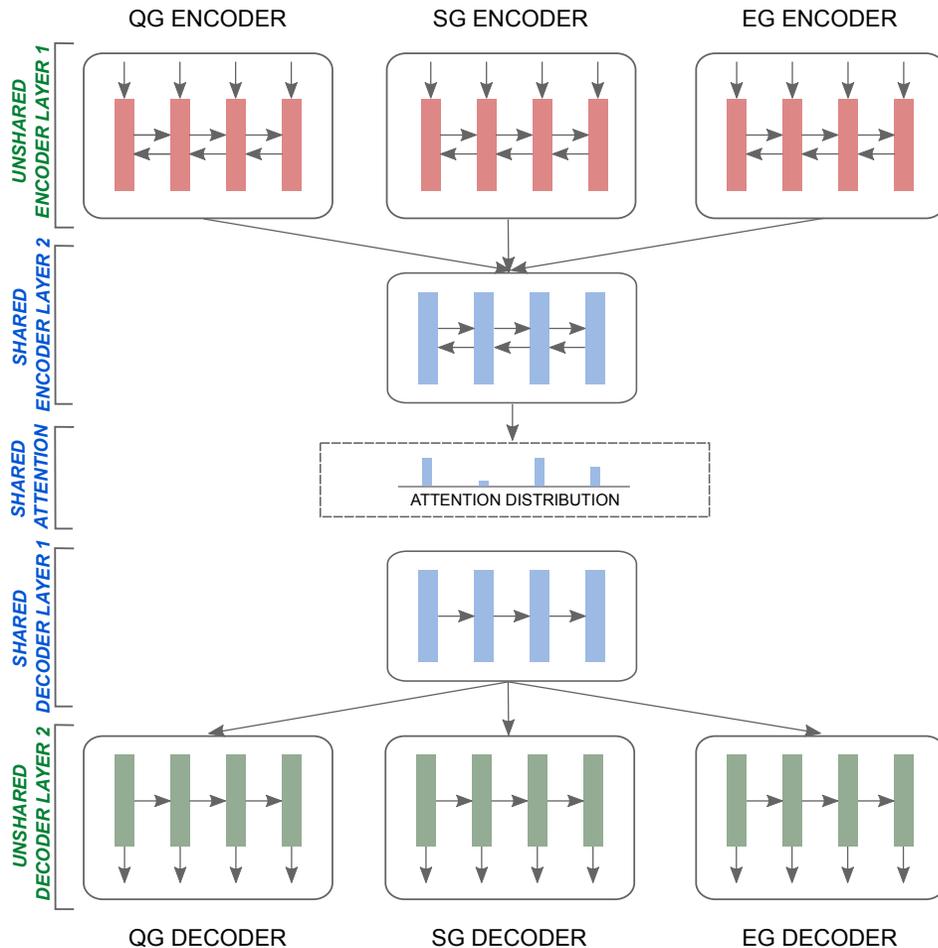
Figure 2: Commonsense selection approach.

Figure 1: Architecture for our Multi-Hop Pointer-Generator Model, and the NOIC commonsense reasoning cell.

Auxiliary Knowledge in Language Generation



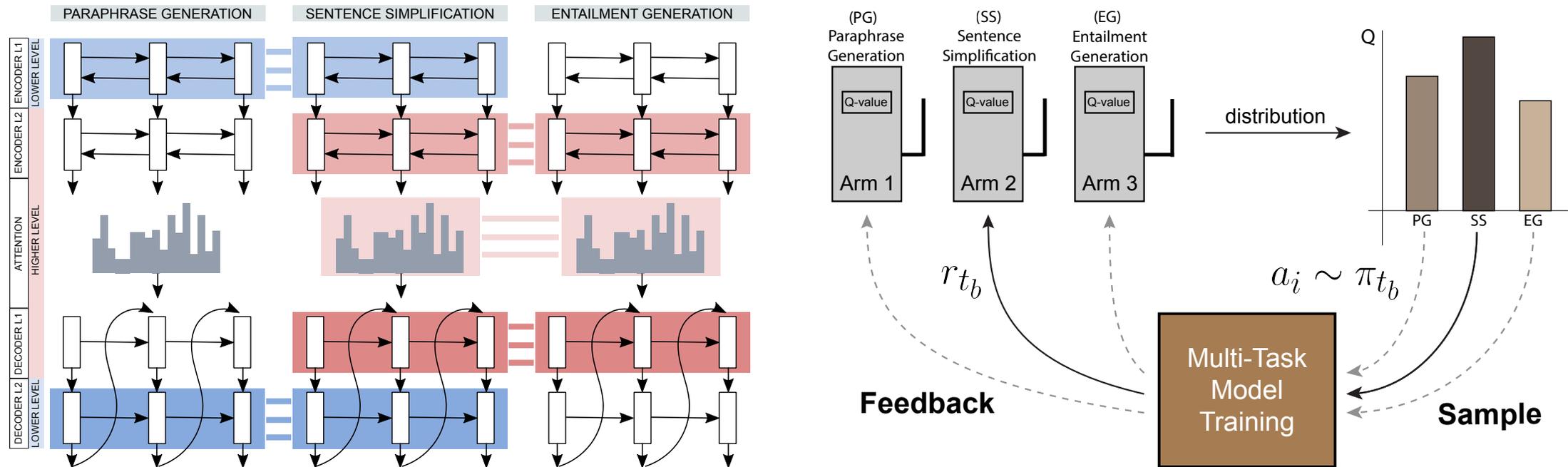
- Multi-Task & Reinforcement Learning with Entailment+Saliency Knowledge for Summarization



