Carnegie Mellon University
School of Computer Science

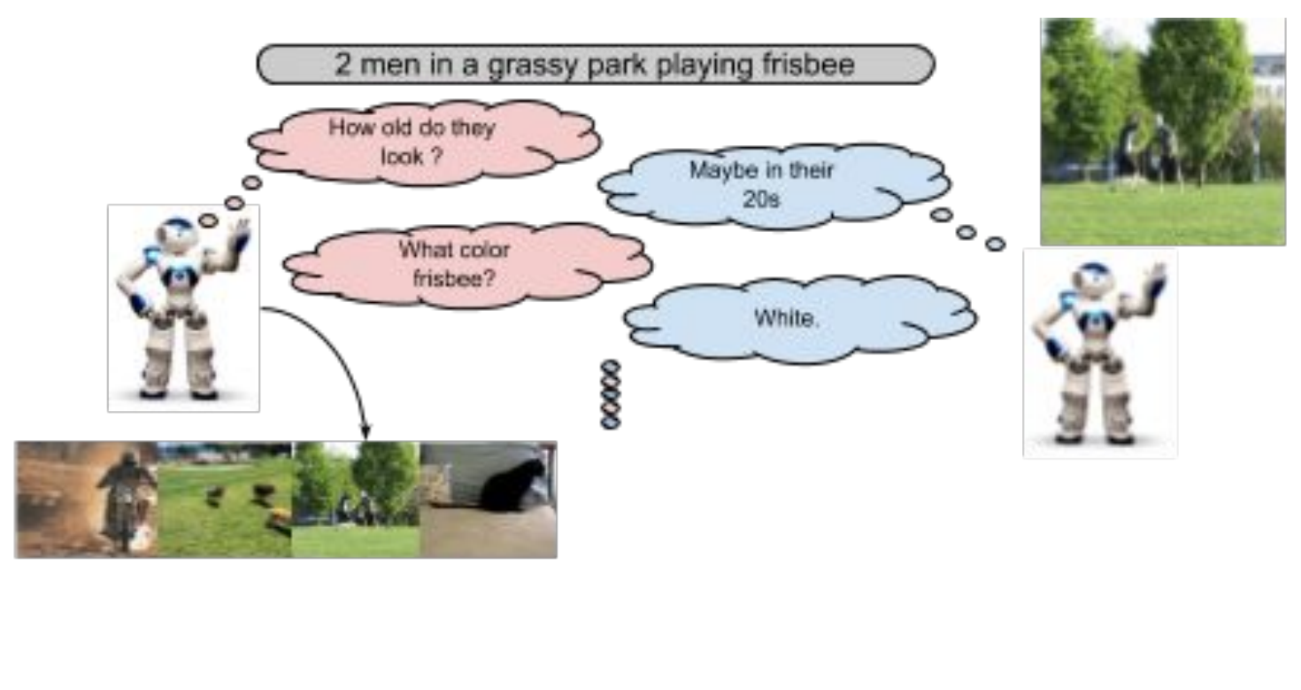
Akshat Agarwal*, Swaminathan G.*, Vasu Sharma*, Katia Sycara

