Discussion #2

Student-led discussions

Ranjay Krishna | ranjay@cs.washington.edu

“Power to the People: The Role of Humans in Interactive Machine Learning”
&
“An Interaction Framework for Human-Machine Relationship in NLP”

Discussion leader

Power to the People: The Role of Humans in Interactive Machine Learning

Summary

- Generally, ML engineers design the system, then take feedback from users
- Process is generally slow, frustrating for both sides:
  - They got this insight from work between ML practitioners and biochemists -- having users interactively build the ML system led to faster development.
- The paper asks: what do we observe when we try to use IML approaches?
  - Users are not Oracles that we can harass with questions
  - People provide more than just labels: they can also provide suggestions on features to consider/alternate reps
  - People want to demonstrate how learners should behave
  - (and so on)

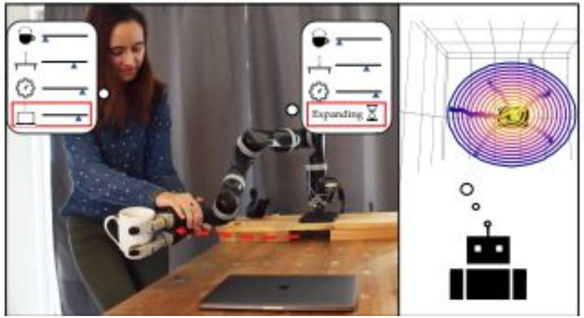
Power to the People: The Role of Humans in Interactive Machine Learning

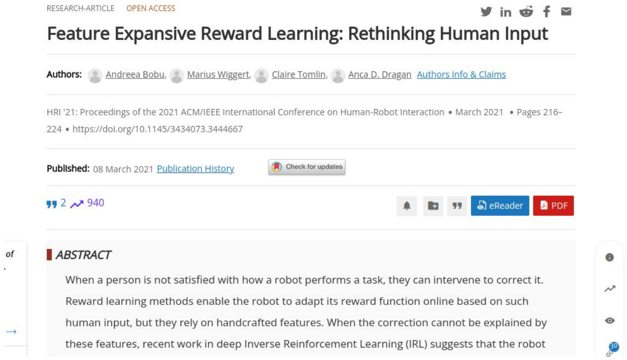
Discussion points

- I like that the authors:
  - Discuss idea of having users interactively build the ML systems
  - Discuss the diverse set of domains the authors said this could be applied to
  - Admit more work to be done here
  - Provide good high-level take away points
- I wish they:
  - Talked about why it is difficult to involve people (expenses, recruiting delays..)
  - Does it scale, or work well in the real world?
  - Talked about whether these methods be used where humans are not experts? (noise removal)

Power to the People: The Role of Humans in Interactive Machine Learning.

- The idea of having systems “pick-up” stuff from human feedback has been explored in inverse reinforcement learning:
  - Work in this paper addresses how a robot can learn a human’s preferences for Movement.
  - They address how:
    - Weights for learned features are optimized
    - How the system learns new features -- if the correction shown by the user does not “align” with any of the features that the system knows of, this is probably a new feature and I should ask the human for more demonstrations To learn this feature.





Feature Expansive Reward Learning: Rethinking Human Input

Authors: Andreea Bobu, Marius Wiggert, Claire Tomlin, Anca D. Dragan Authors Info & Claims

HRI '21: Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction • March 2021 • Pages 216–224 • https://doi.org/10.1145/3434073.3444667

Published: 08 March 2021 Publication History

2 940

ABSTRACT

When a person is not satisfied with how a robot performs a task, they can intervene to correct it. Reward learning methods enable the robot to adapt its reward function online based on such human input, but they rely on handcrafted features. When the correction cannot be explained by these features, recent work in deep Inverse Reinforcement Learning (IRL) suggests that the robot

An Interaction Framework for Human-Machine Relationships in NLP

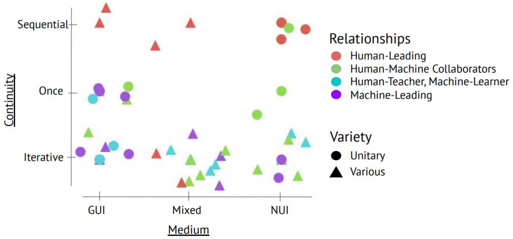
Summary

- Systematic survey on existing human-machine interactions in NLP
- Framework
    - Properties: How does human-machine interaction happen in NLP?
        - Continuity, Variety of Interaction Actions, Medium of Interactions
    - Relationships: How do humans and machines interact with each other in NLP?
        - Human-Teacher and Machine-Learner
        - Machine-Leading
        - Human-Leading
        - Human-Machine Collaborators
- The framework could be used to guide future interaction design

An Interaction Framework for Human-Machine Relationships in NLP

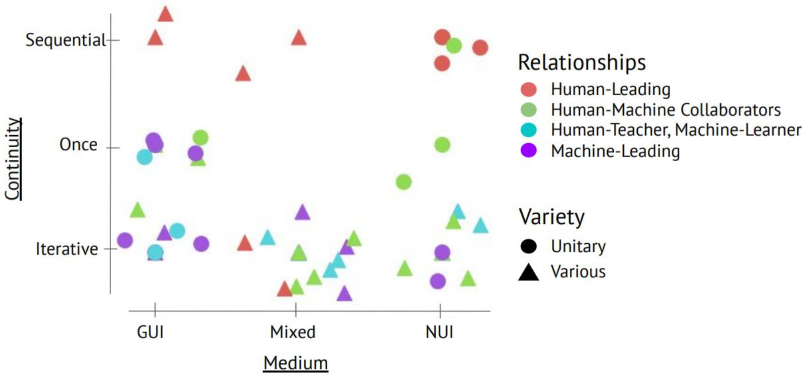
Discussion points

- **I like...**
  - The initiative taken to survey and propose the framework
  - Clarify the nuances in different interactions with examples
  - Visualization of where the existing work lies
- **I wish...**
  - The paper includes more work than last 2 years
  - The paper discuss clearer guidelines, e.g., what tasks → what interactions
  - The paper discuss the *type of feedback* between humans and machines



An Interaction Framework for Human-Machine Relationships in NLP

- **How would you use the framework to help with your research?**
  - E.g., Using LLMs
    - Prompting LLMs with few-shot examples lies in (Human-Leading, NUI, Once)
    - We can make the relationships different?
      - Human-Teacher, Machine Learner → Further train LLMs?
      - Machine Leading → Ask user to denoise post-hoc
      - Human-Machine Collaborators → LLMs learn from user-denoised examples



Scientific Peer Reviewer (Advocate)

The Role of Humans in Interactive Machine Learning

- **Main Idea**: This paper states the **importance of user studies** in interactive machine learning (before the era of DL) and demonstrate how it can result in better user experiences and more effective learning systems.

Expertise

Very Knowledgeable

Originality

Low originality

Significance

Very high significance

Rigor

Medium rigor

Recommendation

I recommend Accept with Minor Revisions

Guidelines for Human-AI Interaction

The Role of Humans in Interactive Machine Learning

- **Main Idea**: This paper states the **importance of user studies** in interactive machine learning (before the era of DL) and demonstrate how it can result in better user experiences and more effective learning systems.
- **Body**: This paper achieves the above goal by surveying and presenting existing works as case studies in three different directions: **Interactive ML (I-ML)**, **user interaction in I-ML** and **novel interface in I-ML**.

Expertise

Very Knowledgeable

Originality

Low originality

Significance

Very high significance

Rigor

Medium rigor

Recommendation

I recommend Accept with Minor Revisions

Guidelines for Human-AI Interaction

The Role of Humans in Interactive Machine Learning

- **Strength**:
    - This paper clearly states its purpose and successfully demonstrate its idea into 3 directions.
    - This paper survey 30+ publications and discuss 15+ of them in detail as case studies.
    - The taxonomy of the paper proposed to cluster the methods are clear, easy to follow and thorough.
    - This paper discusses the potential challenges / potential improvements based on the existing literature and gives insightful suggestions.

Expertise

Very Knowledgeable

Originality

Low originality

Significance

Very high significance

Rigor

Medium rigor

Recommendation

I recommend Accept with Minor Revisions

Guidelines for Human-AI Interaction

The Role of Humans in Interactive Machine Learning

- **Weakness**:
  - The originality of this paper is not very strong (since it’s more like a survey / review paper).
  - This paper could discuss more about the connections between the three main clusters of existing literatures.
- **Other factors**:
  - Test of times: (I’m not sure) is case study a good way to summarize and reflect the time-insensitive contribution of the existing literature ?

Expertise

Very Knowledgeable

Originality

Low originality

Significance

Very high significance

Rigor

Medium rigor

Recommendation

I recommend Accept with Minor Revisions

Guidelines for Human-AI Interaction

User or Labor: An Interaction Framework for Human-Machine Relationships in NLP

Review form IJCAI-ECAI 2022

https://ijcai-22.org/wp-content/uploads/2022/01/Review-form-IJCAI-ECAI-2022-2.pdf

2. Main strengths of paper

Novelty: Summary papers done before, but not for categorizing interaction types

Soundness: Not many technical details, but categories clearly defined

Significance: Limited to 33 papers, but could inform future work within subfield

Relevance to AI: Very relevant to AI, especially interactive Human-AI systems

Clarity of exposition: Clear writing and explanatory examples

Reproducibility: Paper categorization transparent, although reasoning often not

Improvements

3. What opportunities are there to improve the paper?

- *"Wang et al. (2021) summarized recent human in-the-loop NLP work based on their tasks, goals, human interactions, and feedback learning methods."*
  - Could improve by showing the interactions between these types of classifications and the classifications mentioned in this paper

4. What pressing questions do you have for the authors in the rebuttal ?

- Can you include the reasoning behind how you categorized each paper in the appendix?

