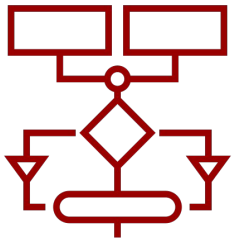
Lecture 5

The challenge with interacting with models

Ranjay Krishna | ranjay@cs.washington.edu

A shift in AI: From algorithms to machine learning



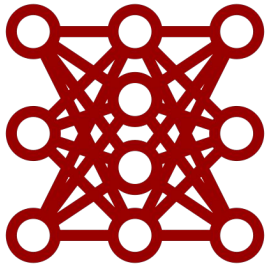
Classical algorithms

Problems: precisely defined algebraically

Example: Graphcut algorithm

Accuracy: measured by correctness

Artifact: provably correct, transparent process



Empirical machine learning

Problems: loosely defined by datasets

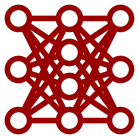
Example: ResNet50 trained on ImageNet 1K

Accuracy: measured using test set

Artifact: stochastic black box model

Ranjay Krishna | ranjay@cs.washington.edu

The last few decades of AI



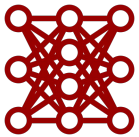
Model-centric AI

Goal: improve accuracy

Output: robust model

Artifact: Training algorithm / model

Current shifts within machine learning



Model-centric AI

Goal: improve accuracy

Output: robust model

Artifact: Training algorithm / model

Data-centric AI

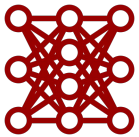
Goal: improving data efficiency

Output: quality data

Artifact: active learning algorithms

Ranjay Krishna | ranjay@cs.washington.edu

The new shift we should move towards.



Model-centric AI

Goal: improve accuracy

Output: robust model

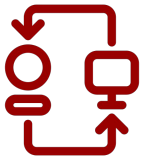
Artifact: Training algorithm / model

Data-centric AI

Goal: improving data efficiency

Output: quality data

Artifact: active learning algorithms



Interaction-centric AI

Goal: success on new goals (sometimes defined by user)

Output: useful model

Artifact: human-AI interaction

Ranjay Krishna | ranjay@cs.washington.edu

Interaction-centric AI

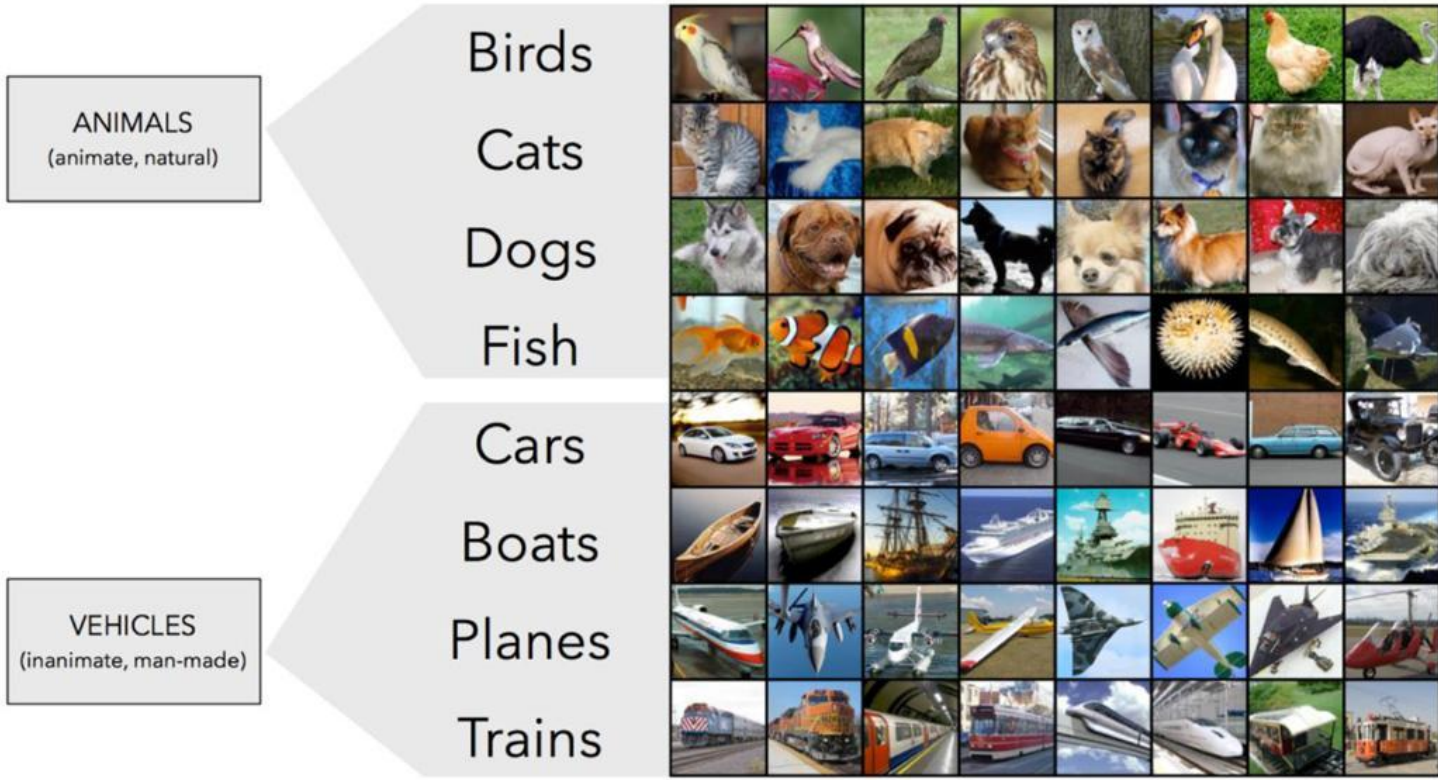
- Placing the human directly within the training cycle
- Expecting to see and operate on inputs never before seen
- Continuously improve model with more interactions

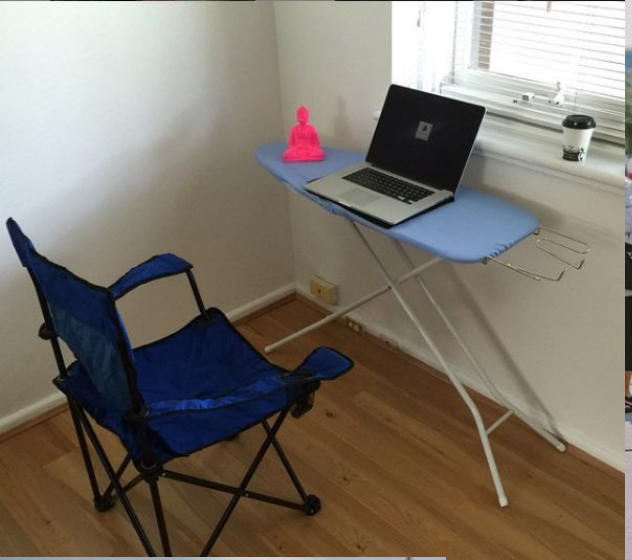
Deep dive into how we can apply this framework in computer vision

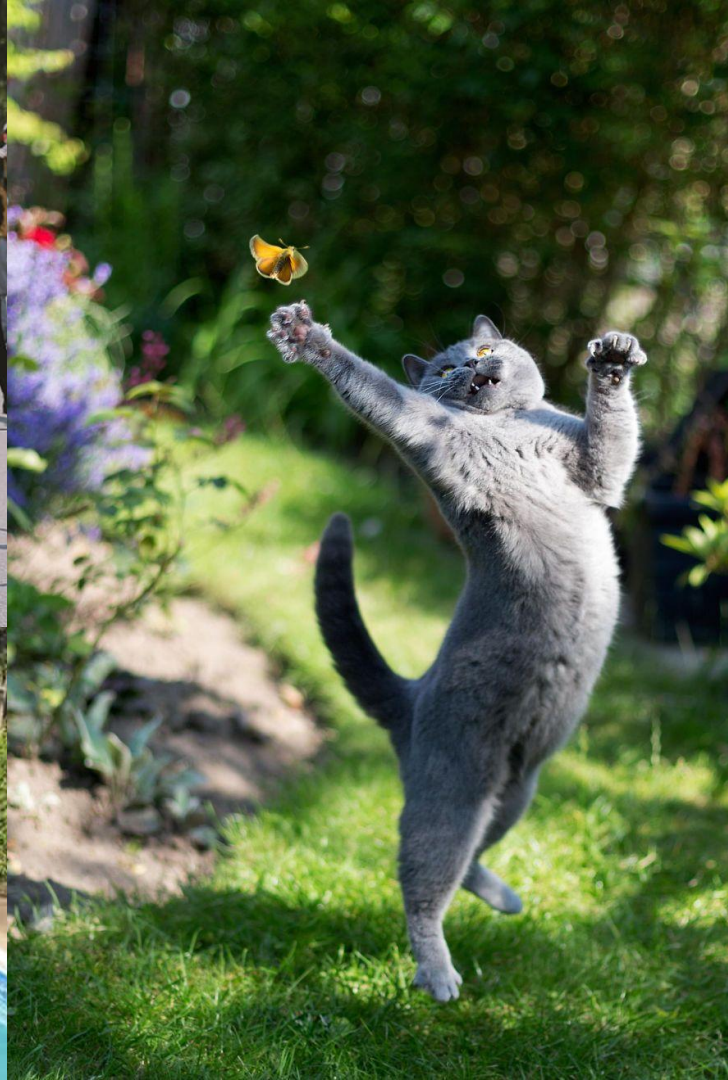
Ranjay Krishna | ranjay@cs.washington.edu

Krishna, Lee, Fei-Fei, Bernstein. Socially Situated Artificial Intelligence. PNAS 2021

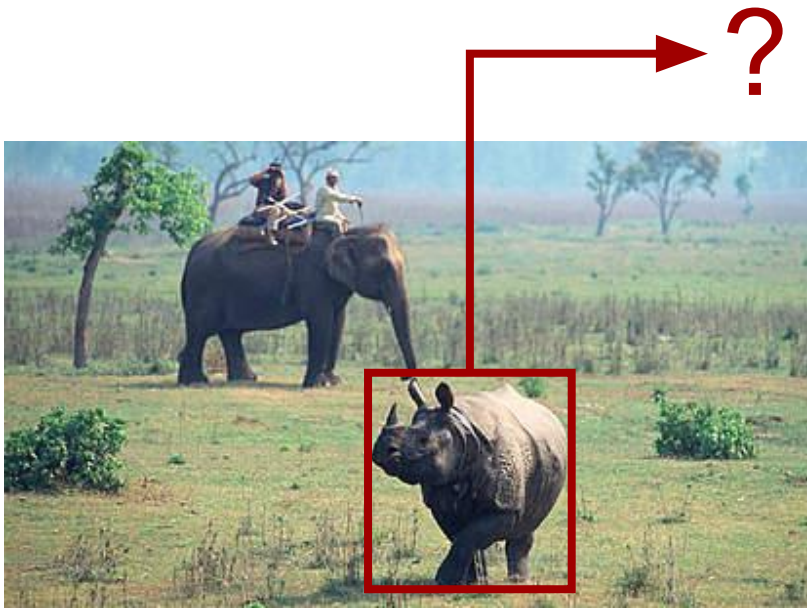
Our world is more than a collection of objects







Models fail when they encounter new concepts



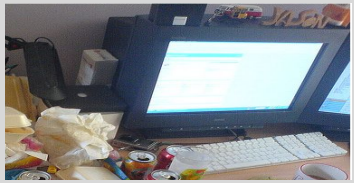
- Deng et al. CVPR 2009; Lin et al. ECCV 2014; Antol et al. CVPR 2015;
- image sources: old and new telephone, old and new computer, airpods, salamander, rhino

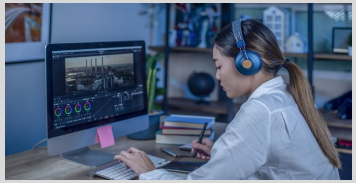
Ranjay Krishna | ranjay@cs.washington.edu

Existing concepts evolve over time

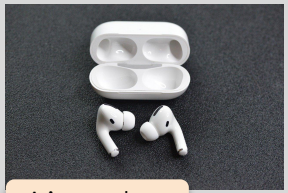
Images from MSCOCO (Lin et al. 2014)

How they look today





New concepts invented or discovered



Airpods

Carolina Salamander

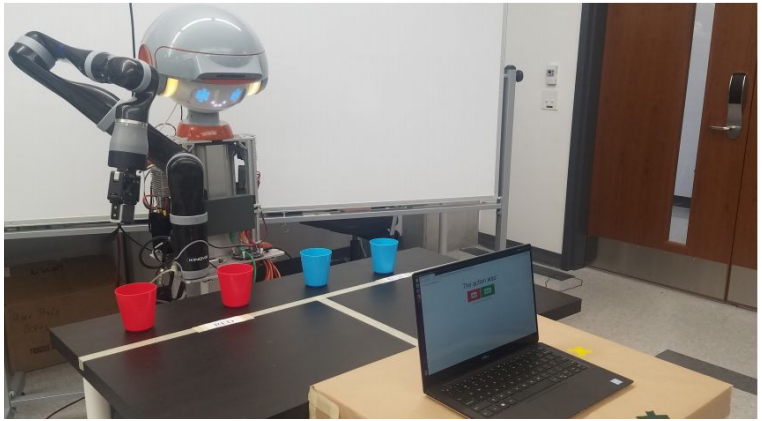
Human development is a socially mediated process

Vygotsky. 1962; Reber. Journal of experimental psychology 1989; Gelman. Annual review of psychology 2009; Breazeal. MIT 2000; Greeno. American psychologist 1998; Darwiche. CACM 2018;

Ranjay Krishna | ranjay@cs.washington.edu

Why is achieving visual intelligence a challenging problem?