Github Code
https://goo.gl/gc6dGZ

Paper
https://goo.gl/pYWUV6

Mind Your Language
Learning Visually Grounded Dialog in a Multi-Agent Setting

MOTIVATION



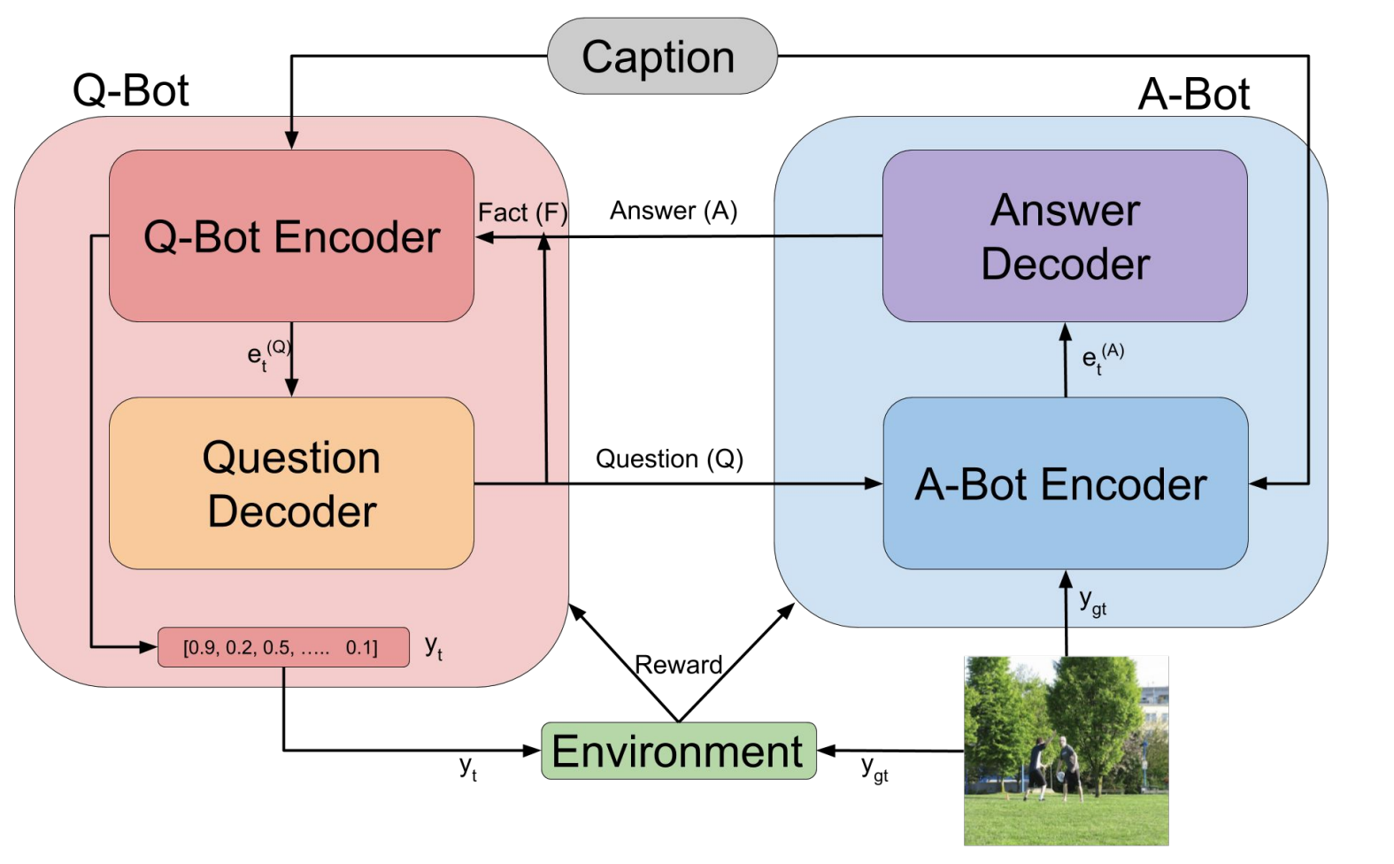
Goal-oriented dialog involves agents conversing with each other to achieve a particular goal, like transferring information

- Interpretability of these conversations desirable for transparency, motivating the use of Natural Language (NL)
- For AI, exchanging info by communicating in NL is inherently suboptimal
- Humans adhere to NL because they have to interact with an entire community, and having a private language for each person would be inefficient
- Hence, we propose a multi-agent dialog framework where each agent interacts with and learns from multiple agents, resulting in coherent and interpretable dialog