Assessment

5/6. Overall assessment / Justify:

Clear Accept. Not as novel and groundbreaking to get a stronger accept, but a useful contribution to the literature

7. Reproducibility: Convincing

8/9. Ethics issues: Not large because it is a review of existing work

10. Alignment with my expertise: Knowledgeable

11. Confidence in evaluation: Confident

Scientific Peer Reviewer (Skeptic)

Power to the People: The Role of Humans in Interactive Machine Learning

Main Contribution:

- Case studies across many CS fields that highlight the role of interactive machine-learning systems and demonstrate the feasibility of richer interactions with users
- Potential future direction to develop Interactive Machine Learning systems

Expertise

Knowledgeable

Originality

Low originality

Significance

High significance

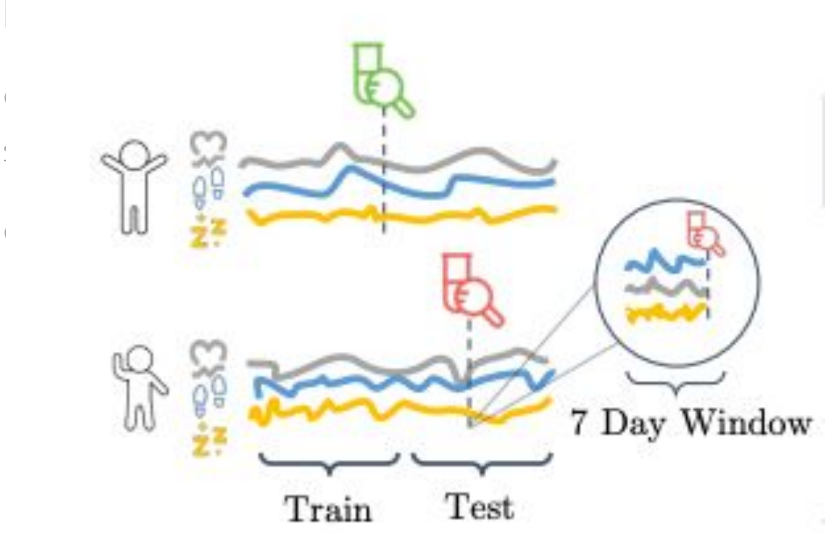
Rigor

Low rigor

Recommendation

I recommend Revise and Resubmit

Pros and Cons



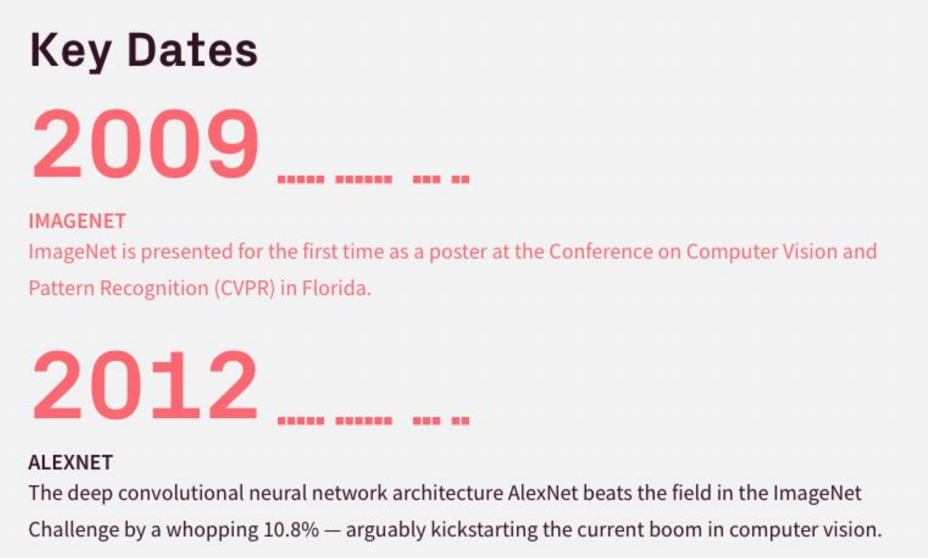
Cons

Did not discuss limitations or potentially counter viewpoints. For example, how would interactive machine learning do if the task is not something a human can easily do such as classifying noisy signals from mobile health sensing data?

(Skeptic reviewer hat on) The work does not seem significant since many models do not perform well on human labeled tasks yet. ImageNet is just human labels and we are starting to see success there

Similar overview paper has been published before. Any value of rehashing to same info as a full paper? Saleema Amershi. 2011. Designing for effective end-user interaction with machine learning. In Proceedings of the 24th annual ACM symposium adjunct on User interface software and technology (UIST '11 Adjunct).

Pros and Cons



Key Dates

2009

IMAGENET

ImageNet is presented for the first time as a poster at the Conference on Computer Vision and Pattern Recognition (CVPR) in Florida.

2012

ALEXNET

The deep convolutional neural network architecture AlexNet beats the field in the ImageNet Challenge by a whopping 10.8% — arguably kickstarting the current boom in computer vision.

Cons

Did not discuss limitations or potentially counter viewpoints. For example, how would interactive machine learning do if the task is not something a human can easily do such as classifying noisy signals from mobile health sensing data?

(Skeptic reviewer hat on) The work does not seem significant since many models do not perform well on human labeled tasks yet. ImageNet is just human labels and we are starting to see success there

Similar overview paper has been published before. Any value of rehashing to same info as a full paper? Saleema Amershi. 2011. Designing for effective end-user interaction with machine learning. In Proceedings of the 24th annual ACM symposium adjunct on User interface software and technology (UIST '11 Adjunct).

Pros and Cons

Cons

Did not discuss limitations or potentially counter viewpoints. For example, how would interactive machine learning do if the task is not something a human can easily do such as classifying noisy signals from mobile health sensing data?

(Skeptic reviewer hat on) The work does not seem significant since many models do not perform well on human labeled tasks yet. ImageNet is just human labels and we are starting to see success there

Similar overview paper has been published before. Any value of rehashing to same info as a full paper?
Saleema Amershi. 2011. Designing for effective end-user interaction with machine learning. In Proceedings of the 24th annual ACM symposium adjunct on User interface software and technology (UIST '11 Adjunct).

User or Labor: An Interaction Framework for Human-Machine Relationships in NLP

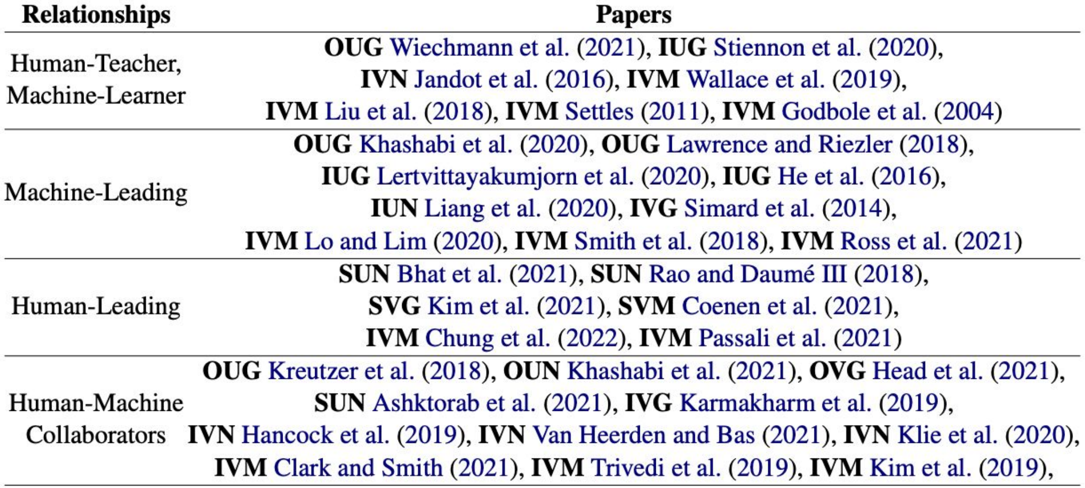
Based on ACL review form I found online

Paper Summary

- Paper surveys the last two years of NLP research for human-machine interaction and build a framework for human-machine interactions:

Summary of Strengths

- A survey like this as a necessary contribution to the field. I do find the need to understand the interaction framework of human and machine relationships important in NLP especially as the field has grown
- The framework the authors build is helpful to discuss future research in this space

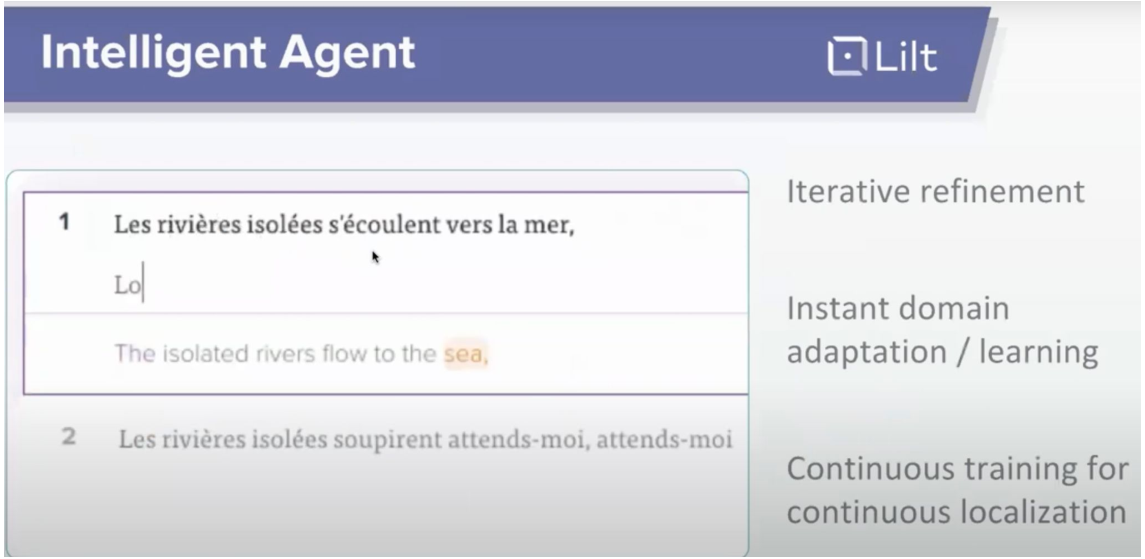


Relationships	Papers
Human-Teacher, Machine-Learner	**OUG** Wiechmann et al. (2021), **IUG** Stiennon et al. (2020), **IVN** Jandot et al. (2016), **IVM** Wallace et al. (2019), **IVM** Liu et al. (2018), **IVM** Settles (2011), **IVM** Godbole et al. (2004)
Machine-Leading	**OUG** Khashabi et al. (2020), **OUG** Lawrence and Riezler (2018), **IUG** Lertvittayakumjorn et al. (2020), **IUG** He et al. (2016), **IUN** Liang et al. (2020), **IVG** Simard et al. (2014), **IVM** Lo and Lim (2020), **IVM** Smith et al. (2018), **IVM** Ross et al. (2021)
Human-Leading	**SUN** Bhat et al. (2021), **SUN** Rao and Daumé III (2018), **SVG** Kim et al. (2021), **SVM** Coenen et al. (2021), **IVM** Chung et al. (2022), **IVM** Passali et al. (2021)
Human-Machine Collaborators	**OUG** Kreutzer et al. (2018), **OUN** Khashabi et al. (2021), **OVG** Head et al. (2021), **SUN** Ashktorab et al. (2021), **IVG** Karmakharm et al. (2019), **IVN** Hancock et al. (2019), **IVN** Van Heerden and Bas (2021), **IVN** Klie et al. (2020), **IVM** Clark and Smith (2021), **IVM** Trivedi et al. (2019), **IVM** Kim et al. (2019),

