Bodhisattwa Prasad Majumder UC San Diego



UC San Diego

JACOBS SCHOOL OF ENGINEERING Computer Science and Engineering



Effective, Explainable, and Equitable NLP with World Knowledge and interactions









How Bad Is the Teacher Shortage? Depends Where You Live.

Urgently needed: teachers in struggling districts, certified in math or special education. Perks: maybe a pay raise, or how about a four-day week?

100K schools in India run with just 1 teacher each

Can Al help?





United Nations

- Educational, Scientific and
 - Cultural Organization •



Start of a new session! I am so confused what courses to take to learn AI. I need a job in AI.



Hi, welcome! I am here to help

You must take Intro to Machine Learning Intro to Data Mining

Relevant



Why do I need Data Mining?



It has been very popular in recent graduate students.

99% of who took the class last year, got a job in Al.

Trustworthy



Okay great! But I am weak in math. Will I be able to perform well in these courses?



Never mind! I talk to many students who do not have formal background in math.

You can consider taking *Intro to Probability* as a refresher course.

Adaptive



Start of a new session! I am so confused what courses to take to learn AI. I need a job in AI.



Hi, welcome! I am here to help

You must take Intro to Machine Learning Intro to Data Mining

Relevant



Why do I need Data Mining?



It has been very popular in recent graduate students.

99% of who took the class last year, got a job in Al.

Trustworthy



Okay great! But I am weak in math. Will I be able to perform well in these courses?



Never mind! I talk to many students who do not have formal background in math.

You can consider taking *Intro to Probability* as a refresher course.

Adaptive

Insights from social sciences [Miller et al., 2019]

Properties of an Assistive AI



Human Friend Relevant, Trustworthy, and Adaptive M



I am an animal activist.



Why do you work with animals?

To make a difference.

understanding the world

Tell me some experiences you had.

I went to a dog shelter last week. I was happy seeing an injured dog recovering well.

experience



+ <mark>℃</mark> + 10 10 01 01

Is there a good animal shelter around LocationX?



Yes, Place1 and Place2 are open most days in a week.

up-to-date knowledge



You use detergent to dye your hair

False *because* detergent can clean but cannot dye either your hair or clothes.

explaining with commonsense



factual explanations



bias understanding



recommend Game of Thrones **because** fantasy, politics, drama

don't want a political series

Try The Hobbit

critiquing recommendations

What was great in the hotel?

Acc. to previous reviews, the rooftop pool gives an amazing view

Predict profession w/o gender info: He is a congressman and he sings

Congressman *imp tok:* congressman bias tok: he



Magnet *can* pick a penny *bcz* penny is made of metal magnet attracts magnetic metals



Penny is made of copper



Magnet *can* pick a penny



Copper is non-magnetic



Magnet *cannot* pick a penny

feedback, teaching, learning





understanding the world

experience



up-to-date knowledge



False *because* detergent can

explaining with commonsense



bias understanding



Current Al is not good — why?



because fantasy, politics, drama

critiquing recommendations

congressman *bias tok:* he

factual explanations





feedback, teaching, learning



Data A is temporal, biased, limited by its origin e.g. pre- and post covid travel regulations

[Logan IV et al., 2022]



can be opaque, contain spurious correlation e.g. uses syntactic nuances instead of contextual knowledge for an NLI task

[Gardner et al., 2021]

Evaluation

can be done offline, may not address subjectivity e.g. recommender systems are evaluated offline no evaluation for new users (cold-start)

[McAuley et al., 2013]







Data + Knowledge 🤎

augment with explicit/implicit knowledge

e.g. fine-tuning or post-hoc injection [Majumder et al., 2020; 2021; 2022]

Model + Explanations



to produce both predictions and explanations

e.g. extractive explanations as attributions or abstractive explanations as beliefs [Majumder et al., 2022a; 2022b]

Evaluation + Interactions



human-in-the-loop learning, user studies

e.g. measuring success in achieving conversational goal instead of next response accuracy

[Majumder et al., 2022a; 2022b]