Learning from people only possible if people provide explicit feedback

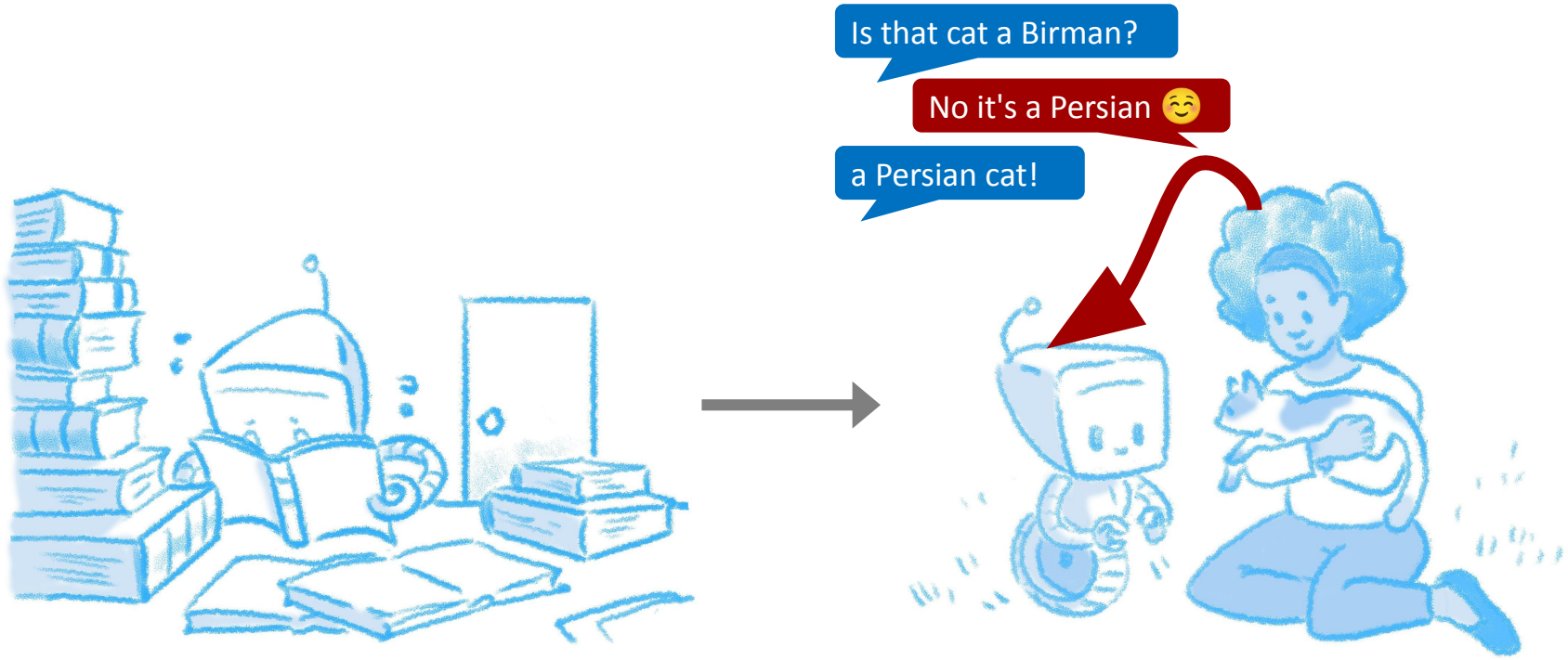


Christiano et al. NeurIPS. 2017; Silver et al. Nature 2016; Misra, et al. CVPR. 2018; Thomaz and Breazeal. AAAI 2006; Thomaz and Breazeal. AI 2008; Cakmak and Thomaz. HRI 2012; Deng et al. CVPR, 2009; Krizhevsky et al. NeurIPS 2012; Devlin et al. NAACL-HLT 2019; Radford et al. ArXiv 2021; Brown et al. ArXiv 2020; Gray. Eamon Donlan Books 2019; Mitchell CACM 2018;

Research question:

How models learn from social interactions with people?

By learning to socially interact with people → models can learn new concepts



Breazeal. MIT 2000; Grudin and Jacques. CHI 2019; Marcus. Pantheon 2019; Winograd. Stanford 1987; Vygotsky. MIT 1962
Drawing credit: Cathy Yuan

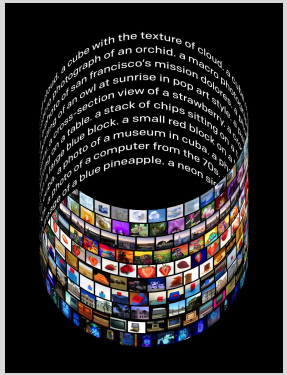
Today's AI agents learn in isolation

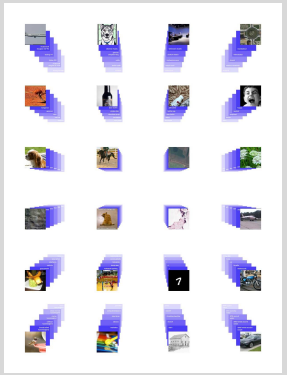
Breazeal. MIT 2000; Grudin and Jacques. CHI 2019; Marcus. Pantheon 2019; Winograd. Stanford 1987; Vygotsky. MIT 1962
Drawing credit: Cathy Yuan

Ranjay Krishna | ranjay@cs.washington.edu

Datasets

Web scaping

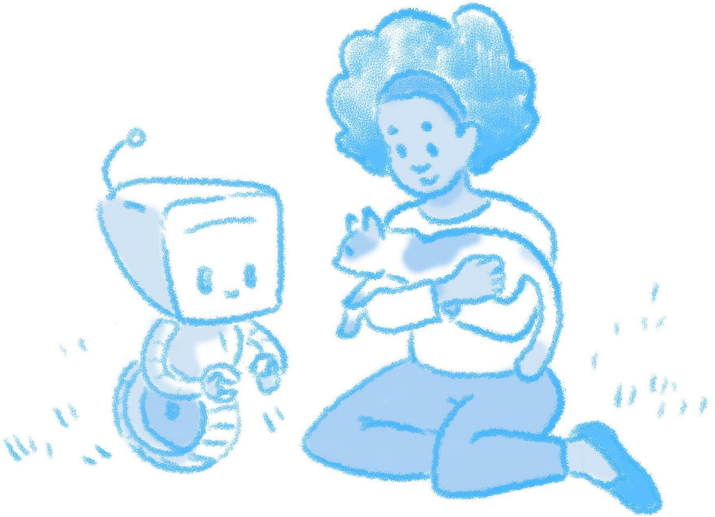




Deng et al. CVPR, 2009; Krizhevsky et al. NeurIPS 2012; Devlin et al. NAACL-HLT 2019; Radford et al. ArXiv 2021; Brown et al. ArXiv 2020; Gray. Eamon Donlan Books 2019; Mitchell CACM 2018;

The walls of the room ossify agents to the evolving world

?



Learning new concepts in social environments

Agent: What is in front of the elephant?

Human: that's a rhinoceros ☺️

Grice. Academic Press 1975

Agents must first learn how to interact with people

Some questions are not informative

Agent: What color is the grass?

Some questions will not receive response

Agent: is the person sitting behind the person guiding the elephant holding their right hand or their left hand up to their eyes?

Ranjay Krishna | ranjay@cs.washington.edu

Agents must tradeoff informative interactions with social interactions

To enable agents that can interact to learn new concepts, they must first learn to interact

Without learning how to interact, the agent will generate questions that people will refuse to respond to

Q: Is that food?
A: ------

Q: Is the color white?
A: ------

Q: How many branches are there?
A: ------

Q: Are there more black balls on top of the table or pink balls under it?
A: ------

Our Socially situated AI framework generates interactions that people want to interact with

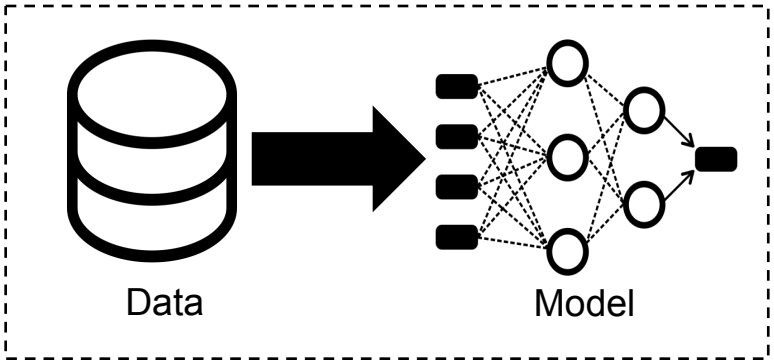
Q; What is the green vegetable?
A: it's bok choy!! So yummy 🥬

Q: What type of dessert is that in the picture?
A: hi dear it's coconut cake, it tastes amazing :)

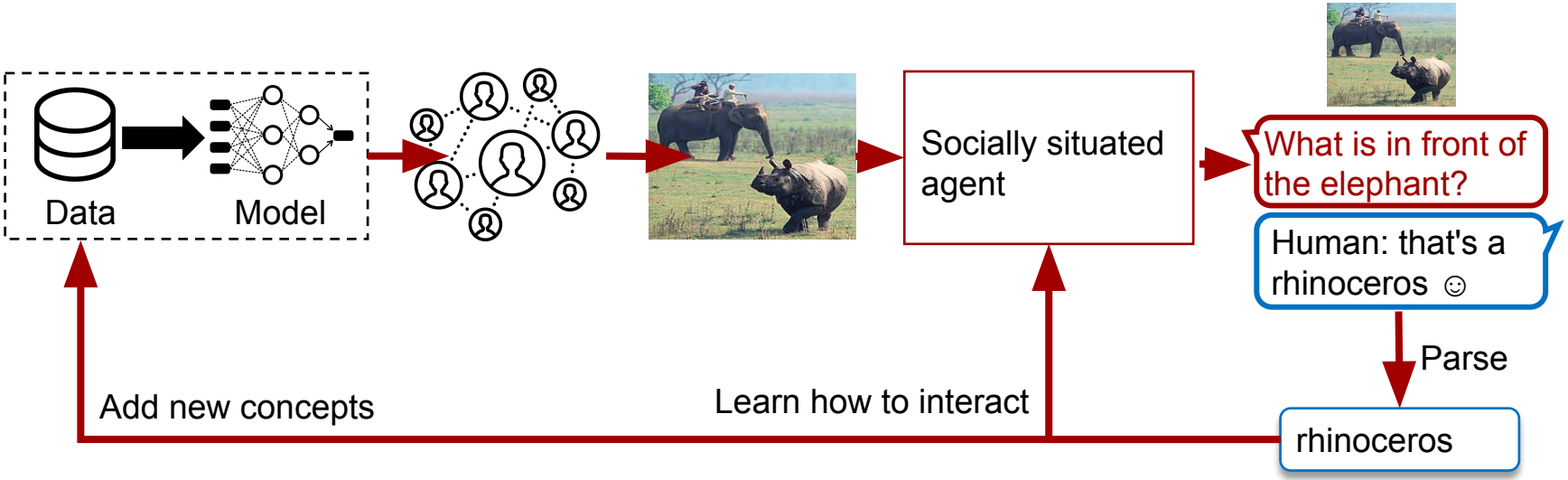
Q: What kind of bird is this?
A: A Barn Owl.

Q: What is on the counter?
A: On the counter you can find a wide variety of chocolates, dragees and all kinds of refined sweets!

Traditional AI training paradigm



Socially situated artificial intelligence framework



Krishna, Lee, Fei-Fei, Bernstein. in submission 2021

Formalizing as a question generation reinforcement learning task

Socially situated agent

What is in front of the elephant?

policy
$q \sim \pi_\theta(i)$

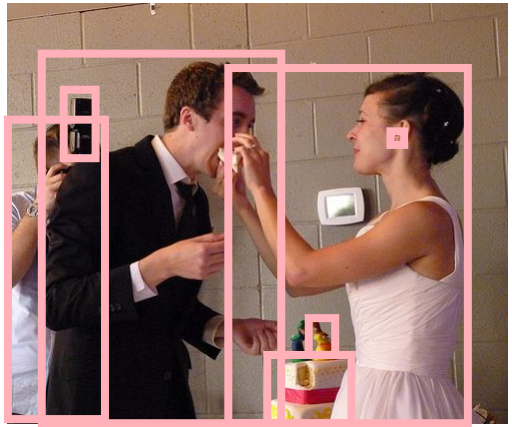
q_1	q_2	q_3	q_4
in	in	**in**	in
front	front	front	**front**
did	did	did	did
what	what	what	what
⋮	⋮	⋮	⋮
from	from	from	from
is	**is**	is	is
umbrella	umbrella	umbrella	umbrella
where	where	where	where

…

q : What is in front of the elephant?