Table 1: Human-Machine Relationships Mapping Interaction Properties: The properties of the interaction in each paper are coded by the first letter of their three interaction properties: **O/S/I** represents **One-time/Sequential/Iterative**. **U/V** represents **Unitary/Various**. **G/N/M** represents **GUI/NUI/MUI**.

Summary of Weaknesses

- The authors ask how does human-interaction happen in NLP but I am not sure if the survey format matches this RQ.
- The author's do not justify well why it was only the last two years and so I am not convinced it is a thorough survey.
  - For example, Jeff Heer's work in Adaptive language translation covers "tools that interleave human & machine translation" and involved published works in 2013-2015
  - What about commercial products?
- Lack of implications, not super convinced of the categories



Intelligent Agent

1 Les rivières isolées s'écoulent vers la mer,

Lo

The isolated rivers flow to the sea,

2 Les rivières isolées soupirent attends-moi, attends-moi

Iterative refinement

Instant domain adaptation / learning

Continuous training for continuous localization

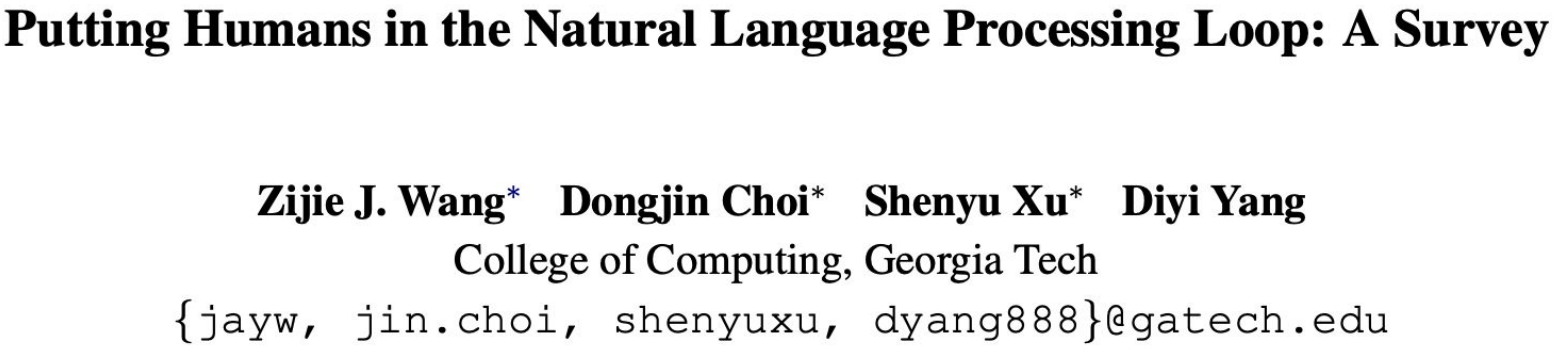
Overall Assessment

- 2 = Revisions Needed: This paper has some merit, but also significant flaws, and needs work before it would be of interest to the community

Archaeologist (Zoom)

One older paper cited within the current paper

- ***Wang et al. (2021)*** *summarized recent human in-the-loop NLP work based on their tasks, goals, human interactions, and feedback learning methods. According to* ***Wang et al. (2021),*** *a good human in-the-loop NLP system must clearly communicate to humans what the model requires, provide user friendly interfaces for collecting feedback, and effectively learn from it.*



Putting Humans in the Natural Language Processing Loop: A Survey

Zijie J. Wang* Dongjin Choi* Shenyu Xu* Diyi Yang
College of Computing, Georgia Tech
{jayw, jin.choi, shenyuxu, dyang888}@gatech.edu

Putting Humans in the Natural Language Processing Loop: A Survey

Categorizes surveyed HITL paradigms:

- Text Classification
- Parsing and Entity Linking
- Topic Modeling
- Summarization and Machine Translation
- Dialogue and Question Answering

Discusses the mediums that users use to interact with HITL systems and different types of feedback that the system collect:

- Medium:
  - Graphical User Interface
  - NL interface

- User Feedback Types
  - Binary Feedback
  - Scaled Feedback
  - NL Feedback

Putting Humans in the Natural Language Processing Loop: A Survey

Summarizes how existing HITL NLP systems utilize different types of feedback:

- Data Augmentation: consider the feedback as a new ground truth data sample.
- Model Direct Manipulation:
  - Li et al. (2017) collect binary feedback as rewards for reinforcement learning of a dialogue agent
  - Kreutzer et al. (2018) uses a 5-point scale rating as reward function of reinforcement and bandit learning for machine translation

Comparisons between these two survey papers

- **Wang et al. (2021)** discusses different tasks, feedback types, feedback utilization methods

- **Wan et al. (2022)** discusses different interaction types, categorizes tasks into paradigms, and talked a bit about it’s impact and limitations

One newer paper that cites this current paper

Unfortunately, this paper has not yet been cited.

However, I notice that this paper hasn't spent much effort talking about how it's work can help improve future interactive NLP.

A nice follow up work could be
(1) an extension that discusses the pros and cons of each dimension (Properties of Interaction, Relationships of Human and Machine)
(2) establish some design principles for future HCI+NLP.

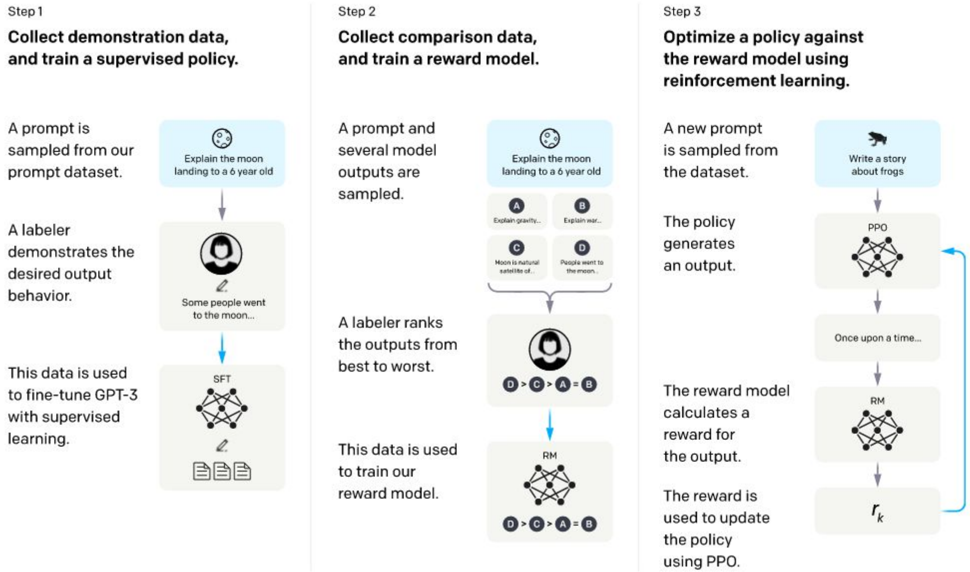
"Our goal was to define a generalizable human-machine interaction framework in NLP to explain current implementation, ***guide the design of human-machine interaction, and inspire future research in interactive NLP systems****."*

Academic Researcher

The Alignment problem

…. Aligning AI objectives to Human Expectation.

- **Traditional ML:**
  - Domain Knowledge Injection
  - Inflexible data labelling at start
- **Interactive ML (IML) Process:**
  - Users-In-the-Loop



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 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 reinforcement learning.

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.

Explain the moon landing to a 6 year old
Some people went to the moon...
SFT
Explain the moon landing to a 6 year old
A
Explain gravity...
B
Explain war...
C
Moon is natural satellite of...
D
People went to the moon...
D > C > A = B
RM
D > C > A = B
Write a story about frogs
PPO
Once upon a time...
RM
r_k

Where does new methods from RL like RL with Human Feedback (RLHF) fall?

Are they Traditional because we collect the human/expert knowledge infrequently piecewise or are they IML cause we train an RL model to approximate this feedback and use it as the "user"?

Other Questions?

- Do people/users want to make design decisions?
  - Do users want to hep define/structure the data points to collect
- What happens when humans as user/experts can’t provide feedback? maybe due to complexity of the task.
- Similar to the discussion-1 paper by Eric Horowitz. In practice, timing of queries to the user is key? How might we understand the nuances of this for different types of learning processes and tasks.
  - e.g Does the timing of netflix’s recommendation engine have to be different than the timing of RL agent in the home e.g a robot vacuum

Application proposal

"...human feedback could let us specify a specific goal more intuitively and quickly than is possible by manually hand-crafting the objective."