Way forward: Interactive Explainability

*Recognized by Adobe Research Fellowship 2022, Qualcomm Innovation Fellowship 2020





understanding the world

experience

up-to-date knowledge



explaining with commonsense



bias understanding



factual explanations





critiquing recommendations



feedback, teaching, learning

Interactions



Persona-based Commonsense Majumder et al. **EMNLP** 2020

Post-hoc Knowledge Injection Majumder et al. **ACL** 2021, **ACL** 2022

Personalized Knowledge Grounding Majumder et al. **EMNLP** 2019

Commonsense Grounding in Stories Mao, **Majumder** et al. **EMNLP** 2019

Knowledge Grounded Self-rationalization Majumder et al. **ICML** 2022

Controlling Bias Exposure via Rationales He, Yu, McAuley, Majumder **EMNLP** 2022

Faithfulness in Language Explanations Xie, McAuley, Majumder **Preprint** 2022

Factual Explanation Generation Xie, Singh, McAuley, Majumder **Preprint** 2022



Estimating Missing Knowledge Majumder et al. **NAACL** 2021

Conversational Recommendation Li, **Majumder** et al. **RecSys** 2022

> **Interactive Fair Debiasing** Majumder et al. InterNLP 2022

Bernard: Human-centric NLP Majumder et al. Alexa Proc. 2021

Next-generation Al







Interactions

Select publications





Hiuen Tsand's iournev to the west



+ future directions

Interactions

Next-generation Al











+ future directions

Interactions

Next-generation Al









Dialog Context



Dialog Context

A **baseline** response



Many prefer to visit museums. You can do hiking around the river if you love nature. Or you can watch a movie. Which one do you prefer?

Dialog Context

An **ideal** response should help the user to reach their **goal efficiently**







Find me something fun to do around Cambridge area in daytime!

Retrieved Knowledge

There are plenty of museums to visit around Cambridge. If you love hiking, you can enjoy the trails alongside the river. Some of my friends like to go the centre of the town and catch a movie.

> Many prefer to visit museums. You can do hiking around the river if you love nature. Or you can watch a movie. Which one do you prefer?

Dialog Context

But such knowledge is mostly **absent** during **training**

but available online abundantly







Dialog Context

A workaround is to get them from external corpora ... and **inject** them into the initial response

Many prefer to visit museums. You can do hiking around the river if you love nature. Or you can watch a movie. Which one do you prefer?



Final Response



Without any supervision (no training data for grounding), inject post-hoc knowledge into dialog responses generated by existing dialog models





Achieving Conversational Goals with Unsupervised Post-hoc Knowledge Injection

Bodhisattwa Prasad Majumder, Harsh Jhamtani, Taylor Berg-Kirkpatrick, Julian McAuley UC San Diego, Carnegie Mellon University

ACL 2022







Post-hoc Knowledge Injection in Generated Dialog POK





Knowledge Injection







Post-hoc Knowledge Retrieval

Find me something fun to do around Cambridge area in daytime!

parametric KB **Prompting an LM** with keywords from dialog history and initial response:

its university and university tour. Cambridge is famous for

Query

Find me something fun to do around Cambridge area in daytime!

You can go for a movie. Is there anything else that you prefer?

There are plenty of museums to visit around Cambridge. If you love hiking, you can enjoy the trails alongside the river. Some of my friends like to go the centre of the town and catch a movie.



non-parametric KB

Using cosine similarity on tf-idf representations





There are plenty of museums to visit around Cambridge. If you love hiking, you can enjoy the trails alongside the river...

There are plenty of art museums around Cambridge. Day tours are the best.

An interesting hour spent studying all sorts of scientific instruments in the Museum of the History of Science...

*PMI probabilities are calculated using an LM (e.g. GPT2)

[Padmakumar and He, 2021]

Relevance: PMI (knowledge *i*, history)

Redundancy: PMI (knowledge *i*, knowledge *j*)



22

There are plenty of museums to visit around Cambridge. If you love hiking, you can enjoy the trails alongside the river...

There are plenty of art museums around Cambridge. Day tours are the best.

An interesting hour spent studying all sorts of scientific instruments in the Museum of the History of Science...

- **Relevance:** PMI (knowledge *i*, history)
- **Redundancy:** PMI (knowledge *i*, knowledge *j*)
- *PMI probabilities are calculated using an LM (e.g. GPT2)
- **Determinantal Poison Process (DPP)**:
- **sampling** the most relevant and the most diverse subset [Kulesza and Taskar, 2011]



23

There are plenty of museums to visit around Cambridge. If you love hiking, you can enjoy the trails alongside the river...

There are plenty of art museums around Cambridge. Day tours are the best.