Lewis et al. EMNLP 2017; Li et al. EMNLP 2016

Initialized using Visual Genome



Questions

1. Why is she giving him cake? To celebrate.
2. What is on the wall? Thermostat.
3. How many people? 3.
4. Where is the wedding? In a building.
5. What is the person in the back holding? Camera.
6. What color is the dress? pink.
7. What is the person doing with the camera? Taking a picture.
8. Who is wearing the suit? Groom
9. Who is holding the cake? Bride.
10. What is on top of the cake? two statues
11. What is the man doing? Eating.
12. What is the girl doing? Feeding the man cake.
13. What color is the middle layer of the cake? Red.
14. is the bride's hair tied up? Yes.
15. Where is the thermostat? On the wall.
16. Who is taking the picture? The person behind the couple.
17. Are they celebrating a wedding? Yes.

108K images

1.7M questions

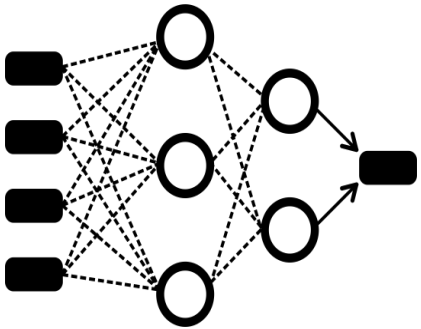
Krishna, Zhu, Hata, Johnson, Kravitz, Chen, Li, Shamma, Bernstein, Fei-Fei. IJCV 2017

Code and dataset available: http://visualgenome.org
Visualization code: https://github.com/ranjaykrishna/graphviz

Two rewards to guide the agent's behavior

What is in front of the elephant?

Knowledge reward:
Produce questions that are hard for the current model



Uncertainty in output space

Yang et al. CVPR 2016; Settles et al. 2012; Lin et al., CoRR 2017; Hunt. The Guardian, 2016; Neff and Nagy. International Journal of Communication, 2016

Ranjay Krishna | ranjay@cs.washington.edu

Interaction reward:
Produce questions that people will respond to

Sarah: What do you mean?

Sarah: <-no response->

Negative interactions

Sarah: that's a rhinoceros ☺️

Informative responses are positive interactions

Webb et al. 1999. Zhao, Zhe, et al. ACM Conference on Recommender Systems. 2019.

Knowledge + interaction are both necessary

Only knowledge reward:

Agent will generate questions people don't want to respond to

Q: is the person sitting behind the person guiding the elephant holding their right hand or their left hand up to their eyes?

Only interaction reward:

Agent will generate questions it already knows the answer to

Q: What color is the grass?

Ranjay Krishna | ranjay@cs.washington.edu

Let’s put all of this together in a workflow

Socially situated AI – Putting it all together

Social Media

Our agent

OUTPUT: dynamically improving AI model


What is the food on the table? **it is a pizza**


Why is the person wearing a helmet? **they are bicycling**


What is the person sitting doing? **reading a book**

Ranjay Krishna | ranjay@cs.washington.edu

Yang et al. "Stacked attention networks for image question answering." CVPR 2016

First, we filter images with a trained model

Social Media

Sarah: Found these little ones outside my apartment.

images

Our agent

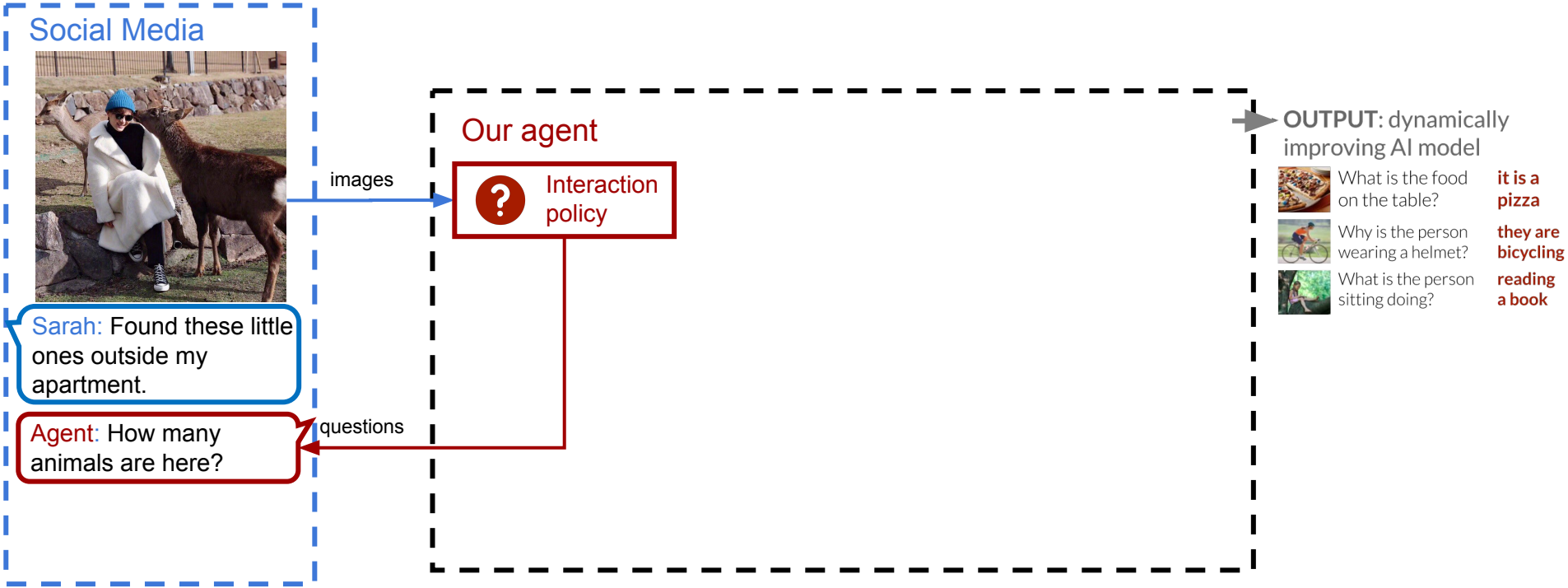
OUTPUT: dynamically improving AI model


What is the food on the table? **it is a pizza**

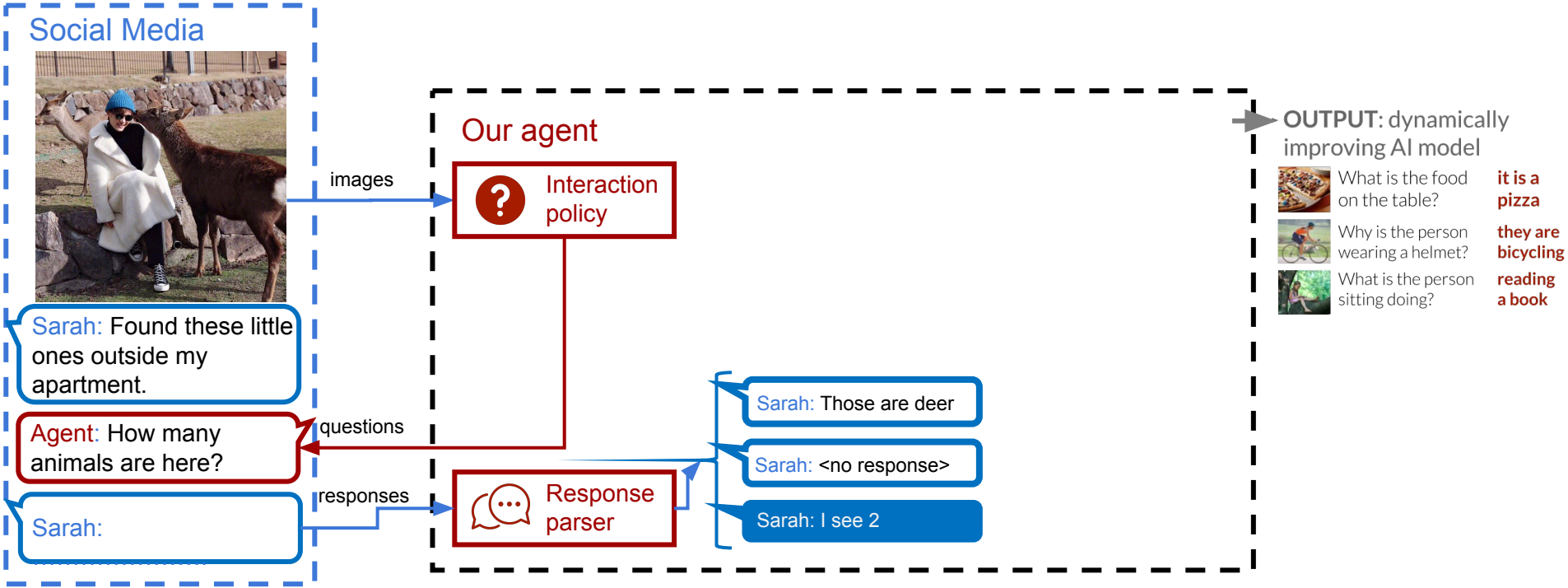

Why is the person wearing a helmet? **they are bicycling**


What is the person sitting doing? **reading a book**

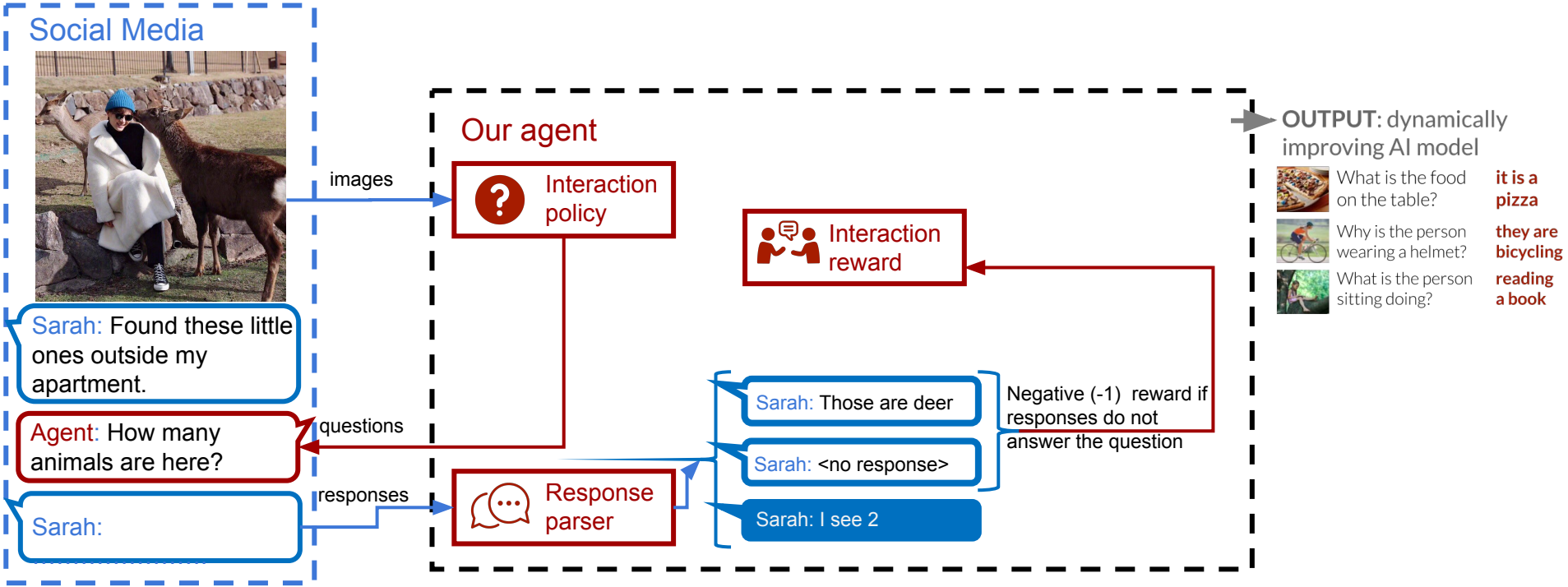
First, we filter images with a trained model