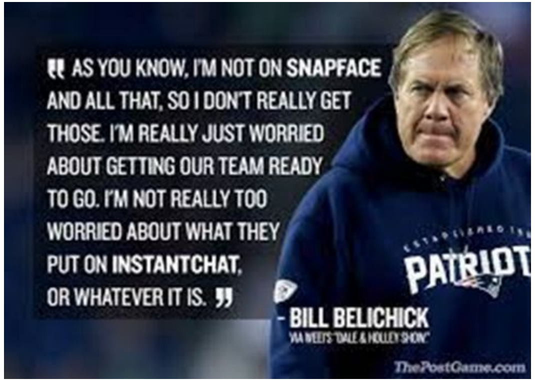
- Users
  - want to provide more data, especially in areas where the model is lacking,
  - May provide data in unstructured ways
- ML practitioners
  - want to have a less noisy and structured data
  - want to collect the data as efficiently as possible in a timely manner

Guide to ML professionals on how to engage Users in varying scenarios and across AI subfields?

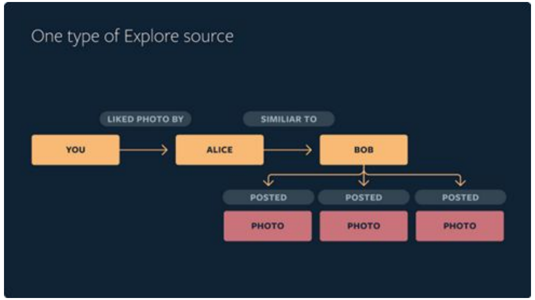
Industry Practitioner (Zoom)

“Power to the People…” – Snapface

- Feed of recommendations based on previous interactions
  - E.g., TikTok, YouTube, Snapchat / Instagram Discover
- **Differentiating feature:** Users can curate feeds by visualizing and editing the interactions that lead to recommendations.
  - Moves past “Similar to posts you interacted with”
  - Transparency and curation of recommender systems
  - ‘Explainability’


“AS YOU KNOW, I'M NOT ON SNAPFACE AND ALL THAT, SO I DON'T REALLY GET THOSE. I'M REALLY JUST WORRIED ABOUT GETTING OUR TEAM READY TO GO. I'M NOT REALLY TOO WORRIED ABOUT WHAT THEY PUT ON INSTANTCHAT, OR WHATEVER IT IS.”
- BILL BELICHICK
VIA WEEI'S 'DALE & HOLLEY SHOW'
ThePostGame.com
PATRIOTS

link



link

“Power to the People…” – Snapface

Positives:

- Transparency
- Allows for Curation + Privacy
- Underlying architectures are explainable

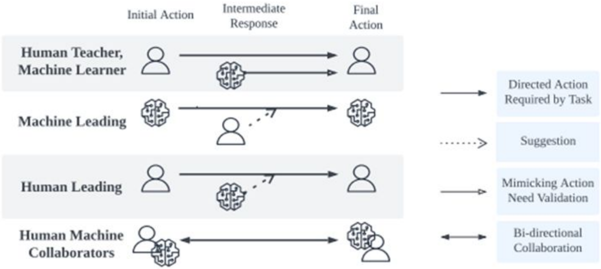
Negatives:

- Large Engineering Costs
  - Explainable and interactive AI is expensive!
- Worse recommendations
  - Performance vs. ‘Explainability’ Tradeoff

“User or Labor…” – Hoh Mechanical Turk

- Crowdsourcing Platform
- **Differentiating feature:** Splits work on along two axes:
  - **Framework:** Human-Teacher and Machine-Learner, Machine-Leading, Human-Leading, Human-Machine Collaborators
  - **Interaction:** Continuity, Variety, Medium

link


Initial Action
Intermediate Response
Final Action
Human Teacher, Machine Learner
Machine Leading
Human Leading
Human Machine Collaborators
Directed Action Required by Task
Suggestion
Mimicking Action Need Validation
Bi-directional Collaboration

link

“User or Labor…” – Hoh Mechanical Turk

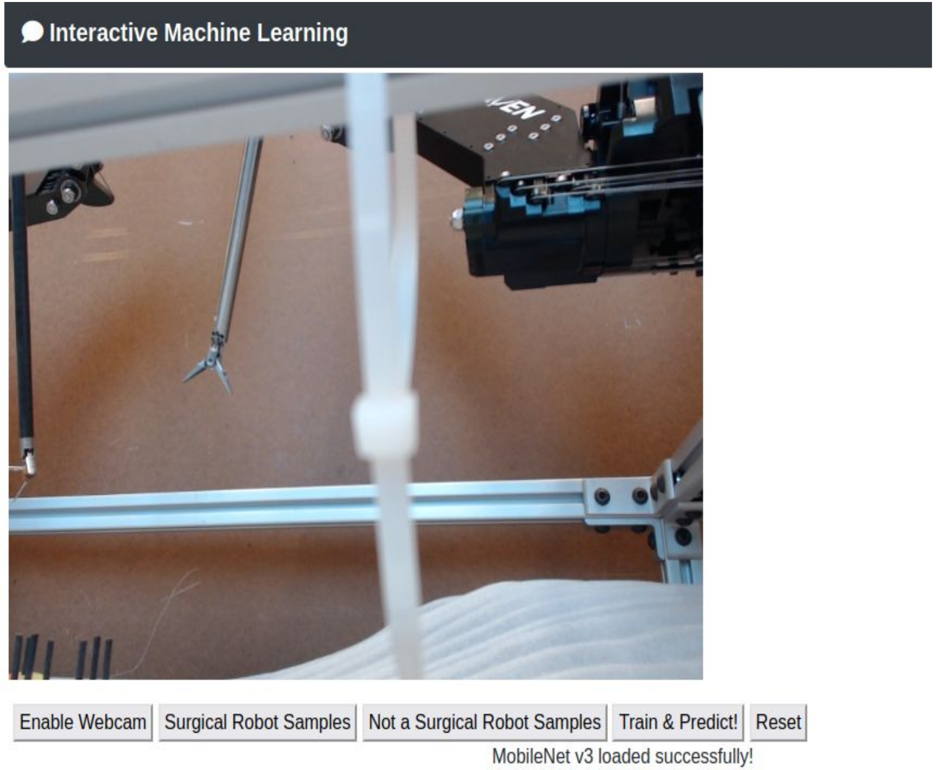
Positives:

- Allows for more diverse types of crowdsourced HCI
  - Interactions with *trained* vs. *training* models
- Drives clarity for crowd workers

Negatives:

- Enforces rigidity in interaction
  - **Question:** Do applications exist outside of these bounds? Will they exist?
- Building out support for harder interaction types
  - Support exists for interacting with *trained* AI system at scale
  - **Question:** Is it easy to parallelize *training* at scale?

Hacker

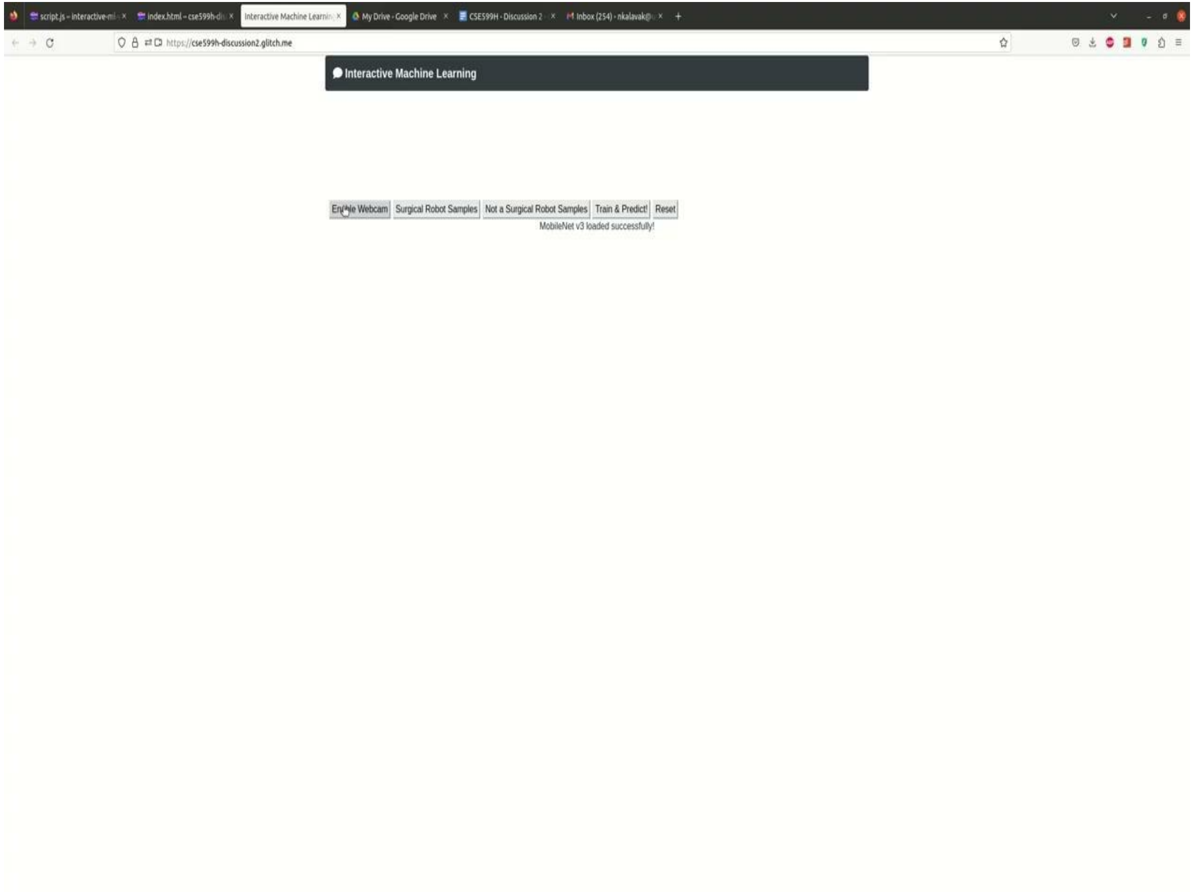


MobileNet v3 loaded successfully!