VISUAL DIALOG TASK

- Formulated as a conversation between two collaborative agents, a Question (Q-) Bot and an Answer (A-) Bot
- A-Bot given an image and a caption, while Q-Bot is given only a caption - both agents share a common objective, which is for Q-Bot to form an accurate mental representation of the unseen image
- Facilitated by exchange of 10 pairs of questions and answers between the two agents, using a shared common vocabulary

BACKGROUND



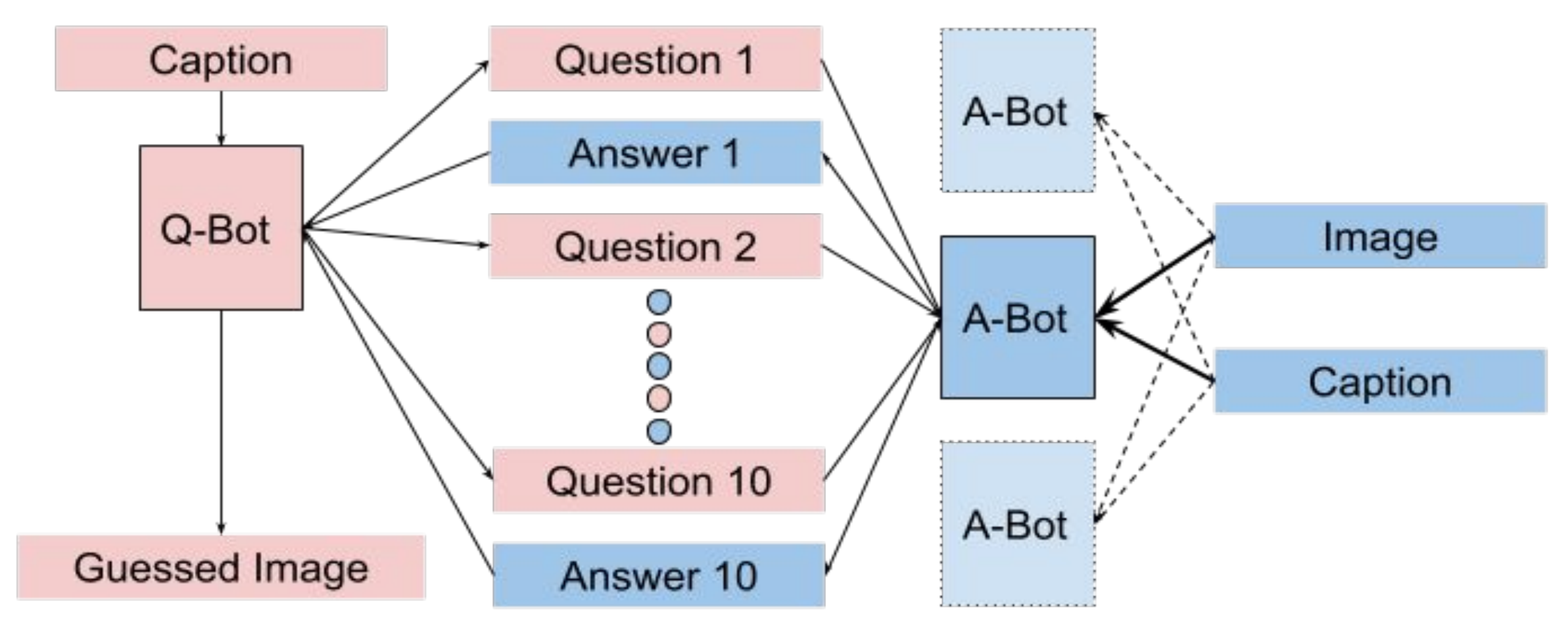
First, agents are trained in isolation via supervision for 15 epochs from VisDial dataset, using Max Likelihood Estimation loss wrt ground truth QA - results in repetitive responses