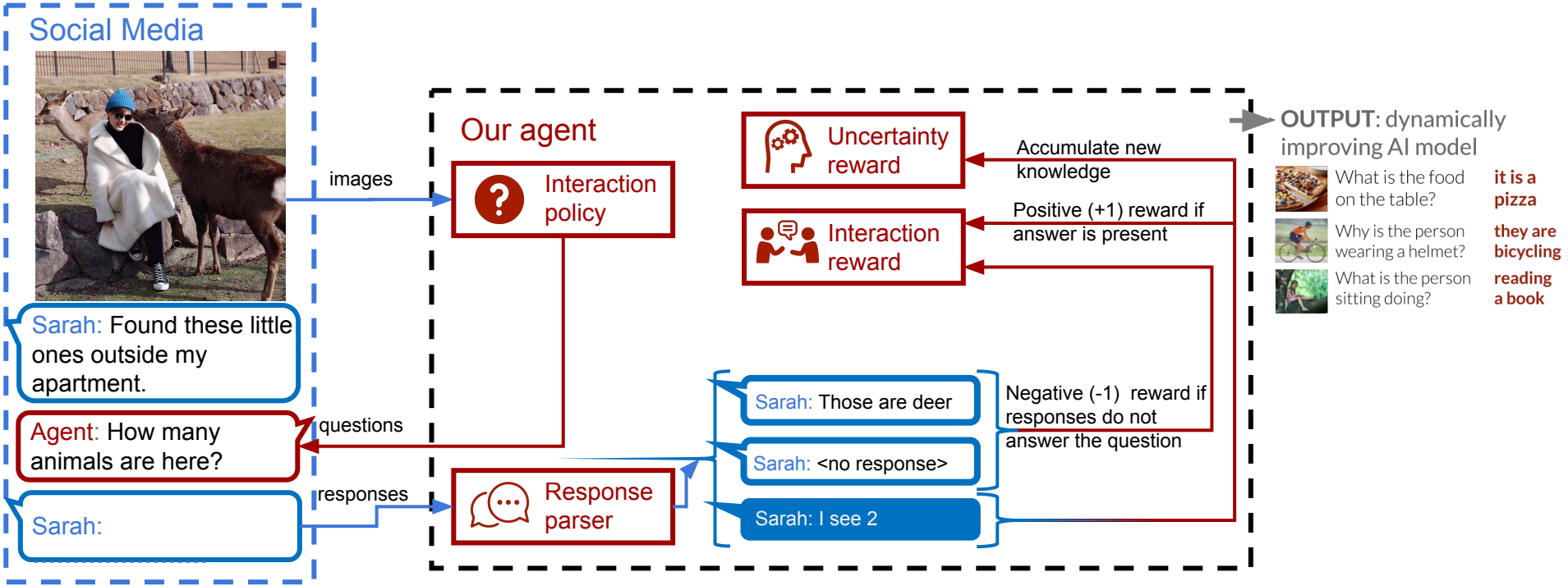
First, we filter images with a trained model



First, we filter images with a trained model

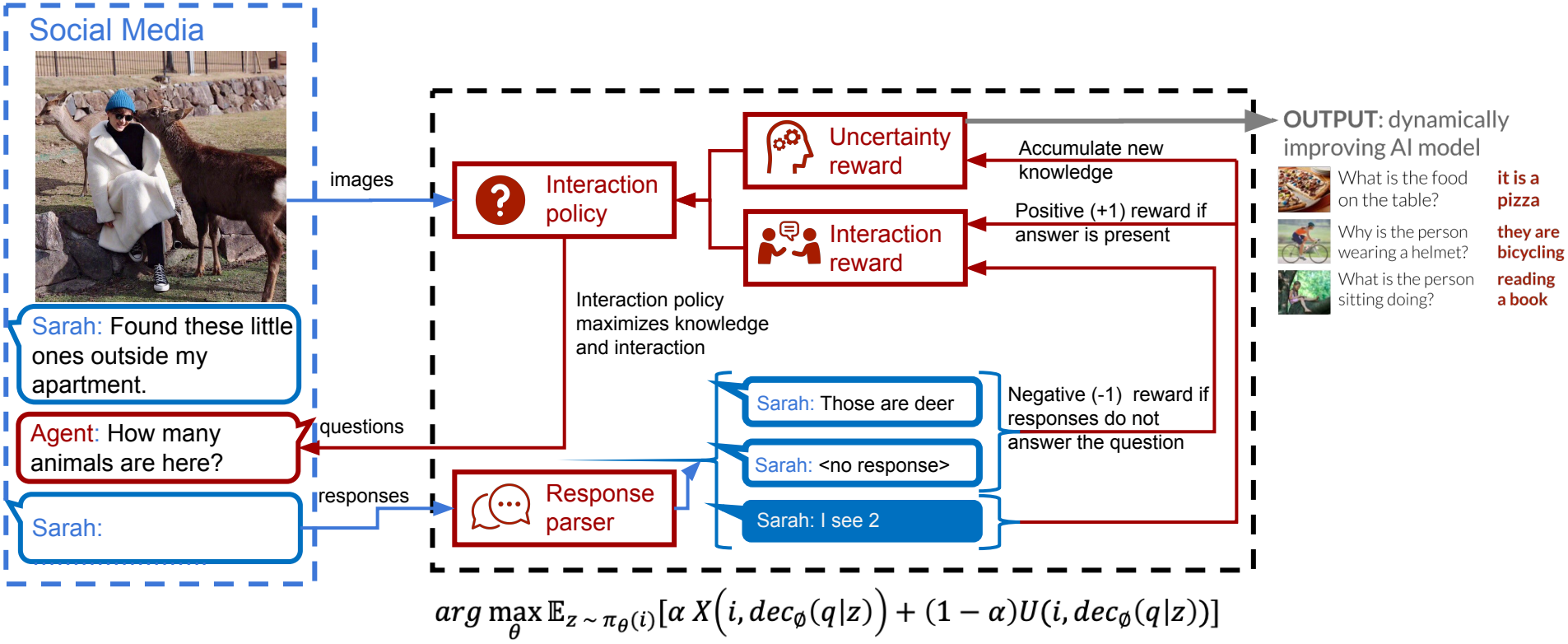


First, we filter images with a trained model



Yang et al. "Stacked attention networks for image question answering." CVPR 2016

Learning to Interact – the complete reinforcement learning framework



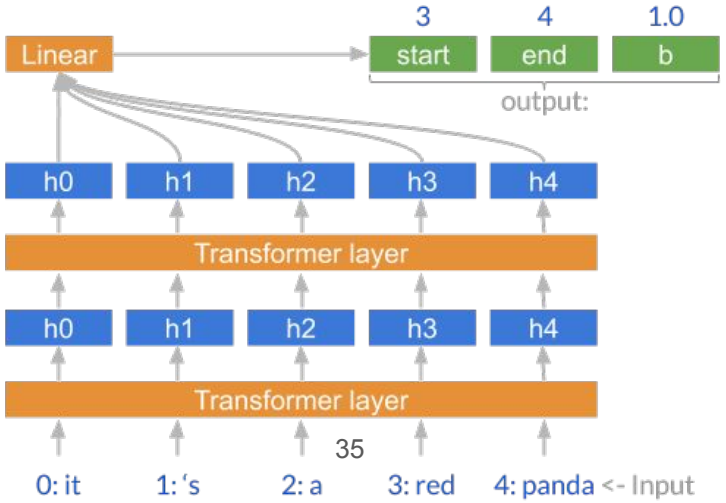
$$arg\max_{\theta} \mathbb{E}_{z \sim \pi_\theta(i)}[\alpha\, X\Big(i, dec_\emptyset(q|z)\Big) + (1-\alpha)U(i, dec_\emptyset(q|z))]$$

Responses parsed using finetuned BERT-small model

- We train on a dataset of 50K human parsed responses from social media (annotated by crowd workers)

Answer exists:
- Precision: 0.73
- Recall: 0.67

Answer Span Prediction
- Precision: 0.73
- Recall: 0.67

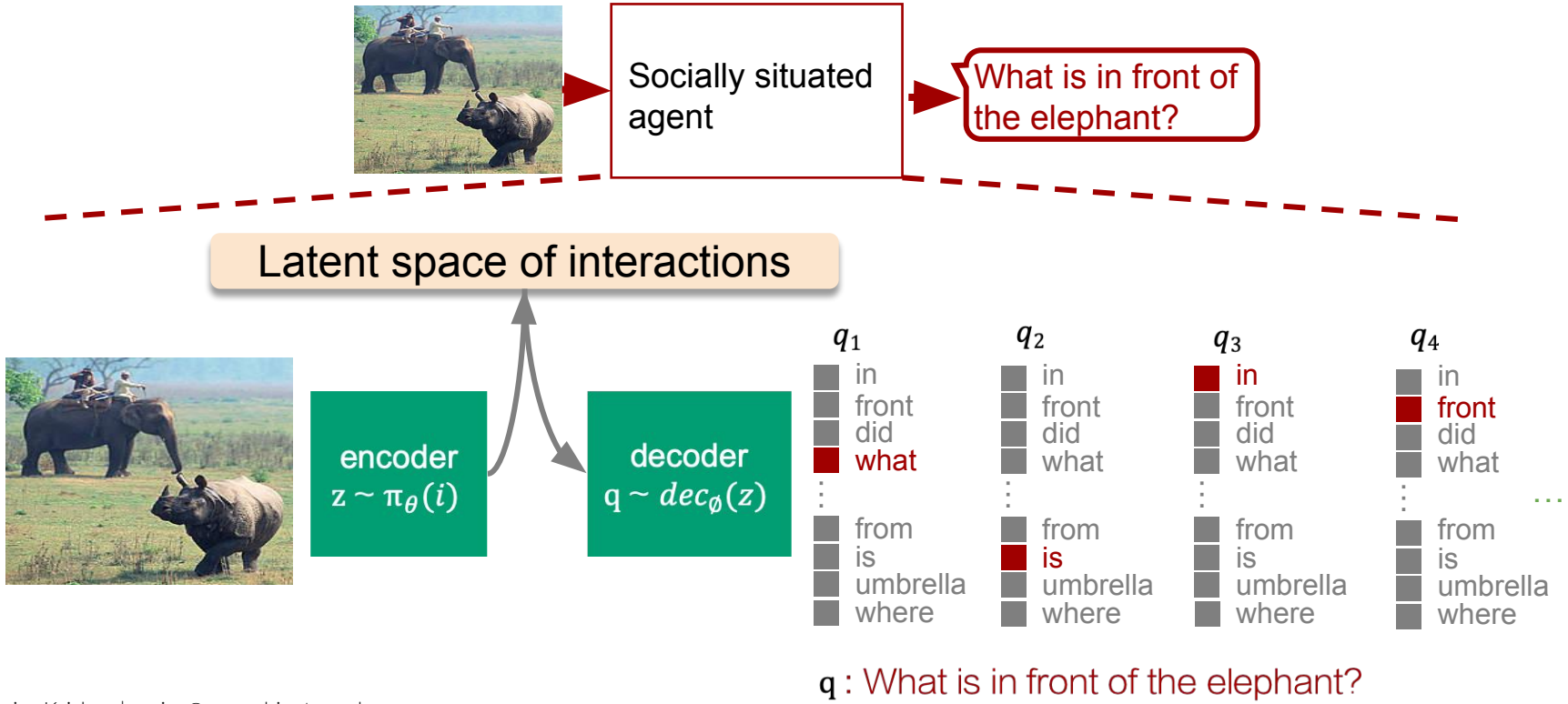


Devlin, et al., Bert: Pre-training of deep bidirectional transformers for language understanding, NAACL 2019

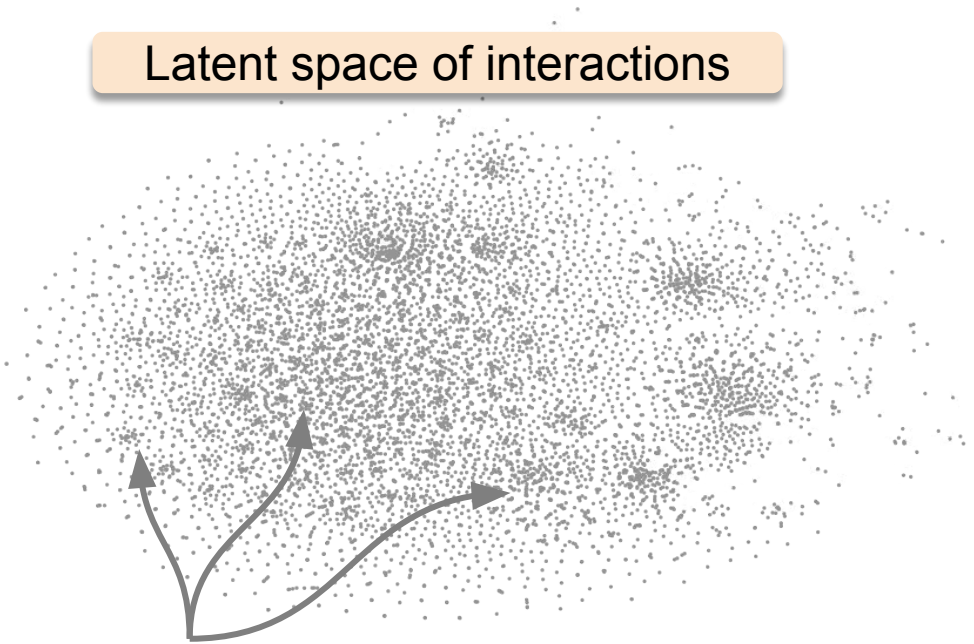
Ranjay Krishna | ranjay@cs.washington.edu