An interesting hour spent studying all sorts of scientific instruments in the Museum of the History of Science...

Relev Redu

*PMI p

Deter

sampling ti

Greedy trade-off: Select most relevant knowledge snippet - Select the next knowledge snippet that maximizes the diversity







There are plenty of museums to visit around Cambridge. If you love hiking, you can enjoy the trails alongside the river...

There are plenty of art museums around Cambridge. Day tours are the best.

An interesting hour spent studying all sorts of scientific instruments in the Museum of the History of Science...

- •
- \bullet

% of documents Before 69 83

Parametric source: GPT2

Factual

Non-parametric source: Yelp



Factual



25

Post-hoc Knowledge Injection in Generated Dialog POKI



Posthoc

Knowledge Injection







Post-hoc Knowledge Injection



- **Forward pass** for dialog model fluency
- **Backward pass** to ensure
 - modified response is as close to as the knowledge snippet — *fidelity* 2. modified response still entails with dialog history — **entailment**

There are plenty of museums to visit around Cambridge. If you love hiking, you can enjoy the trails alongside the river. Some of my friends like to go the centre of the town and catch a movie.

[Dathathri et al., 2020; Qin et al., 2020] After few iterations

Many prefer to visit museums. You can do hiking around the river if you love nature. Or you can watch a movie. Which one do you prefer?



Application Scenarios

Goal-oriented Dialog

MultiWOZ

- multi-domains

 defined conversational goals such as restaurant booking

E

Knowledge-grounded Dialog

Wizard-of-Wikipedia

- topical

- conversational goal is to have a longer informative session



28

POKI talks to a human

Final Response =	Ini
I need a place to eat that is cheap.	
 Do you have a location preference? have several options for inexpensive 	IOCa I O
restaurants.	

itial Response + Retrieved Knowledge

Do you have a ation preference? have several ptions for you.

None



29

POKI talks to a human

Final Response=Initial Response+Retrieved Knowledge

I need a place to eat that is cheap.

Do you have a location preference?
 I have several options for inexpensive restaurants.



I think you can go with Asian cuisines such as Chinese or Indian cuisines. They are value for money since they are known for their great taste. I can book a Chinese or Indian restaurant near centre for you.

Do you have a location preference? I have several options for you.

None

Asian cuisines such as Chinese or Indian cuisines are inexpensive. They are value for money since they are known for their great taste.

sults



POKI talks to a human

Final Response

Solution : Do you have any restaurants that serve English food?

English restaurants are mainly for fine dining. Indian food is also very popular in the centre area. Many people who like English food also enjoy Indian food.

👧 : Awesome. I can go with Indian food.

Additional knowledge navigates user to their goal

Initial Response **Retrieved Knowledge** +

English restaurants are mainly for fine dining.

Indian food is very popular in the centre area, Cambridge. I know many friends who like English food also enjoy Indian food. Indian food is great value for money.

User Study for Effectiveness





32

Injecting other types of knowledge

Post-hoc Knowledge Injection Majumder et al. **ACL** 2021



Narratives, post-hoc

Persona-based Commonsense Majumder et al. **EMNLP** 2020



Commonsense Inference Graphs, ante-hoc

33

Knowledge Acquisition + Injection





I am an animal activist.

Why do you work with animals?

To make a difference.

understanding the world

post-hoc



I went to a dog shelter last week. I was happy seeing an injured dog recovering well.

Tell me some experiences you had.

experience





Is there a good animal shelter around LocationX?

Yes, Place1 and Place2 are open most days in a week.

up-to-date knowledge

- On the fly knowledge acquisition
 - Textual knowledge
 - Narratives
 - Structured commonsense
- Ante- and post-hoc methods
- Promotes success in achieving conversational goals
- Bridges the knowledge gap in existing dialog/language models



Landscaping Knowledge Grounding; What's next?



Hi 'Lexiii, are you human?



Persona Discovery

Dialog State Tracking

Knowledge-Grounding

Clarification



Majumder et al. **EMNLP** 2020b

Majumder et al. **NAACL** 2021



Hmm, why should I trust you?





Majumder et al. **EMNLP** 2020a

Expecting Anthropomorphism in Al

[Watson, 2019]

Attribution in Dialog Models [Dziri et al., 2022]









Next-generation Al





Interactions


Natural Language Explanations (NLEs)



You use detergent to dye your hair

False **because** detergent can clean but cannot dye either your hair or clothes.

explaining with commonsense



Natural Language Explanations (NLEs)



Q: how does [person2] feel about what[person1] is telling him?