Auxiliary Knowledge in Language Generation



- Dynamic-Curriculum Multi-Task Learning with Entailment+Paraphrase Knowledge for Simplification



Students:

PhD Students



Robbie Allen
PhD at UNC



Lisa Bauer
PhD at UNC



Darryl Hannan
PhD at UNC



Hyounghun Kim
PhD at UNC
(co-advised with H. Fuchs)



Jie Lei
PhD at UNC
(co-advised with T. Berg)



Yixin Nie
PhD at UNC



Ramakanth Pasunuru
PhD at UNC



Hao Tan
PhD at UNC



Licheng Yu
PhD at UNC
(co-advised with T. Berg)



Shiyue Zhang
PhD at UNC



Yubo Zhang
PhD at UNC
(co-advised with A. Tropsha)



Xiang Zhou
PhD at UNC

Masters Students



Yen-Chun Chen
MS at UNC



Yichen Jiang
MS at UNC

Undergraduate Students



TSION Coulter
Undergraduate at UNC



Han Guo
Undergraduate at UNC



Sweta Karlekar
Undergraduate at UNC



Yicheng Wang
Undergraduate at UNC

Visiting Students



Nitish Joshi
Undergraduate at IIT Bombay



Tong Niu
MS at Duke

Sponsors:



UNC-NLP Lab: nlp.cs.unc.edu



[HOME](#) [PEOPLE](#) [PUBLICATIONS](#) [SOFTWARE](#) [SPONSORS](#) [TEACHING](#) [COLLOQUIA](#) [JOIN](#) [PHOTOS](#)



Welcome to the UNC-NLP Research Group

Our lab has research interests in statistical natural language processing and machine learning, with a focus on multimodal, grounded, and embodied semantics (i.e., language with vision and speech, for robotics), human-like language generation and Q&A/dialogue, and interpretable and structured deep learning. We are a group of PhD, MS, BS, and visiting students who work with [Prof. Mohit Bansal](#) and collaborators in the [Computer Science department](#) (lab located in [Brooks Building FB-241C](#)) at the [University of North Carolina \(UNC\) Chapel Hill](#).

News

Aug 2018. [7 new papers](#) (6 in EMNLP; 1 in CoNLL).

July 2018. Thanks to Army Research Office for the [ARO Young Investigator Program \(YIP\) Award](#).

July 2018. 1st rank in EMNLP [FEVER \(Fact Extraction & VERIFICATION\) shared task](#) (congrats Yixin and Haonan)!

June 2018. COLING paper on dynamic-MTL selected as "[Area Chair Favorites](#)" (congrats Han+Ram)

Apr 2018. [4 new papers](#) (2 in ACL; 1 in TACL; 1 in WinNLP).

Apr 2018. Congrats to [Lisa Bauer](#) for winning the 3-year [NSF PhD Fellowship!](#)

Mar 2018. Thanks to Adobe for the [Adobe Research Award](#).

Feb 2018. [9 new 2018 papers](#) in NAACL, CVPR, AAAI, WACV.

Sept 2017. Thanks to DARPA for the [DARPA Young Faculty Award \(link\)](#).

Sept 2017. Thanks to Facebook for the [Facebook ParIAI Research Award](#).

July 2017. 3 papers at [EMNLP 2017](#) and 2 papers at the [Summarization-Frontiers](#) and [RepEval](#) workshops.

June 2017. Top single model results on the [RepEval-NLI Shared Task](#) at EMNLP 2017 (congrats Yixin!).

June 2017. [Outstanding Paper Award](#) at ACL 2017 (congrats Ram!).

Feb 2017. Thanks to Google for a [Google Faculty Research Award \(link\)](#).

Nov 2016. 3 papers on [navigational instruction generation](#), [coherent dialogue w/ attn-LMs](#), and on [context-RNN-GAN models](#) to appear at [AAAI 2017](#) and [HRI 2017](#).

July 2016. [5 papers](#) to appear at [EMNLP 2016](#): visual story sorting, visual question relevance, neural network interpretation (for politeness), machine comprehension/Q&A, and character-based sentence embedding models.

July 2016. Best paper award at [ACL 2016 Repl4NLP workshop](#) for [paper](#) on mapping unseen words, and [paper](#) on end-to-end joint neural models for entity+relation extraction to appear at [ACL 2016](#).

Mar 2016. Thanks to Bloomberg for a [Bloomberg Data Science Research Grant \(link\)](#).

Feb 2016. [Paper](#) on universal sentence embeddings selected as an [oral](#) at [ICLR 2016](#).

Jan 2016. Our [work](#) on AI for computational humor was covered in [MIT Technology Review](#) and [Newsweek](#).

Nov 2015. Nvidia paper award at [NIPS 2015 Multimodal ML](#) workshop for [paper](#) on navigational instruction following.

Dec 2014. Thanks for an [IBM Faculty Award](#) and a [Google Faculty Research Award \(link\)](#).

Thank you!

Webpage: <http://www.cs.unc.edu/~mbansal/>

Email: mbansal@cs.unc.edu

UNC-NLP Lab: <http://nlp.cs.unc.edu/>