Reinforcement learning in language space is really hard to train

Formalizing as a question generation reinforcement learning task



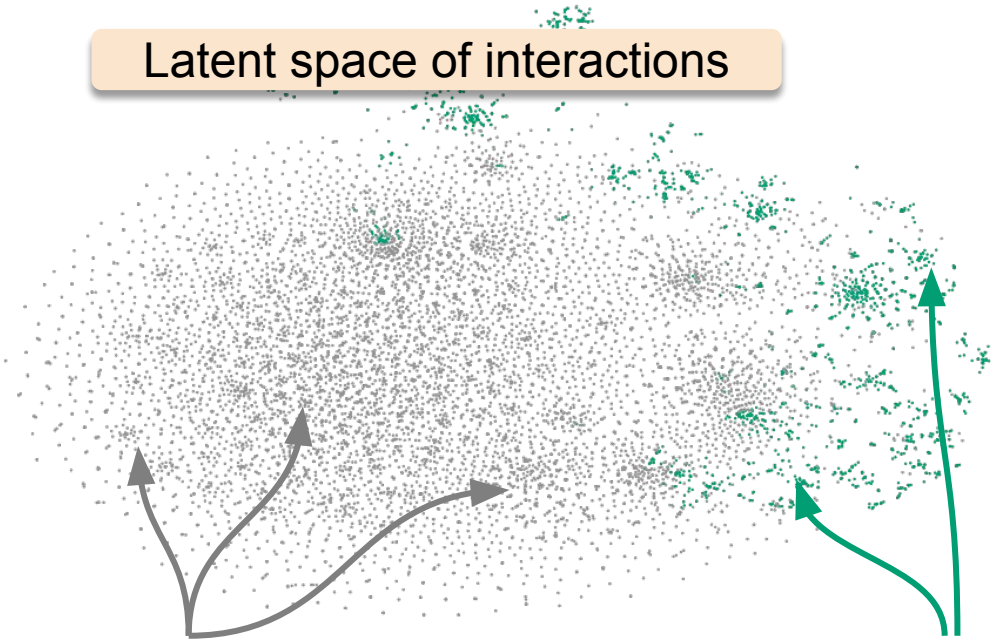
The space of interactions is combinatorically vast



Language reward

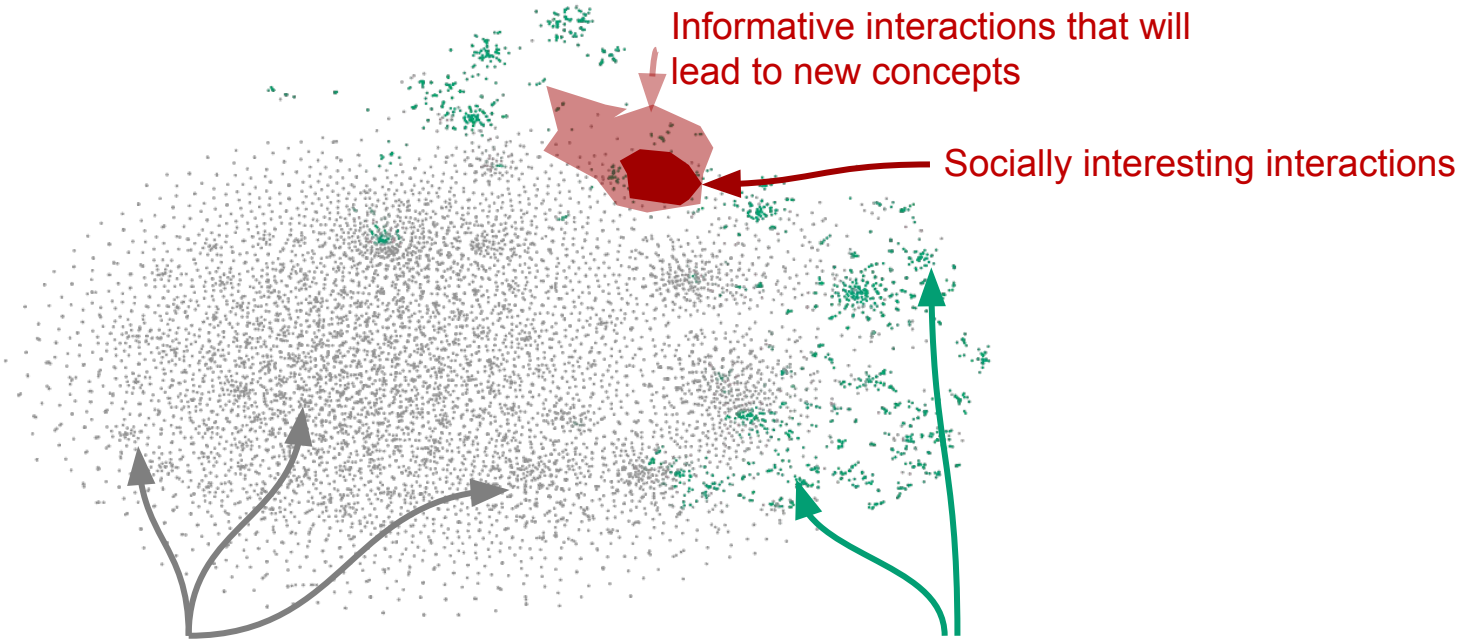
in white:
the space of all possible sentences

Questions aiming to learn is a small subset of all interactions

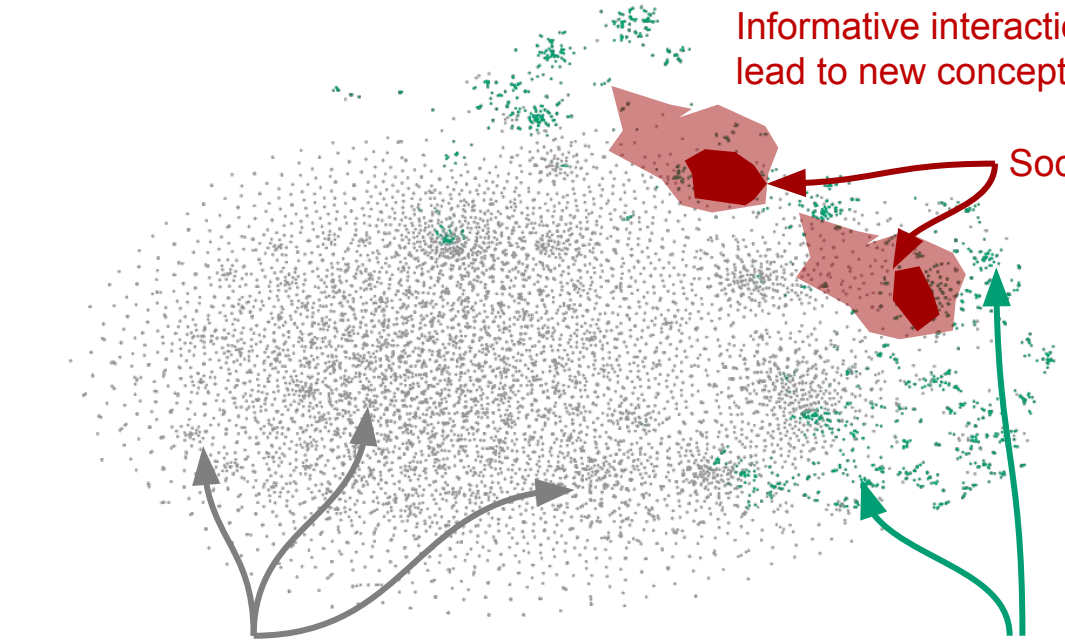


in white:
the space of all possible sentences

Agent must uncover informative AND social interactions



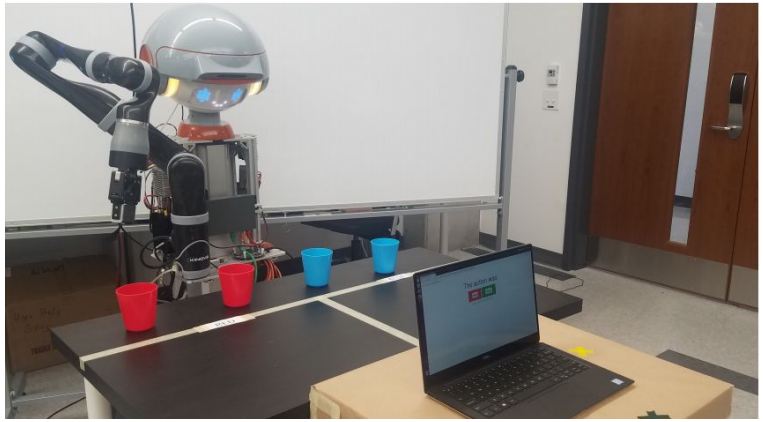
As the model learns, the informative space will shift



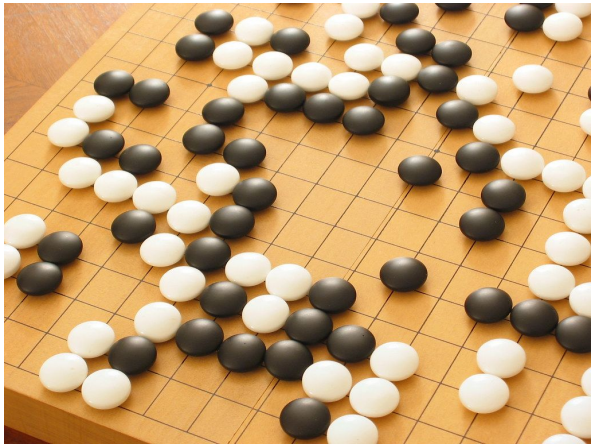
The optimal behavior is non-stationary

Existing methods only successful under 2 conditions:

(1) Feedback from people is explicit

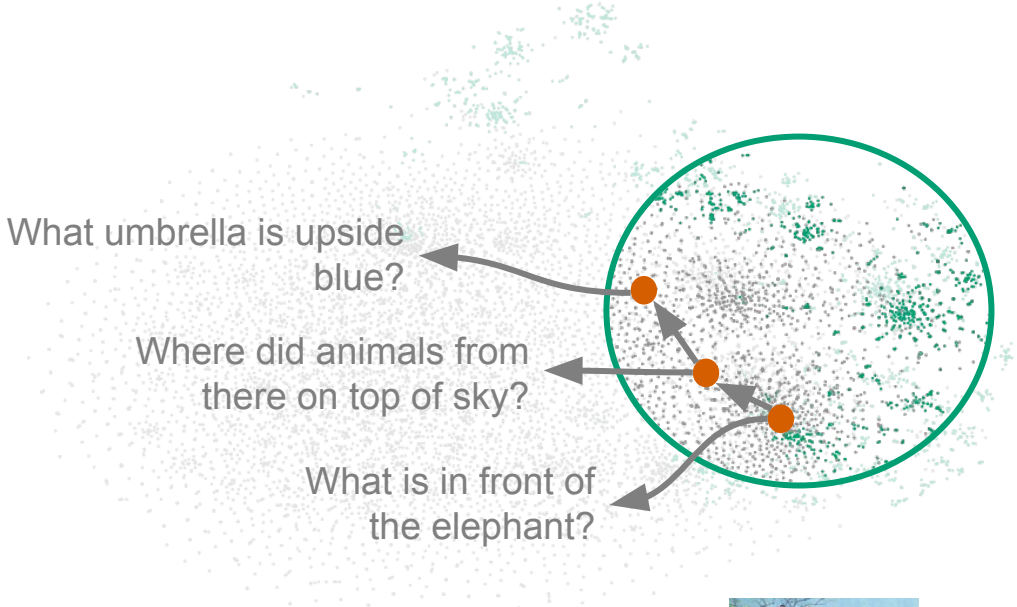


(2) Small action spaces



Christiano et al. NeurIPS. 2017; Silver et al. Nature 2016; Misra, et al. CVPR. 2018; Thomaz and Breazeal. AAAI 2006; Thomaz and Breazeal. AI 2008; Cakmak and Thomaz. HRI 2012

Existing methods struggle to explore this space of interactions



Rewards

Interaction reward

Knowledge reward

Language reward

Our insight: Large percentage of the variance of human interactions lie on a low-dimensional manifold

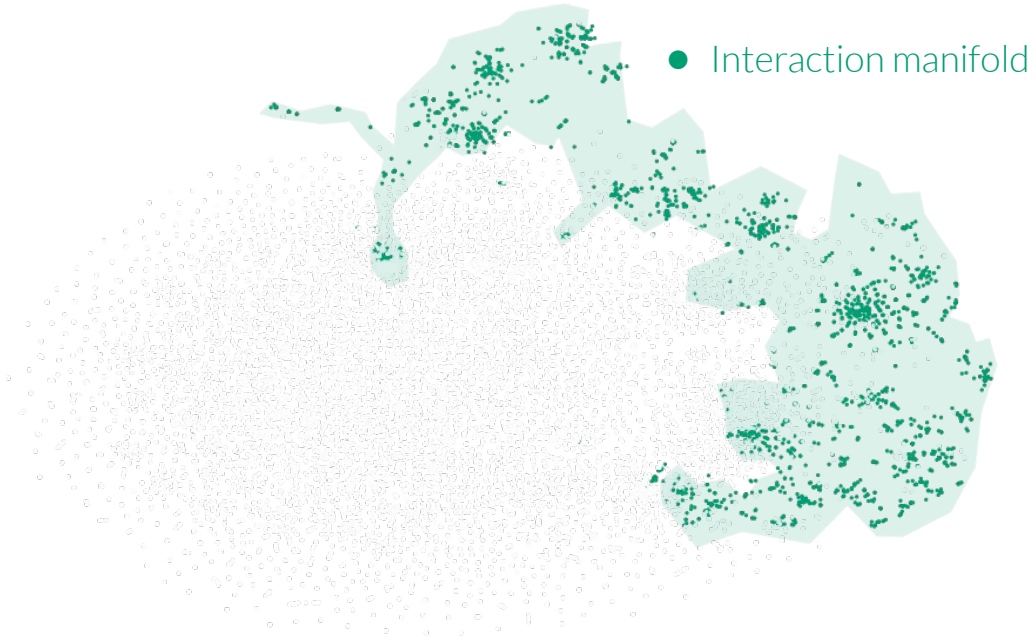
Interactions that people will understand