Classify if a frame has a surgical robot or not; in the browser, trained with minimal samples (and a MobileNetv3 Pre-trained model)

Loosely based on Google’s [Teachable Machine] tool

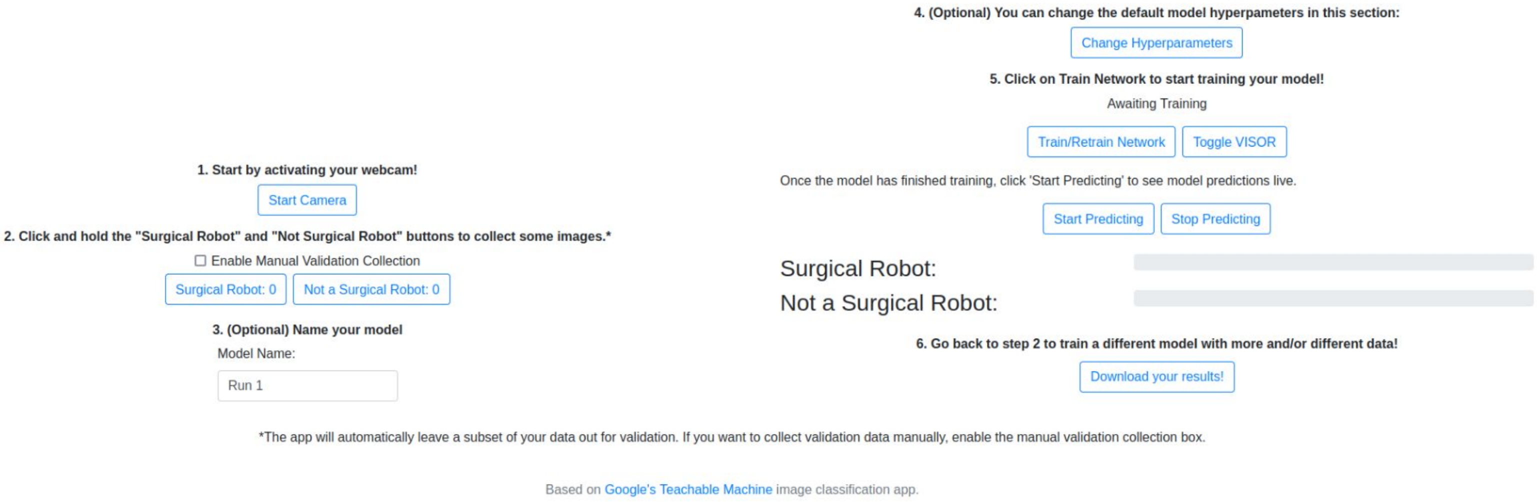
Interactive Surgical Robot Classifier (Power to the People)



MobileNet v3 loaded successfully!

Interactive Surgical Robot Classifier

Ranjay Krishna | ranjay@cs.washington.edu



1. Start by activating your webcam!

2. Click and hold the "Surgical Robot" and "Not Surgical Robot" buttons to collect some images.*

☐ Enable Manual Validation Collection

3. (Optional) Name your model

Model Name: Run 1

4. (Optional) You can change the default model hyperpameters in this section:

5. Click on Train Network to start training your model!

Awaiting Training

Once the model has finished training, click 'Start Predicting' to see model predictions live.

Surgical Robot:

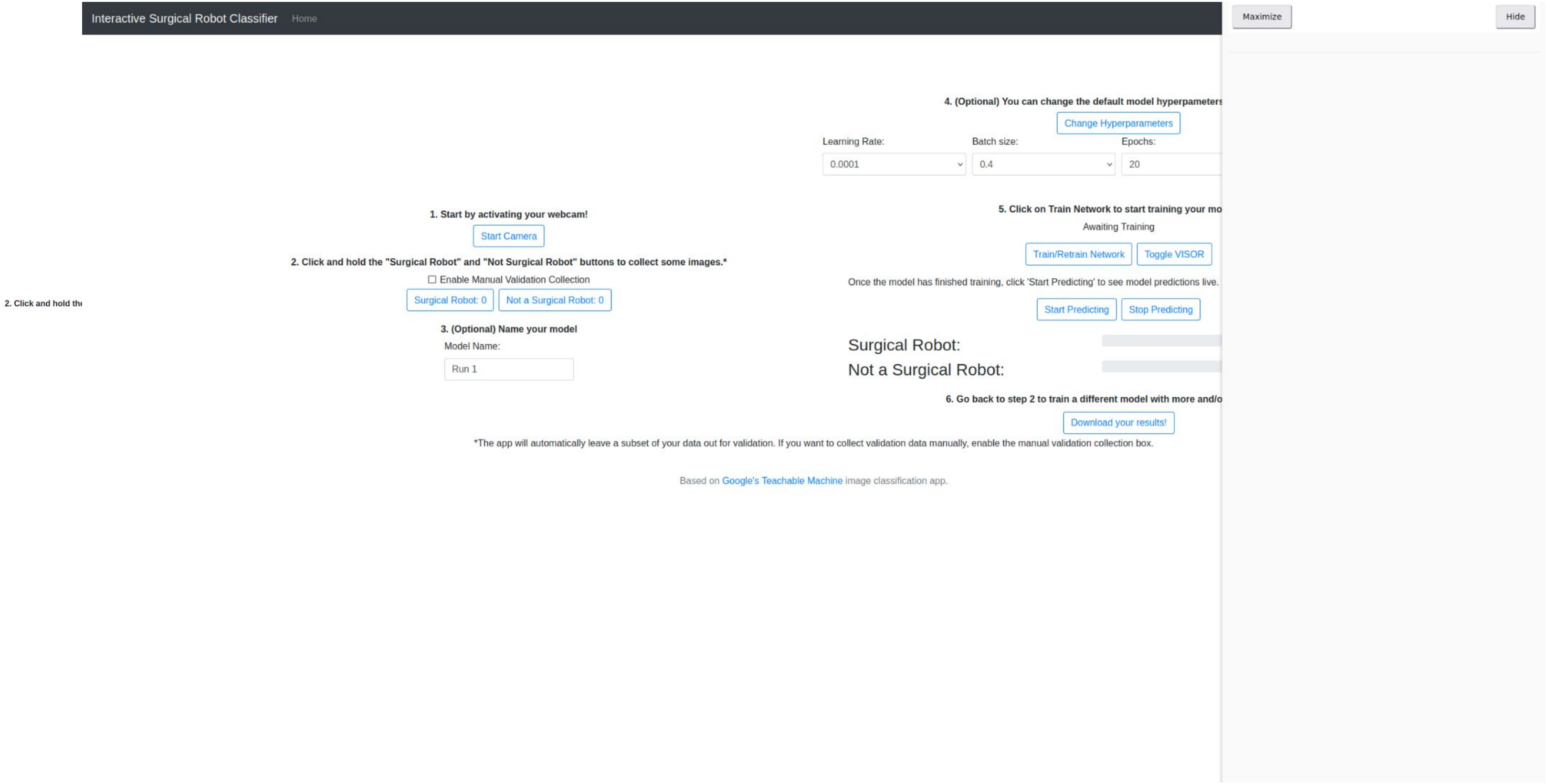
Not a Surgical Robot:

6. Go back to step 2 to train a different model with more and/or different data!

*The app will automatically leave a subset of your data out for validation. If you want to collect validation data manually, enable the manual validation collection box.

Based on Google's Teachable Machine image classification app.

Incorporate functionality to retrain models and interactive features that could be interesting (explicitly defining a validation set, visualize graphs and the ability to download a model)



4. (Optional) You can change the default model hyperpameters

Learning Rate: 0.0001

Batch size: 0.4

Epochs: 20

1. Start by activating your webcam!

2. Click and hold the "Surgical Robot" and "Not Surgical Robot" buttons to collect some images.*

☐ Enable Manual Validation Collection

2. Click and hold the

3. (Optional) Name your model

Model Name: Run 1

5. Click on Train Network to start training your mo

Awaiting Training

Once the model has finished training, click 'Start Predicting' to see model predictions live.

Surgical Robot:

Not a Surgical Robot:

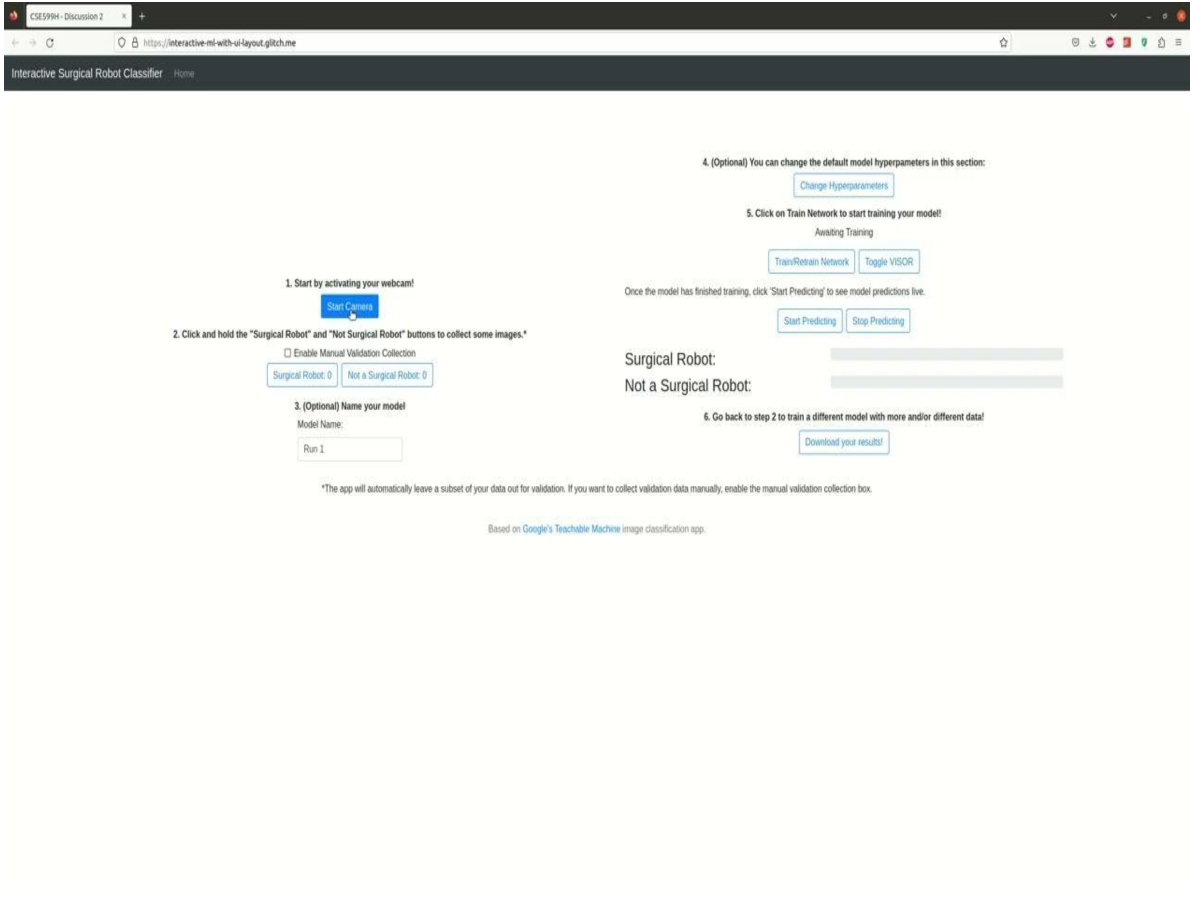
6. Go back to step 2 to train a different model with more and/o

*The app will automatically leave a subset of your data out for validation. If you want to collect validation data manually, enable the manual validation collection box.