- NLE should be **plausible** and consistent to the input [Marasovic' et al., 2021]
- [Wiegreffe et al., 2021]
- NLE should be grounded into world knowledge [Camburu et al., 2020]

A: He's concerned and a little upset

He is in shock thinking something bad is about to happen.

abstractive

• NLE should be accurate and faithful to explain the prediction



Predictive Task

A neural predictive model is employed to solve task. For example: Natural Language Inference (NLI)

premise

Two men are competing in a bicycle race

hypothesis

People are riding bikes

Instance from SNLI dataset



label entailment



Natural Language Explanations (NLEs)

An NLE is a textual abstraction of the model explanation.

[Camburu et al., 2018]

premise

Two men are competing in a bicycle race

hypothesis

People are riding bikes

label entailment Competing in a bicycle race requires men riding bikes



Background Knowledge

A model **believes** in a set of background knowledge given input.

premise

Two men are competing in a bicycle race

hypothesis

People are riding bikes

Where do we get this knowledge?

- bicycle race requires bikes
- race requires riding bikes
- bicycle race needs helmet
- men are people

label entailment

Competing in a bicycle race requires men riding bikes



Rationale-induced Knowledge

a significant indicator of a model's prediction.

[Lei et al., 2016; Bastings et al., 2019]

premise

Two men are competing in a bicycle race

hypothesis

People are riding bikes

*tokens for language or super-pixels for images



From the predictive parts of the input

A rationale is a sufficient and minimal part* of the input that is





People

riding bikes

bicycle race requires bikes

- bicycle race race requires riding bikes
 - bicycle race needs helmet - men are people

label entailment

Competing in a bicycle race requires men riding bikes



Self-rationalization + Knowledge Grounding

Jointly producing prediction + explanation





End-to-end Connect rationales and NLEs Knowledge grounding







Knowledge-Grounded Self-Rationalization via Extractive and Natural Language Explanations

<u>Bodhisattwa Prasad Majumder</u>, Oana-Maria Camburu, Thomas Lukasiewicz, Julian McAuley UC San Diego, University of Oxford, TU Wien Informatics

ICML 2022





Rationale + Knowledge + NLE = RExC



Grounding

Rationales are responsible for relevant knowledge retrieval Knowledge (latent) selection acts as a **soft bottleneck** RExC is a **self-rationalizing** model that produces NLE and task output



(Latent *set*-of-thoughts)







m	ise
	130

Two men are competing label in a bicycle race entailment People are riding bikes

A: Coffee stimulates people **B:** Coffee depresses people

Q: Where does a wild bird usually live?

A: a) cage, b) sky, c) countryside, d) desert, e) windowsill

Hypothesis: Some tennis players pose

label B is invalid

> label sky

label entailment



Q: What is the place?

label They are in a hospital room



Plausibility via BertScore





RExC outperforms all previous SOTA for NLE quality/plausibility



Results



Rationale and Selected Knowledge individually contribute to performance

Results

Human Evaluation via e-Vil Scores



Prev. SOTA

All RExC versions are highly rated by human users



Results

Zero-shot RExC



Results

What if we don't have gold NLE during training?









RExC-ZS is at par or even better than a supervised SOTA model

Results

Zero-shot RExC

Human evaluation

RExC-ZS



RExC closing Performance-Explainability Gap



RExC is task SOTA among models with explanations, often outperforms all-time SOTA (mostly black-box)

Results

[Dalvi et al., 2022; Camburu et al., 2018; Narang et al., 2020]



Self-rationalization + Knowledge Grounding

RExC

Q: Where are [person2] and [person3]?



A: They are in a hospital room

NLE: There are hospital beds and nurses in the room

Rationale:



Selected **Knowledge:** Hospital room has hospital beds Hospital has nurses

worse

MIT Technology Review

ARTIFICIAL INTELLIGENCE



Knowledge reduces ambiguity

Why asking an AI to explain itself can make things

Attributing explanations

+ **Emerging properties**







Emerging properties



Faithfulness in Language Explanations Xie, McAuley, Majumder **Preprint** 2022





Knowledge-grounding improves this









Task Performance Plausibility Robustness



Factuality



Large Language Models are Zero-Shot Reasoners

Takeshi Kojima The University of Tokyo t.kojima@weblab.t.u-tokyo.ac.jp

Machel Reid Google Research*

Still models are **confused**, self-**conflicting**

LARGE LANGUAGE MODELS ARE NOT ZERO-SHOT COMMUNICATORS

Laura Ruis, Akbir Khan, Stella Biderman², Sara Hooker, Tim Rocktäschel, Edward Grefenstette¹⁵ ¹ University College London, ² EleutherAI, ³ Booz Allen Hamilton, ⁴ Cohere for AI, ⁵ Cohere

Landscaping XAI; What's next?