Questions asking about visual content

Ranjay Krishna | ranjay@cs.washington.edu

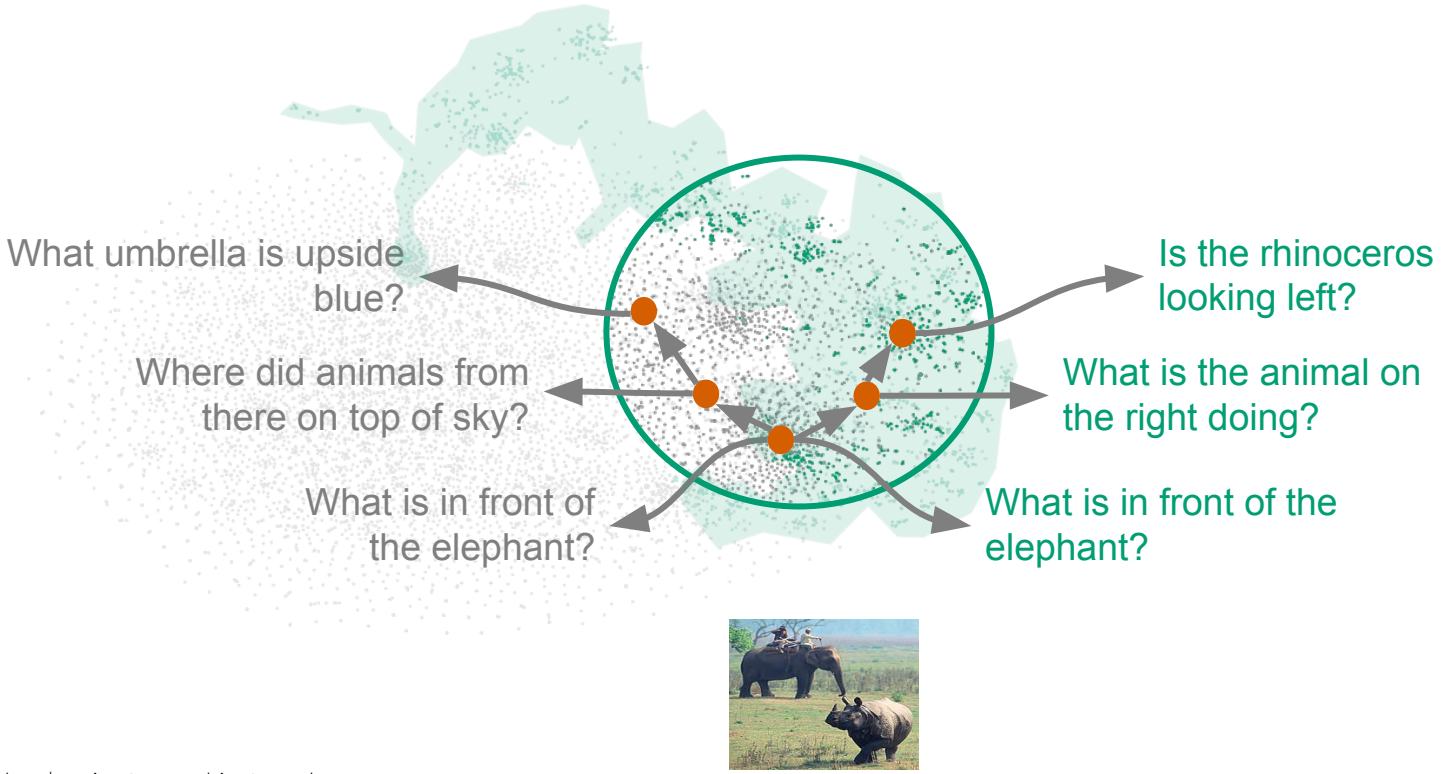
Sherif. 1936; Zipf. 1949; Goffman. Harmondsworth, 1978; Schank et al. Psychology Press, 2013; Fehr et al. Trends in cognitive sciences 2004;

Our algorithm: an interaction manifold as a surrogate action space

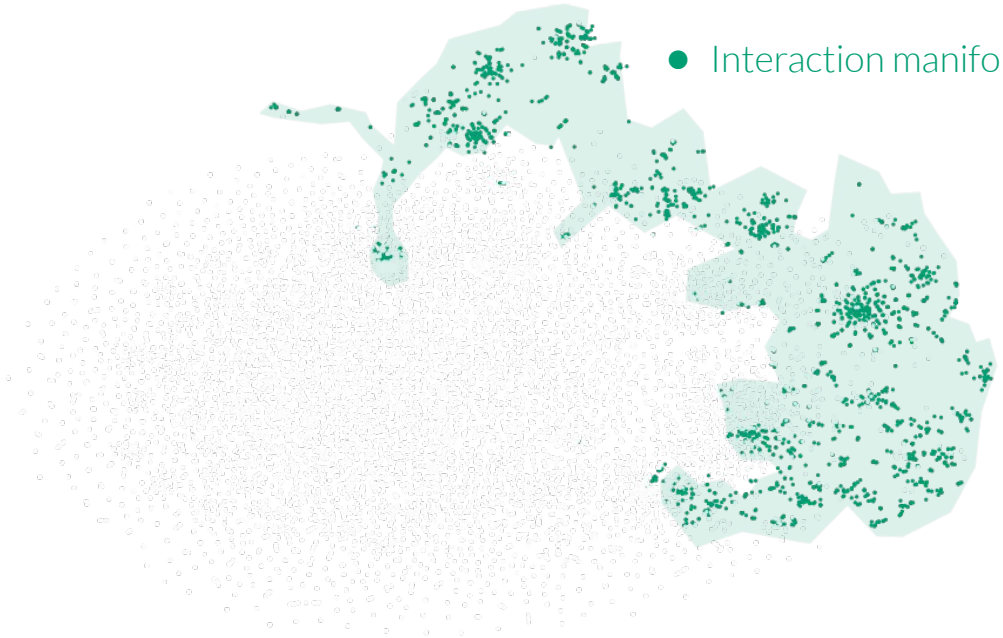


Krishna, Bernstein, Fei-Fei. Information maximizing visual question generationCVPR 2019