Based on Google's Teachable Machine image classification app.

Interactive Surgical Robot Classifier

Ranjay Krishna | ranjay@cs.washington.edu



1. Start by activating your webcam!

2. Click and hold the "Surgical Robot" and "Not Surgical Robot" buttons to collect some images.*

☐ Enable Manual Validation Collection

3. (Optional) Name your model

Model Name: Run 1

4. (Optional) You can change the default model hyperparameters in this section:

5. Click on Train Network to start training your model!

Awaiting Training

Once the model has finished training, click 'Start Predicting' to see model predictions live.

Surgical Robot:

Not a Surgical Robot:

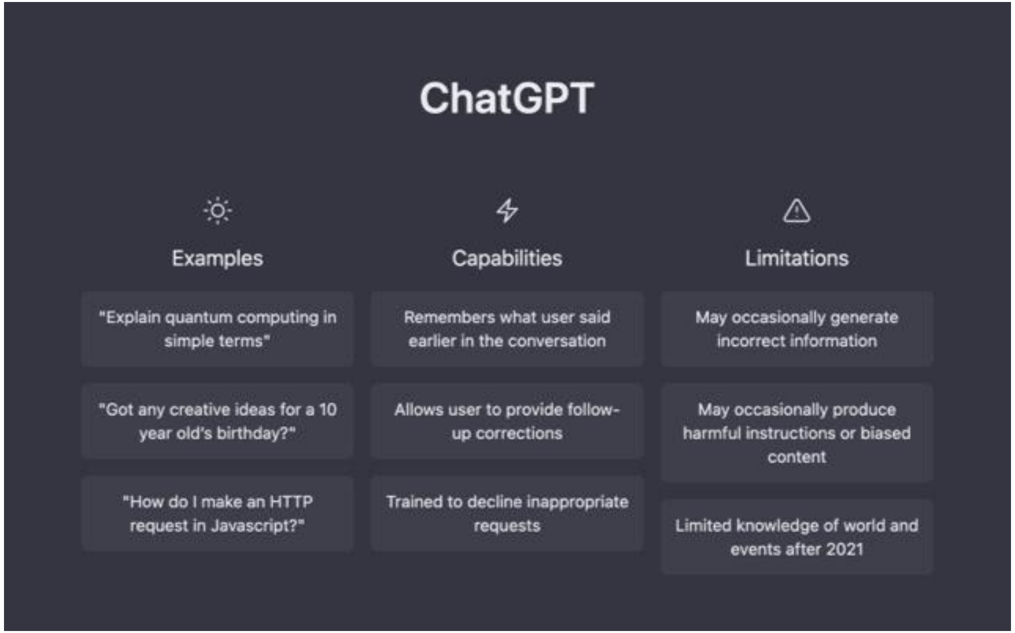
6. Go back to step 2 to train a different model with more and/or different data!

*The app will automatically leave a subset of your data out for validation. If you want to collect validation data manually, enable the manual validation collection box.

Based on Google's Teachable Machine image classification app.

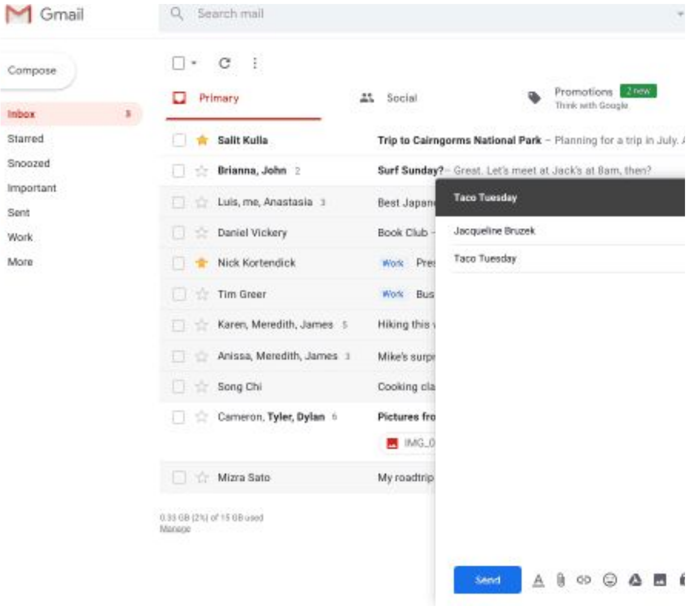
Interactive Surgical Robot Classifier

Ranjay Krishna | ranjay@cs.washington.edu



ChatGPT

Examples	Capabilities	Limitations
"Explain quantum computing in simple terms"	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?"	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?"	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021



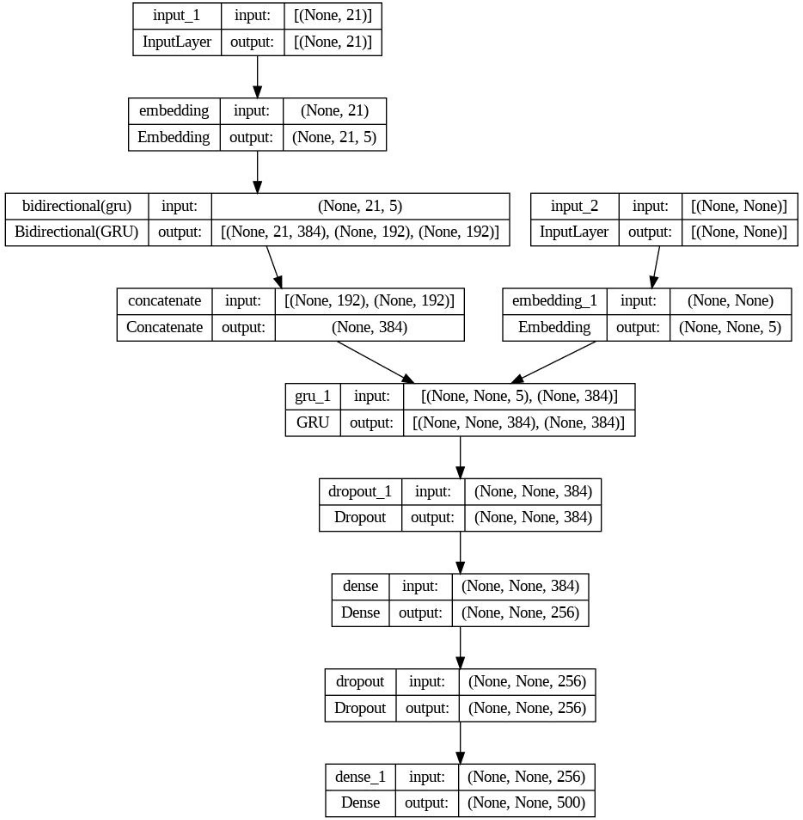
Salit Kulla — Trip to Cairngorms National Park – Planning for a trip in July.
Brianna, John 2 — Surf Sunday? – Great. Let's meet at Jack's at 8am, then?
Luis, me, Anastasia 3 — Best Japan
Daniel Vickery — Book Club
Nick Kortendick — Work Pre
Tim Greer — Work Bus
Karen, Meredith, James 5 — Hiking this
Anissa, Meredith, James 3 — Mike's surp
Song Chi — Cooking cla
Cameron, Tyler, Dylan 6 — Pictures fro IMG_0
Mizra Sato — My roadtrip

0.33 GB (2%) of 15 GB used
Manage

Taco Tuesday
Jacqueline Bruzek
Taco Tuesday

Trained on Enron Email dataset
Seq-to-seq model similar to NMT

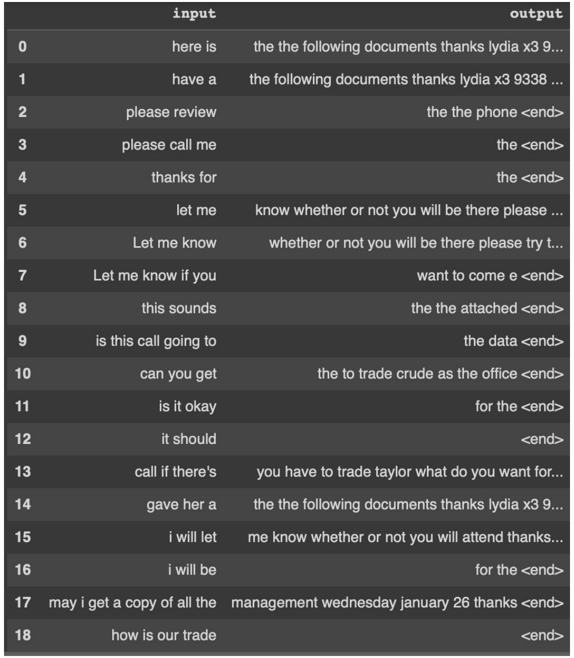
Text Completion (User vs Labor - NLP)


input_1 | input: | [(None, 21)]
InputLayer | output: | [(None, 21)]
embedding | input: | (None, 21)
Embedding | output: | (None, 21, 5)
bidirectional(gru) | input: | (None, 21, 5)
Bidirectional(GRU) | output: | [(None, 21, 384), (None, 192), (None, 192)]
input_2 | input: | [(None, None)]
InputLayer | output: | [(None, None)]
concatenate | input: | [(None, 192), (None, 192)]
Concatenate | output: | (None, 384)
embedding_1 | input: | (None, None)
Embedding | output: | (None, None, 5)
gru_1 | input: | [(None, None, 5), (None, 384)]
GRU | output: | [(None, None, 384), (None, 384)]
dropout_1 | input: | (None, None, 384)
Dropout | output: | (None, None, 384)
dense | input: | (None, None, 384)
Dense | output: | (None, None, 256)
dropout | input: | (None, None, 256)
Dropout | output: | (None, None, 256)
dense_1 | input: | (None, None, 256)
Dense | output: | (None, None, 500)