Then, they are smoothly transitioned to RL via a curriculum

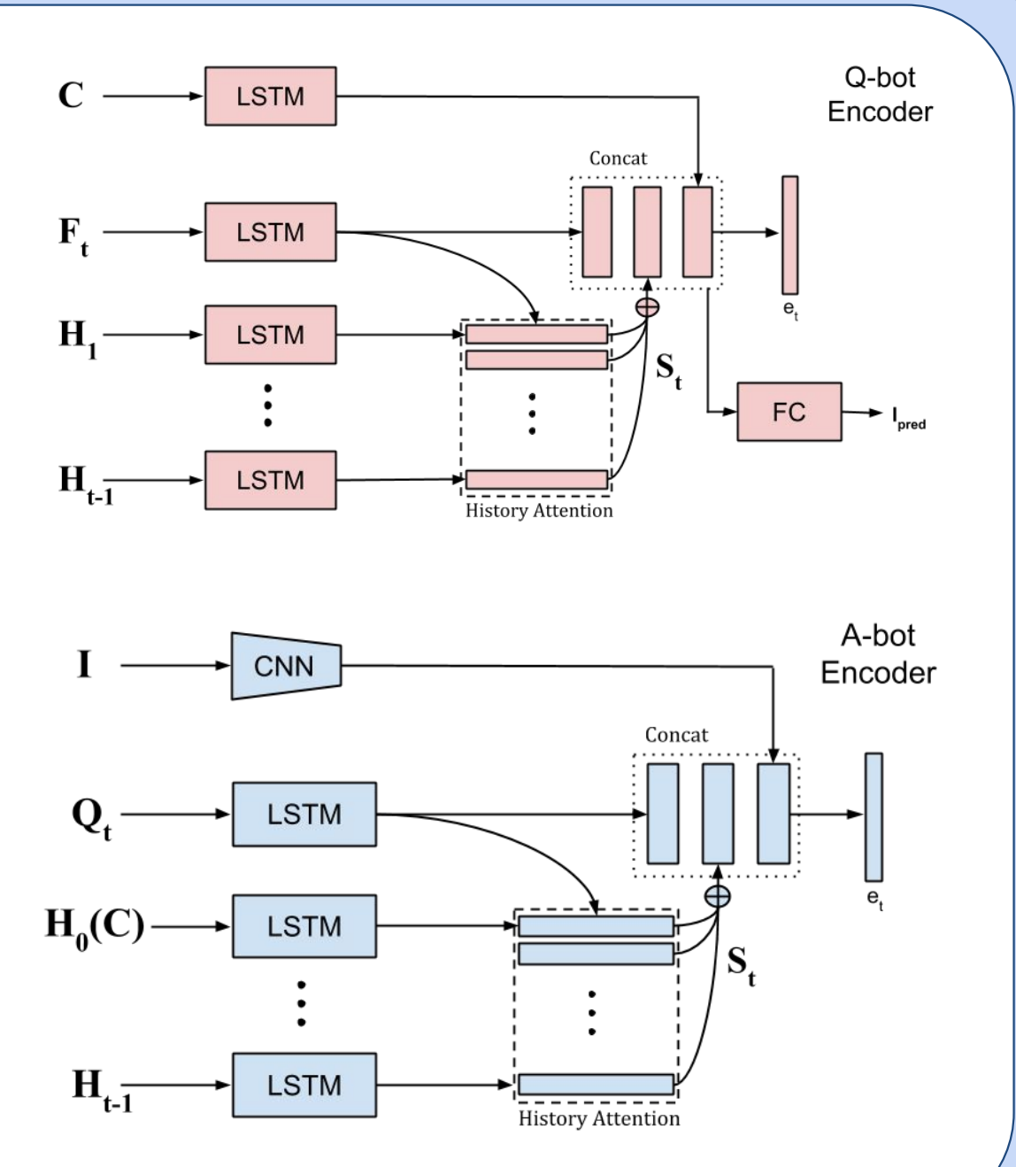
1. Both agents interact with each other and learn by self-play
2. Q-Bot observes $\{c, q_1, a_1, \ldots, q_{10}, a_{10}\}$, A-Bot also observes I
3. Action: Predict words sequentially until a stop token is encountered (or max length reached)
4. Reward: Incentivizing information gain from each round of QA, measured using the predicted image embedding y_t

$$r_t(s_t^Q, (q_t, a_t, y_t)) = l(y_{t-1}, y^{gt}) - l(y_t, y^{gt})$$

5. No motivation to stick to rules and conventions of English language, making the RL optimization problem ill-posed



Multi-Agent Dialog Framework (MADF)