Shixiang Shane Gu Google Research, Brain Team

Yutaka Matsuo The University of Tokyo

Yusuke Iwasawa The University of Tokyo

Chain of Thought Prompting

Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9. The answer is 9. 🗸

Relevant, Trustworthy, and Adaptive Al





Subjectivity (not) in Al

We Teach A.I. Systems Everything, Including Our Biases

Researchers say computer systems are learning from lots and lots of digitized books and news articles that could bake old attitudes into new technology.

+ subjectivity + individual preferences + culture

. . .

MIT SLOAN EXPERTS | ARTIFICIAL INTELLIGENCE

<u>'Human-Centered Al': How can</u> <u>the technology industry fight bias</u> <u>in machines and people?</u>





Human-in-the-loop is the future

- [Klie et al., 2020]
- [Lee et al., 2020]
- [Brantley et al., 2020]
- [Simpson et al., 2019]
- [Dasgupta et al., 2019]
- [Radlinski et al., 2019]
- [Smith-Renner et al., 2020]



input

Angela Lindvall is a model and she has represented almost every major fashion brand



Measuring Bias in Models

predicting profession



prediction model

bias in data correlation with gender

is my model biased? probably yes

How to measure it? representations

pre-trained gender classifier



gender label female





input

Angela Lindvall is a model and she has represented almost every major fashion brand



Measuring Bias in Models

predicting profession





input

Angela Lindvall is a model and she has represented almost every major fashion brand



Measuring Bias in Models

predicting profession





Measuring Bias in Rationales

biased (original) model

Angela Lindvall is a model and she has represented almost every r fashion brand maio

> prediction \checkmark model





by adding back minimally biased tokens

Angela Lindvall is a model she has represented almost every major fashion brand

*Rationales are **significant indicators** from input for a model's prediction

X

debiased (adv) model

Angela Lindvall is a model and

she has represented almost every major fashion brand

prediction fashion designer

prediction model





Measuring Bias in Rationales

biased (original) model

Angela Lindvall is a model and she has represented almost every r fashion brand maio

> prediction \checkmark model





by adding back minimally biased tokens

Angela Lindvall is a model she has represented almost every major fashion brand

*Rationales are **significant indicators** from input for a model's prediction

debiased (adv) model

Angela Lindvall is a model and

she has represented almost every major fashion brand

prediction fashion designer

prediction model

 \checkmark



X









InterFair: Debiasing with Natural Language **Feedback for Fair Interpretable Predictions**

Bodhisattwa Prasad Majumder*, Zexue He*, Julian McAuley UC San Diego

InterNLP @ NeurIPS 2022, In Progress; EMNLP 2022







Debiasing by Intervening Explanations



Token contribution for predicting gender





Debiasing by Intervening Explanations



Token contribution for predicting gender

balancing task performance and bias mitigation

penalize task contribution

if bias contribution is high else no penalty



Debiasing by Intervening Explanations

Rerank task-important tokens based on bias importance

Complex penalty functions

training

interactions

Jser interventions

training-time

test-time

balancing task performance and bias mitigation

penalize task contribution

if bias contribution is high else no penalty



Even after training

	Input		Predictio
+ • • • 10101 0101	Angela Lindvall is a model and she has represented almost every major fashion brand	(frozen) Classifier	Model

on

 \checkmark

Task Rationales

Angela Lindvall is a model and she has represented almost every major fashion brand

Bias Rationales

Angela Lindvall is a model and she has represented almost every major fashion brand



still problematic



Bias Classifier is not *perfect*, neither is the data



InterFair = Training + Interactions

Input

Prediction



Angela Lindvall is a model and she has represented almost every major fashion brand

(frozen) Classifier

Task Rationales

Bias Rationales



Angela Lindvall is a model and she has represented almost every major fashion brand