Restricting actions to lie within the interaction manifold



Our algorithm: the interaction manifold as a surrogate action space



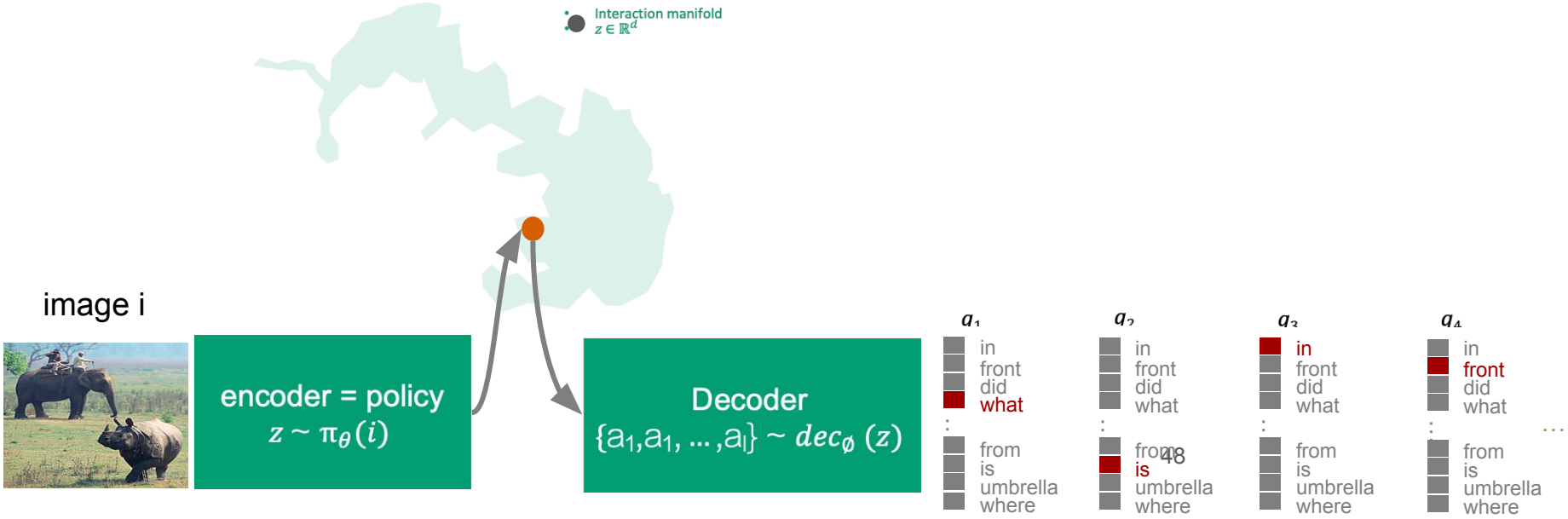
Rewards

Interaction reward

Knowledge reward

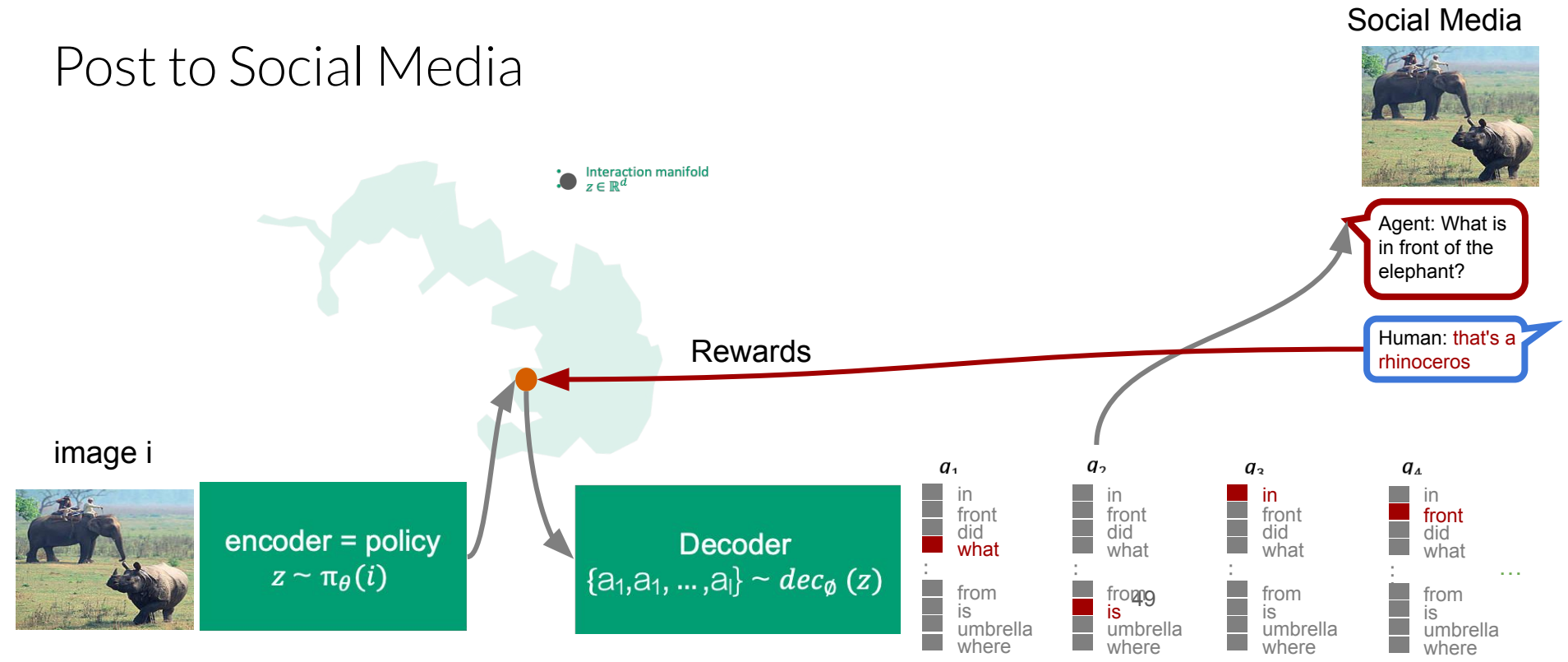
Language reward

Putting it all together

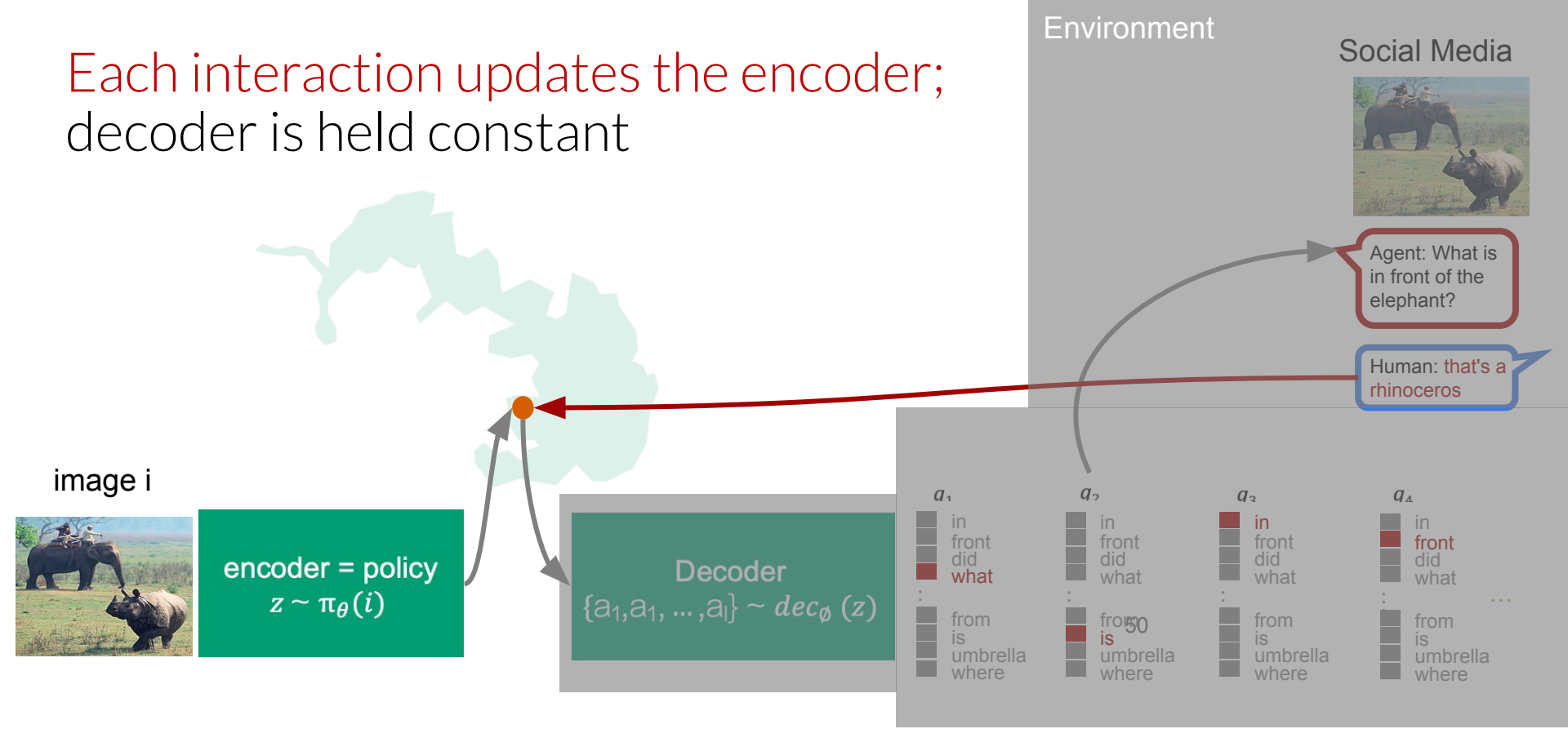


Action space: d = 500

Post to Social Media



Each interaction updates the encoder; decoder is held constant



Interacting with people on social media

Interacted with 230K people over 8 months

Real interactions between our agent & people on social media

Q: What is the dog's tail resting on?
A: it's a public restroom.

Q; Is the board room carpeted?
A: we use wood floor.

Q: Is this person wearing a life vest?
A: Ahahah not at all ! She is wearing a big coat 😉

Questions about objects

Q: What is the white stuff on the table?
A: mayo!.

Q: Is that a real bird?
A: 😂😂😂 it's a crocodile

Q: Is this a commercial plane?
A: yes, you're right! Flight coming from Amsterdam to Saint Martin!

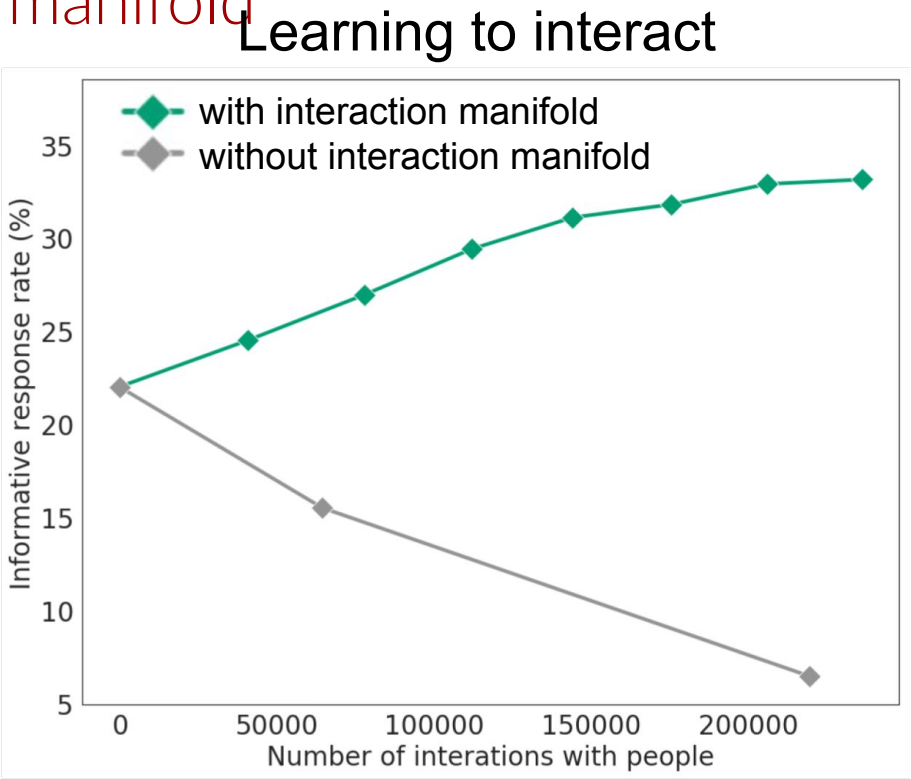
Questions about attributes

Q: What is the shape of the sink?
A: It's a square.

Q: What material is the counter?
A: It looks as though it is marble, however this isn't my design so I can't be 100%. It's gorgeous though isn't it!

Q: What kind are the pink ones?
A: Hi, it is japanese cherry flower - Prunus serrulata, the others: paeonia, anemone, ranuculus.

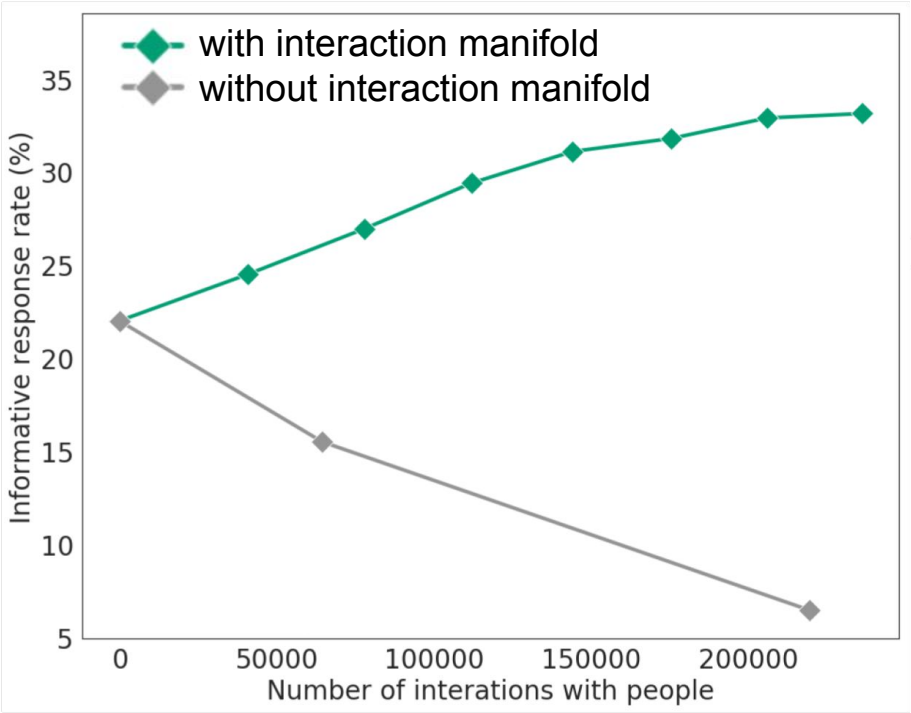
Informative response rate with and w/o the interaction manifold



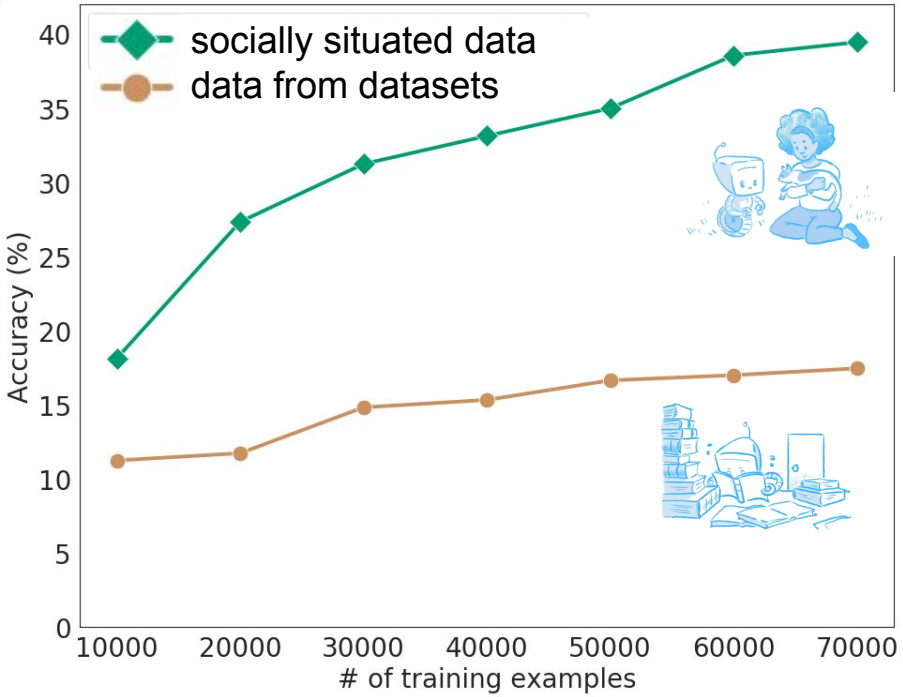
Krishna, Lee, Fei-Fei, Bernstein. Socially Situated Artificial Intelligence. PNAS 2022

Learning with traditional versus socially situated data

Learning to interact



Interacting to learn



Krishna, Lee, Fei-Fei, Bernstein. Socially Situated Artificial Intelligence. PNAS 2022

We acquire new concepts that need expert knowledge

Q: What kind of bird is that?

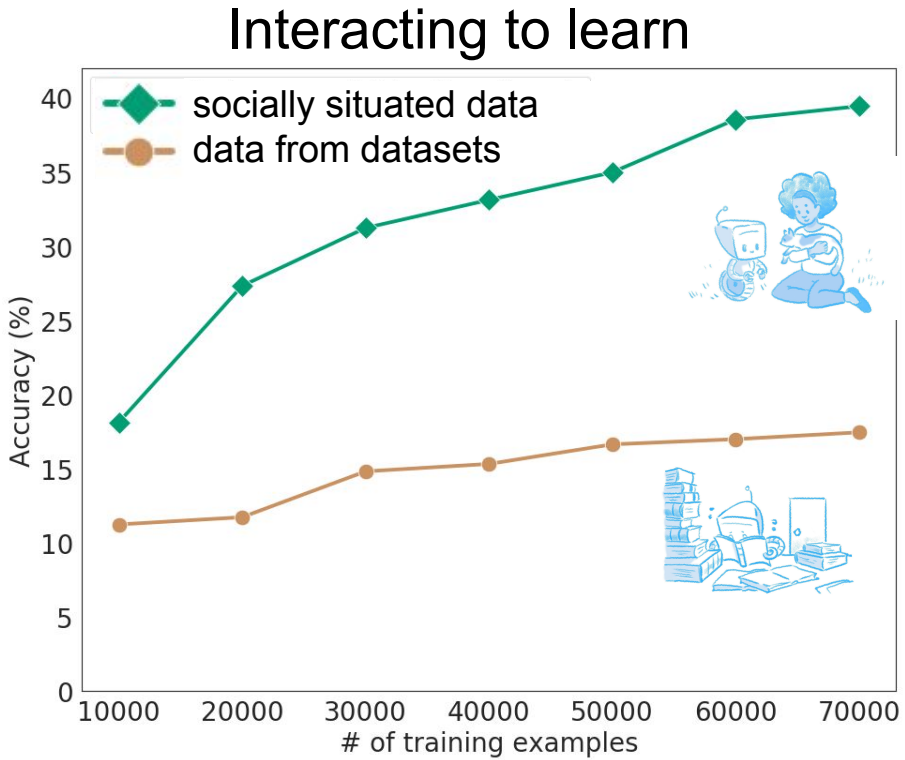
A (by AI): Magpie

Q: What kind of flower is that?