- We solve the problem using our multi-agent setup where 1 Q-Bot communicates with 1 of multiple A-Bots (or vice-versa) for a batch of training, then chooses another A-Bot and repeats
- Much harder for the agents to deviate from natural language since coming up with a new language for each pair would be inefficient

Algorithm 1 Multi-Agent Dialog Framework (MADF)

```
procedure MULTIBOTTRAIN
    while train_iter < max_train_iter do                                   ▷ Main Training loop over batches
        $Qbot \leftarrow random\_select\ (Q_1, Q_2, Q_3....Q_q)$
        $Abot \leftarrow random\_select\ (A_1, A_2, A_3....A_a)$           ▷ Either $q$ or $a$ is equal to 1
        $history \leftarrow (0, 0, ...0)$                                   ▷ History initialized with zeros
        $fact \leftarrow (0, 0, ...0)$                                      ▷ Fact encoding initialized with zeros
        $\Delta image\_pred \leftarrow 0$                                   ▷ Tracks change in Image Embedding
        $Qz_1 \leftarrow Ques\_enc(Qbot, fact, history, caption)$
        for t in 1:10 do                                                   ▷ Have 10 rounds of dialog
            $ques_t \leftarrow Ques\_gen(Qbot, Qz_t)$
            $Az_t \leftarrow Ans\_enc(Abot, fact, history, image, ques_t, caption)$
            $ans_t \leftarrow Ans\_gen(Abot, Az_t)$
            $fact \leftarrow [ques_t, ans_t]$                               ▷ Fact encoder stores the last dialog pair
            $history \leftarrow concat(history, fact)$                     ▷ History stores all previous dialog pairs
            $Qz_t \leftarrow Ques\_enc(Qbot, fact, history, caption)$
            $image\_pred \leftarrow image\_regress(Qbot, fact, history, caption)$
            $R_t \leftarrow (target\_image - image\_pred)^2 - \Delta image\_pred$
            $\Delta image\_pred \leftarrow (target\_image - image\_pred)^2$
        end for
        $\Delta(w_{Qbot}) \leftarrow \frac{1}{10}\sum_{t=1}^{10} \nabla_{\theta_{Qbot}} [G_t \log p(ques_t, \theta_{Qbot}) - \Delta image\_pred]$
        $\Delta(w_{Abot}) \leftarrow \frac{1}{10}\sum_{t=1}^{10} G_t \nabla_{\theta_{Abot}} \log p(ans_t, \theta_{Abot})$
        $w_{Qbot} \leftarrow w_{Qbot} + \Delta(w_{Qbot})$                 ▷ REINFORCE and Image Loss update for Qbot
        $w_{Abot} \leftarrow w_{Abot} + \Delta(w_{Abot})$                 ▷ REINFORCE update for Abot
    end while
end procedure
```

The little girl is standing with skis on her feet

Human-Human

what color are the skis ?	A UNK color
Are there any other people?	Not that i can see
Is this outdoors?	Yes
Do you see snow?	Yes
Is it currently snowing?	No I don't think so
Is she on a slope or hill?	No i don't think so
Do you see trees?	Yes
Do you see the sky?	No
Is she wearing gloves?	Yep
Is she wearing a hat?	yes