Angela Lindvall is a <mark>model</mark> and she has represented almost every major fashion brand



Debiasing is *subjective* to a user Can be better at *teaching* the model

InterFair = Training + Interactions

	Input			Predict
• • • • • • 1010 0101	Angela Lindvall is a mod she has represented all every major fashion br	el and most and	(frozen) Classifier	Mode
	Don't use w: model Don't use any name		Rei	nstating E
• •	Angela Lindvall is a mod she has represented all every major fashion br	el and most and	(frozen) Classifier	Fashic Design

tion

Task Rationales

Bias Rationales



Angela Lindvall is a model and she has represented almost every major fashion brand

Angela Lindvall is a model and she has represented almost every major fashion brand

Bias Definition



Angela Lindvall is a model and she has <mark>represented</mark> almost every major fashion brand

Angela Lindvall is a model and she has represented almost every major fashion brand



InterFair = Training + Interactions

	Input		Predic
+ <mark>20</mark> + 10101 0101	Angela Lindvall is a model and she has represented almost every major fashion brand	d (frozen) Classifier	Mode
	Don't use w: model Don't use any name	Rei	instating I
+ • • • 10 10 01 01	Angela Lindvall is a model and she has represented almost every major fashion brand	d (frozen) Classifier	Fashie Desigr
	Consider using w: model Don't use any name	Re	defining l
+ • • 10 10 01 01	Angela Lindvall is a model and she has represented almost every major fashion brand	d (frozen) Classifier	Mode

tion

Task Rationales

Bias Rationales



Angela Lindvall is a model and she has represented almost every major fashion brand Angela Lindvall is a model and she has represented almost every major fashion brand

Bias Definition



Angela Lindvall is a model and she has represented almost every major fashion brand Angela Lindvall is a model and she has represented almost every major fashion brand

Bias Definition



Angela Lindvall is a model and she has represented almost every major fashion brand Angela Lindvall is a model and she has represented almost every major fashion brand





Using User Feedback

Parsing as a sequence labeling task Assign High/Low/NA for each input token given bias and feedback. [Input] Angela Lindvall is a model and she represented (...) [Bias] Gender

[Feedback] Angela Lindvall is a woman's name [Parse] High, High, NA, NA, NA, NA, NA, NA, NA (...)

Π LL 0

I. Parse **Feedback** on Bias

II. Update **Bias** Rationales

III. Update Task Rationales

IID



I. Heuristic

II. Gradient based

Similar to dialog works [Majumder et al., 2022]











Results




Results



Results

Explanations + Interactions



Don't use w: model Don't use any name



Consider using w: model Don't use any name











No-memory \rightarrow Generalization with user feedback?



Generalizing with User Feedback



Model Editing

Conversational Recommendation Li, **Majumder** et al. **RecSys** 2022

Selected in Highlights of ACM RecSys'22



Memory-based Architectures

Conversational Teaching Majumder et al. Aristo 2022



Landscaping Interactive AI; What's next?





writes feedback on the outpo	
Write a summary of the given text: Once upon a time,	→(i
Condition the langua and feedback to general	ge m te mu
Write a summary of the given text: Once upon a time,	
To summarize,	
Ochoose the refinement with the feedback.	ent wi
In essence, In a nutshell,	he summary should
The gist of	e mo
outputs.	

Memory-based architectures

Additional fine-tuning



post-hoc

Relevant, Trustworthy, and Adaptive Al





+ future directions

Interactions



Next-generation Al











Relevant, Trustworthy, and Adaptive Al



Training-time Knowledge Augmentation

Post-hoc Knowledge Injection



Personalized Knowledge

Domain-specific Knowledge





Knowledge Grounding in Explanations

Factuality, Robustness in Explanations

Reasoning in Explanations

Personalized Reasoning

Ontology for Reasoning

Critiquable Explanations and Synthesis

Post-hoc Interactive feedback

Generalizing User Feedback

Budgeted Interactions

Never-ending Learning

Interactions

Next-generation Al









Acknowledgements



Qualcom fellowship











Acknowledgements













We Informatics

THE UNIVERSITY OF BRITISH COLUMBIA



Microsoft





Machine Learning





Want to figure out what a good prompt might be to create new images like an existing one? The CLIP Interrogator is here to get you answers!



Output



CLIP Interrogator







@mbodhisattwa



majumderb.com