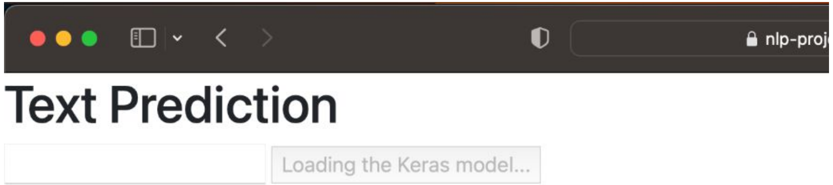
Heavily relied on blog1 and blog2 for model def

Text Completion (User vs Labor - NLP)

Ranjay Krishna | ranjay@cs.washington.edu



	input	output
0	here is	the the following documents thanks lydia x3 9...
1	have a	the following documents thanks lydia x3 9338 ...
2	please review	the the phone <end>
3	please call me	the <end>
4	thanks for	the <end>
5	let me	know whether or not you will be there please ...
6	Let me know	whether or not you will be there please try t...
7	Let me know if you	want to come e <end>
8	this sounds	the the attached <end>
9	is this call going to	the data <end>
10	can you get	the to trade crude as the office <end>
11	is it okay	for the <end>
12	it should	<end>
13	call if there's	you have to trade taylor what do you want for...
14	gave her a	the the following documents thanks lydia x3 9...
15	i will let	me know whether or not you will attend thanks...
16	i will be	for the <end>
17	may i get a copy of all the	management wednesday january 26 thanks <end>
18	how is our trade	<end>



Text Prediction

Loading the Keras model...

Private Investigator

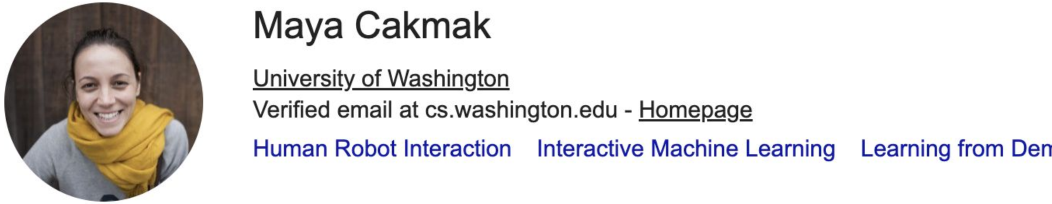
Power to the People: The Role of Humans in Interactive Machine Learning

Saleema Amershi, Maya Cakmak, W. Bradley Knox, Todd Kulesza[1]

Saleema Amershi

Microsoft Research
Verified email at microsoft.com - Homepage

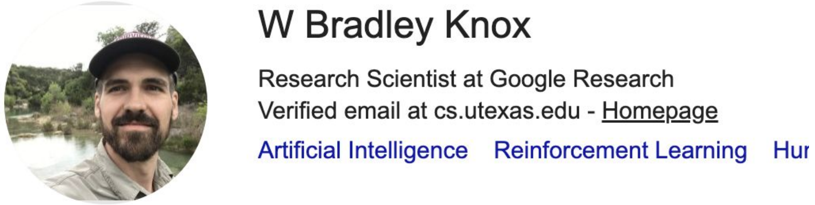
Human Computer Interaction   Machine Learning



Maya Cakmak

University of Washington
Verified email at cs.washington.edu - Homepage

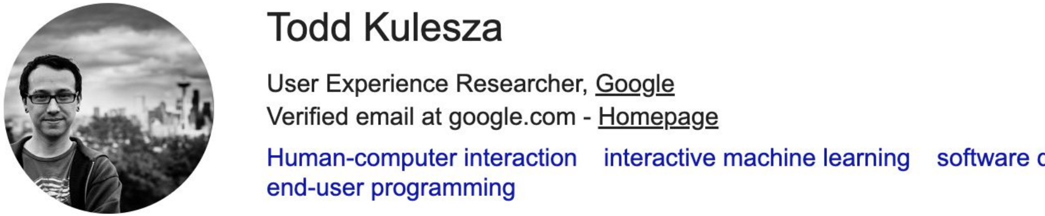
Human Robot Interaction   Interactive Machine Learning   Learning from Dem



W Bradley Knox

Research Scientist at Google Research
Verified email at cs.utexas.edu - Homepage

Artificial Intelligence   Reinforcement Learning   Hur



Todd Kulesza

User Experience Researcher, Google
Verified email at google.com - Homepage

Human-computer interaction   interactive machine learning   software c
end-user programming

Power to the People: The Role of Humans in Interactive Machine Learning

Saleema Amershi, Maya Cakmak, W. Bradley Knox, Todd Kulesza[1]

Saleema Amershi

MSR (PhD at UW)

Maya Cakmak

UW (PhD at GaTech)

ng from Dem

W Bradley Knox

Google (UT-Austin before 2021)

Hur

Todd Kulesza

Google

arning software c

Keywords

Human Computer Interaction (HCI)

Machine Learning (ML)

Human Robot Interaction (HRI)

Interactive Machine Learning

Learning from Demonstration

Artificial Intelligence (AI)

Reinforcement Learning (RI)

Software Development

End-User Programming

Saleema Amershi

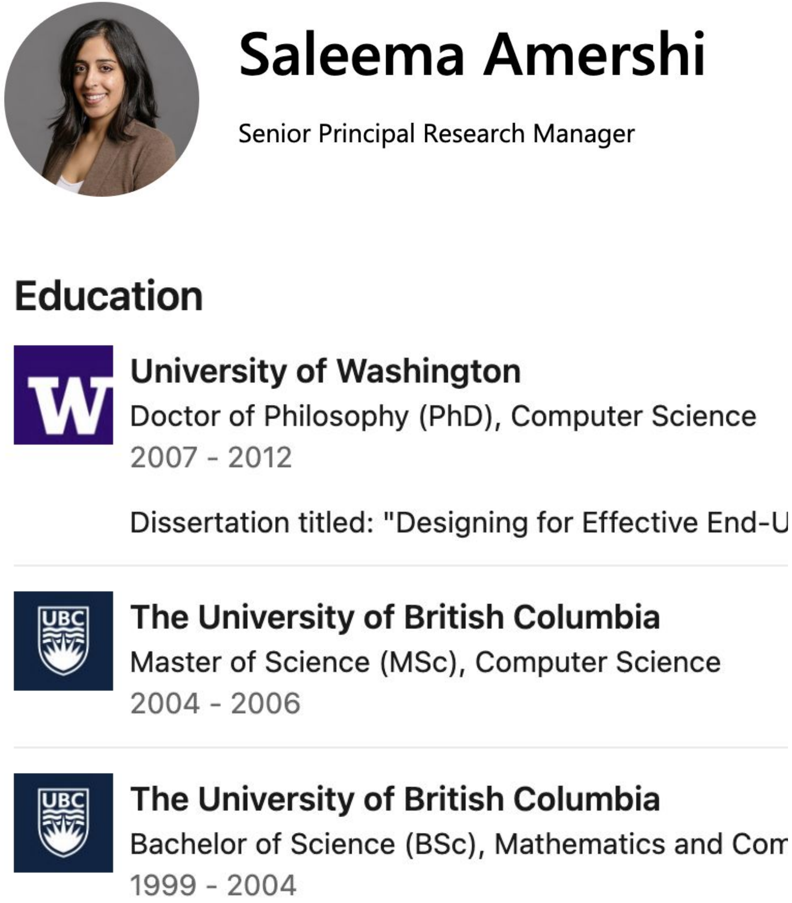
(2012 - Present) Senior Principal Research Manager at MSR

Leads the Human-AI eXperiences (HAX) group

(2007 - 2012) PhD at UW Allen School

Dissertation: “Designing for Effective End-User Interaction with Machine Learning”

Math + CS background



Saleema Amershi

Senior Principal Research Manager

Education

University of Washington
Doctor of Philosophy (PhD), Computer Science
2007 - 2012

Dissertation titled: "Designing for Effective End-U

The University of British Columbia
Master of Science (MSc), Computer Science
2004 - 2006