A (by AI): Dahlias

Q: What is the white stuff on the plate?

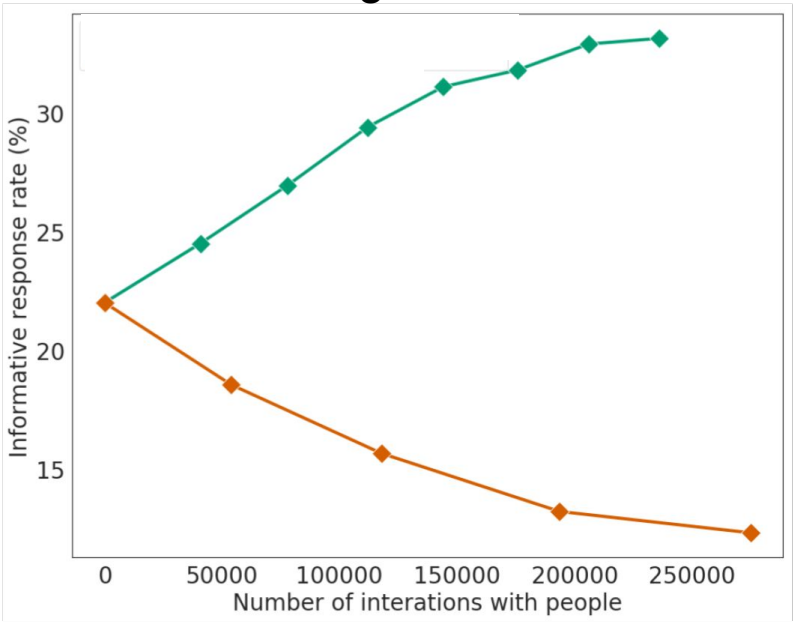
A (by AI): Feta cheese



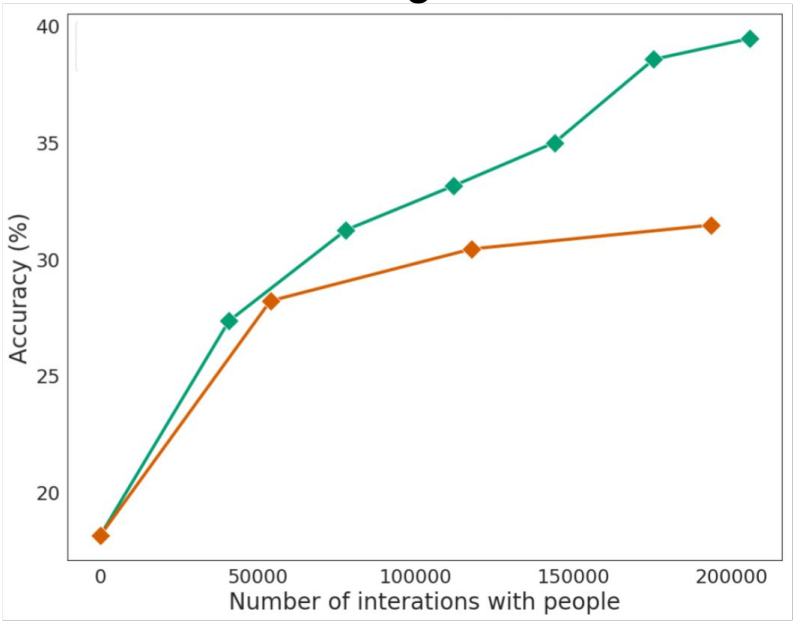
Krishna, Lee, Fei-Fei, Bernstein. Socially Situated Artificial Intelligence. PNAS 2022

Why the interaction reward is important – not following norms results in fewer interactions

Learning to interact



Interacting to learn



with knowledge + interaction reward
only with knowledge reward

Emergent agent behavior was consistent with social science lit

Easy to answer

Existence questions

Is this a restaurant?
Is that a bear?

Color questions

What is the color of the wall behind the fabric?

Questions with multiple correct answers

Questions asking why

Why is the man wearing gloves?

Vague questions

What is the child doing?

Questions that exhibit social proof

Mentioning known concepts

What is in front of the teddy?
What is the orange food?

Questions that require cognitive load

Long questions

is the tool held by the person in the middle made for someone right-handed or left-handed?

Morris et al. CHI 2010; Lakhani, et al. MIT 2004; Krishna, et al. CVPR 2019; Cialdini. Collins 2007; Kraut 2011;

How people ask for help: People augment questions with social strategies

Q: that is very good looking, what is the name of the dish?
A: it is a caribbean dish named << crab pie >>, very tasty!

Q: What type of bread is this? It looks like a sourdough with something in it.
A: yes, there are sun dried tomatoes and beet greens in it.

Q: this type of art is called what, i have seen it before?
A: looks a bit steampunk I suppose, but created well before that term was thought of

Q: I love the colors in this, was it edited in any way or natural?
A: thank you so much, not edited at all, just nature doing its best work! 🏞😁

questions written by hired workers | response from an online user

- Park, Krishna, et al. "AI-Based Request Augmentation to Increase Crowdsourcing Participation." HCOMP 2019

Ranjay Krishna | ranjay@cs.washington.edu

Ethics & design decisions

- Which social media posts should agents interact with?
  - only publicly accessible content with common hashtags that encourage engagement.

Ranjay Krishna | ranjay@cs.washington.edu

Ethics & design decisions

- Which social media posts should agents interact with?
  - only publicly accessible content with common hashtags that encourage engagement.
- How should AI agents self-identify?
  - Yes, before and after the interaction - "We are a computer science research project". Vet interactions with people.

Ranjay Krishna | ranjay@cs.washington.edu

Murgia. "Who's using your face? The ugly truth about facial recognition." Financial Times, 2019
Solon. "Facial Recognition's 'Dirty Little Secret': Millions of Online Photos Scraped without Consent." NBC News, 2019

Ethics & design decisions

- Which social media posts should agents interact with?
  - only publicly accessible content with common hashtags that encourage engagement.
- How should AI agents self-identify?
  - Yes, before and after the interaction - "We are a computer science research project". Vet interactions with people.
- What are the risks associated with diverting people's attention?
  - Stop agents that do not receive positive reactions or responses

Ranjay Krishna | ranjay@cs.washington.edu

Ethics & design decisions

- Which social media posts should agents interact with?
  - only publicly accessible content with common hashtags that encourage engagement.
- How should AI agents self-identify?
  - Yes, before and after the interaction - "We are a computer science research project". Vet interactions with people.
- What are the risks associated with diverting people's attention?
  - Stop agents that do not receive positive reactions or responses
- What are the risks associated with malicious usage of our technology?
  - Rewards that agents use should be made publicly transparent

Cialdini. "Influence: The psychology of persuasion." 1993
Cheng, et al. "Anyone can become a troll: Causes of trolling behavior in online discussions." *CSCW*. 2017
Kramer et al. "Experimental evidence of massive-scale emotional contagion through social networks." *Proceedings of the National Academy of Sciences* 2014

Ranjay Krishna | ranjay@cs.washington.edu

Ethics & design decisions

- Which social media posts should agents interact with?
  - only publicly accessible content with common hashtags that encourage engagement.
- How should AI agents self-identify?
  - Yes, before and after the interaction - "We are a computer science research project". Vet interactions with people.
- What are the risks associated with diverting people's attention?
  - Stop agents that do not receive positive reactions or responses
- What are the risks associated with malicious usage of our technology?
  - Rewards that agents use should be made publicly transparent
- What are the biases inherited from readily available data?
- - Manual cleanup and inspection of datasets used

Ranjay Krishna | ranjay@cs.washington.edu

Antol, et al. "VQA: Visual question answering." *ICCV*. 2015

Krishna, et al. "Visual genome: Connecting language and vision using crowdsourced dense image annotations." *IJCV* 2017

Ethics & design decisions – Biases on social media

Ranjay Krishna | ranjay@cs.washington.edu

Images taken from Emma Sheffer's https://www.instagram.com/insta_repeat

Limitations – not capitalizing on corrective responses

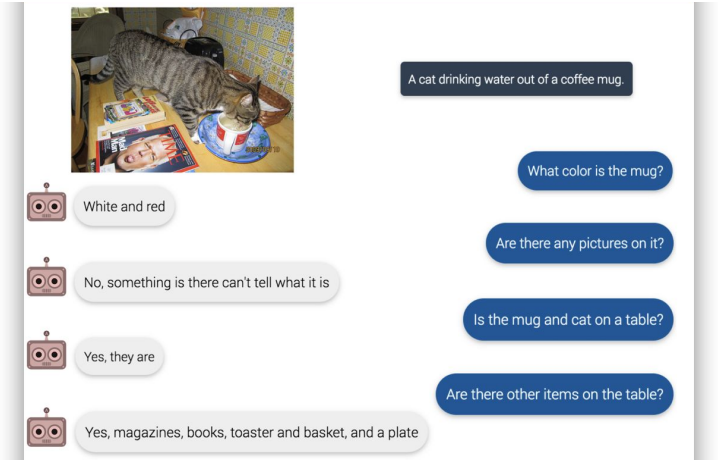
Agent: What is the cat doing?

Sarah: You mean the goat? It is trying to climb a tree.

Ranjay Krishna | ranjay@cs.washington.edu

Stumpf et al. "Toward harnessing user feedback for machine learning." ICIUI 2007

Limitations – longer dialogues



Ranjay Krishna | ranjay@cs.washington.edu

Das, et al. "Visual dialog." CVPR 2017

Surprisingly - ChatGPT has a similar set of training steps

Instruct GPT

- Finetunes GPT-3 using human generated instructions and outputs
  - Similar to our question generation from visual genome)
- Trains a reward model that ranks good versus bad generations
  - Similar to our knowledge and interaction rewards
- Using reinforcement learning to train using the reward model
  - Similar to our step

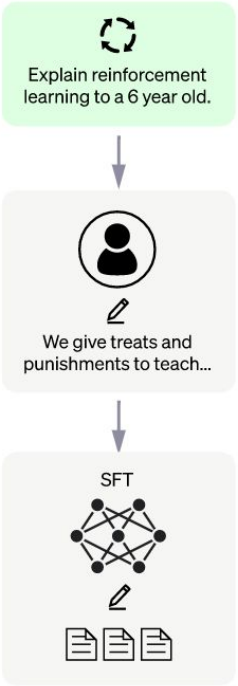
Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.



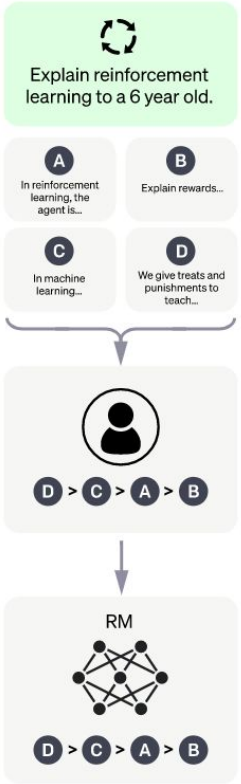
Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

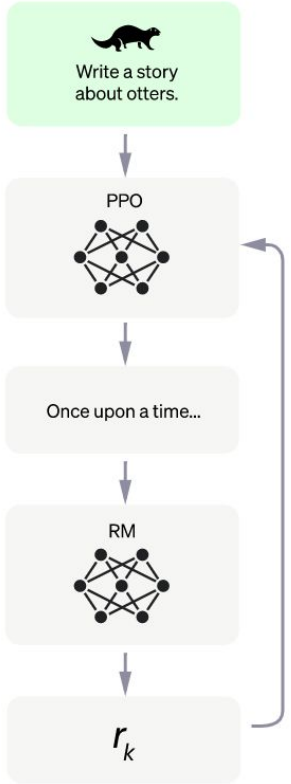
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



What we do differently?

- No foundation models when we started this project
- Models are much smaller
- No rankings used
- Released on real social environment.
  - No training people how to interact but learn interactions from implicit signals
- Only interacted with using questions and not instructions
- Iteratively ran multiple rounds of reinforcement learning and reward model training

Opportunities for so much future research

What is the right interaction modality?

Is pretraining necessary?

How noisy can our reward models be?

How quickly can we bootstrap reward models?

Can we allow a single person to design few-shot reward models and personalize their foundation models?

How do we prevent RL from degrading original model?

Ranjay Krishna | ranjay@cs.washington.edu

Next time: Creativity and generative models