Supervised

is the girl close to a race ?	no , woman is riding
'is she wearing a helmet ?	no
is she wearing a hat ?	She is wearing a headband
what color is it ?	I would say it's black
'is she wearing goggles ?	Can't see her face
'is she wearing goggles ?	Not sure
what color is her skis ?	Dark, hard to tell i would guess
is she wearing a hat ?	No
what color does she have ?	White
and what is the girl wearing ?	A white striped long sleeve shirt and blue

REINFORCE (1 Q-Bot - 1 A-bot)

is there more than 1 skier ?	Yes
oh how many UNK do you see ?	A few
is the person holding a pipe of some sort of basket ?	Yes, there is a bowtie
is lady going fast ?	I can not tell
how many is there ?	It's not silver people are in grocery
is there snow ?	Is covered by snow covered by snow
how many animals do you see ?	Of the 1, and a 2
is there a lot of snow ?	Yes tons of snow everywhere covered
are they all together ?	Yes 1 is a darker 1
is there a lift ?	Just a few skiing out yet

REINFORCE (1 Q-Bot - 3 A-bots)

are there any other people around ?	No
is it snowing ?	Can't tell
is it a sunny day ?	Yes
is the woman wearing a dress ?	Yes
what color is it ?	It has a white meat and girl on
is she by herself of snow ?	Just lady, woman and dog
what color is her jacket ?	Orange
does it look like a competition ?	Yes
is it day or night ?	Daytime
is it sunny ?	yes

REINFORCE (3 Q-Bots - 1 A-bot)

how old is the woman ?	About 16
is she in a ski resort ?	I don't think so
can you see the sky ?	No
is she alone ?	Yes
what color are the gloves ?	Silver
is there snow on the ground ?	No snow
is the photo very clear ?	Yes
is she wearing a helmet ?	No
is it snowing ?	it looks like it is stopped
is it sunny ?	No really

RESULTS

	Metric	N	Supervised	RL 1Q,1A	RL 1Q,3A	RL 3Q,1A
1	Q-Bot Relevance	8	2.5	2.75	**2**	2.75
2	Q-Bot Grammar	8	**2.25**	2.875	2.5	2.375
3	A-Bot Relevance	12	2.5	2.583	2.25	**1.67**
4	A-Bot Grammar	12	1.92	3.5	**1.83**	2.25
5	Overall Coherence	20	2.8	3.05	**2.3**	**1.85**

Quantitative Metrics (below) and Human Evaluations (above; lower is better; 20 evaluators). RL 1Q,3A refers to dialog system trained with 1 Q-Bot, 3 A-Bots

Overall **Dialog Coherence** of RL-1Q,3A and 3Q,1A systems ranked much better according to humans

- Multiple A-Bots interacting with Q- Bot improves relevance, and vice versa
- The grammar improves for both bots in both 1Q,3A and 3Q,1A settings
- Having multiple A-Bots to interact with exposes the Q-Bot to diverse dialog, leading to more stable updates with lower bias

Model	MRR	Mean Rank	R@1	R@5	R@10
Answer Prior (Das et al., 2016)	0.3735	26.50	23.55	48.52	53.23
MN-QIH-G (Das et al., 2016)	0.5259	17.06	42.29	62.85	68.88
HCIAE-G-DIS (Lu et al., 2017)	0.547	14.23	44.35	65.28	71.55
Frozen-Q-Multi (Das et al., 2017)	0.437	21.13	N/A	53.67	60.48
CoAtt-GAN (Wu et al., 2017)	0.5578	14.4	**46.10**	**65.69**	71.74
SL(Ours)	**0.610**	**5.323**	34.74	57.67	**72.68**
RL - 1Q,1A(Ours)	0.459	7.097	16.04	54.69	72.34
RL - 1Q,3A(Ours)	0.601	5.495	34.83	57.47	72.48
RL - 3Q,1A(Ours)	0.590	5.56	33.59	57.73	72.61

We outperform all previous architectures in MRR, Mean Rank and R@10: **consistently good responses**

This research was sponsored in part by AFOSR grant FA9550-15-1-0442