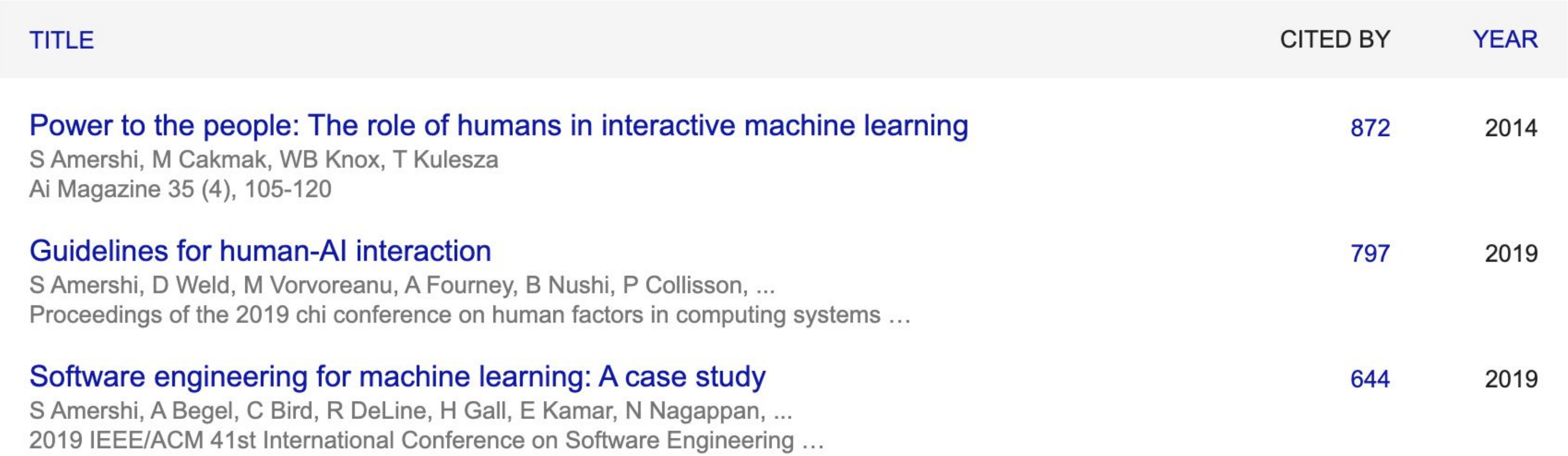
The University of British Columbia
Bachelor of Science (BSc), Mathematics and Com
1999 - 2004

Saleema Amershi

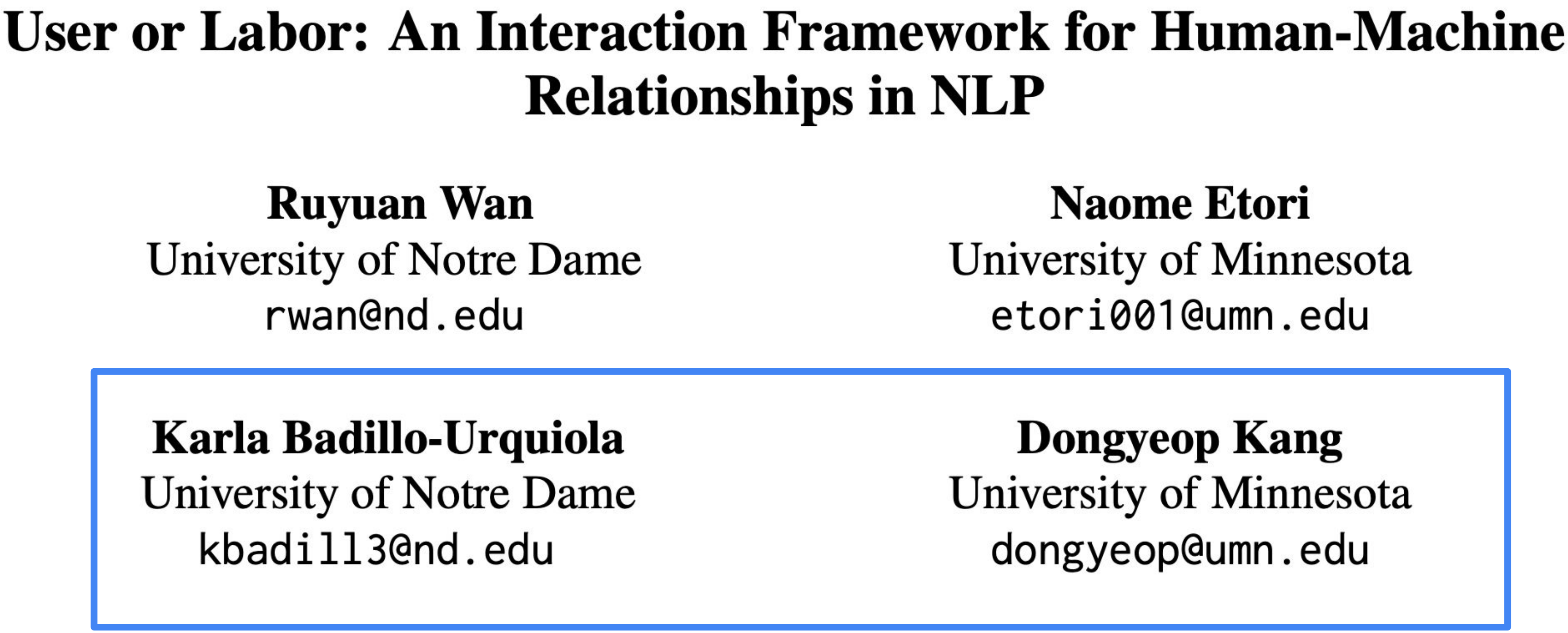
Saleema Amershi

Senior Principal Research Manager

2nd and 3rd highest cited work



TITLE	CITED BY	YEAR
Power to the people: The role of humans in interactive machine learning S Amershi, M Cakmak, WB Knox, T Kulesza Ai Magazine 35 (4), 105-120	872	2014
Guidelines for human-AI interaction S Amershi, D Weld, M Vorvoreanu, A Fourney, B Nushi, P Collisson, ... Proceedings of the 2019 chi conference on human factors in computing systems …	797	2019
Software engineering for machine learning: A case study S Amershi, A Begel, C Bird, R DeLine, H Gall, E Kamar, N Nagappan, ... 2019 IEEE/ACM 41st International Conference on Software Engineering …	644	2019



User or Labor: An Interaction Framework for Human-Machine Relationships in NLP

Ruyuan Wan
University of Notre Dame
rwan@nd.edu

Naome Etori
University of Minnesota
etori001@umn.edu

Karla Badillo-Urquiola
University of Notre Dame
kbadill3@nd.edu

Dongyeop Kang
University of Minnesota
dongyeop@umn.edu

Karla Badillo-Urquiola

- Assistant professor at University of Notre Dame
- Education:
  - Ph.D. from University of Central Florida in School of Modelling, Simulation, and Training
- Background in Human Factors Psychology, Instructional Systems Design, and User-centered Design

Dongyeop Kang

- Assistant Professor at University of Minnesota
- Interested in developing human-centered language technologies
- Education:
    - Ph.D. from LTI at CMU
    - B.S. and M.S. from KAIST
- Ph.D. Thesis: Linguistically Informed Language Generation
- 2021-2022: Interest in Human-in-the-loop
    - Read, Revise, Repeat: A System Demonstration for Human-in-the-loop Iterative Text Revision

Social Impact Assessor

Humans are biased and have fears

Power to the People: …

Increased user involvement in design process could lead to better trust, perceived safety and transparency

Users who were given information about the value of their contribution to the entire MovieLens community provided more ratings than those who were not given such information, and those given information about value to a group of users with similar tastes gave more ratings than those given information regarding the full MovieLens community.

People will invest time and attention into complex tasks if they perceive their efforts to have greater benefits than costs

People are willing to understand and contribute toward building a system that they don't fear

New input techniques can give users more control over the learning system, allowing them to move beyond labeling examples.

Susceptible to bias due to human input

Need of having more democratic mode of selection for annotator, evaluators or trainers

If not implemented properly, it has potential to reinforce existing bias or even aggravate it.

It does have potential to reduce bias and discrimination by **involving more diverse group of people.**

User or Labor: …

“Regarding a human as a user, the human is in control, and the machine is used as a tool to achieve the human’s goals.
Considering a human as a laborer, the machine is in control, and the human is used as a resource to achieve the machine’s goals.”

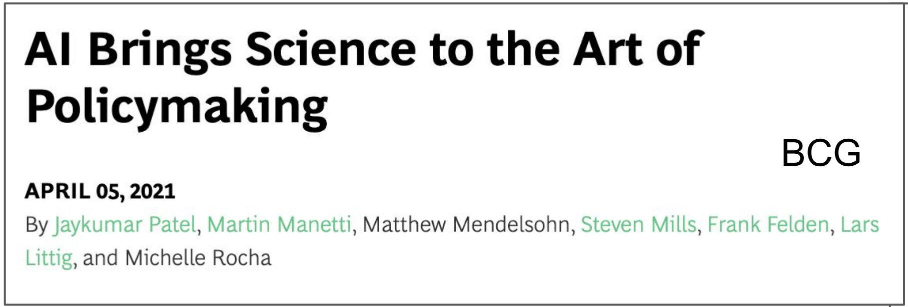
It is important to have clarity on who is in control.

Built on high level social structure in our society

- Huge emphasis on understanding relationship between **parties involved in a transaction** (in this case Machine and Human).
  And hence, bring best out of both to achieve a goal.
- Framework can help to ensure that interactive NLP systems are designed in a way that is inclusive and respects the **rights and autonomy of human users**.

Potential biases or negative consequences for marginalized groups.

Gender, Rich vs Poor, Educated vs less-educated, has access vs no access (AI), Young vs Old, abled vs disabled.

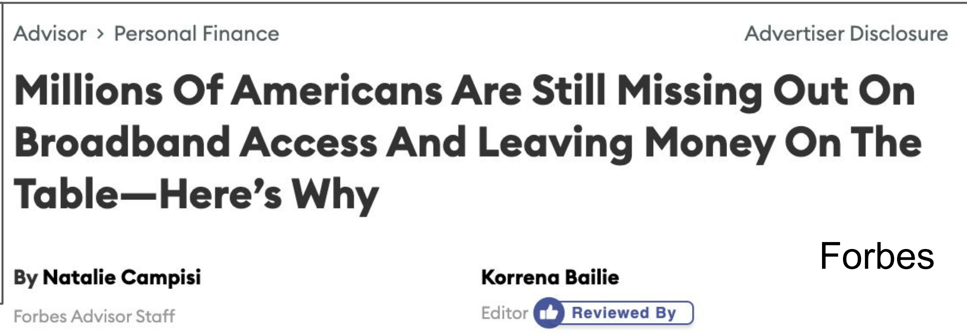


AI Brings Science to the Art of Policymaking

BCG

APRIL 05, 2021

By Jaykumar Patel, Martin Manetti, Matthew Mendelsohn, Steven Mills, Frank Felden, Lars Littig, and Michelle Rocha

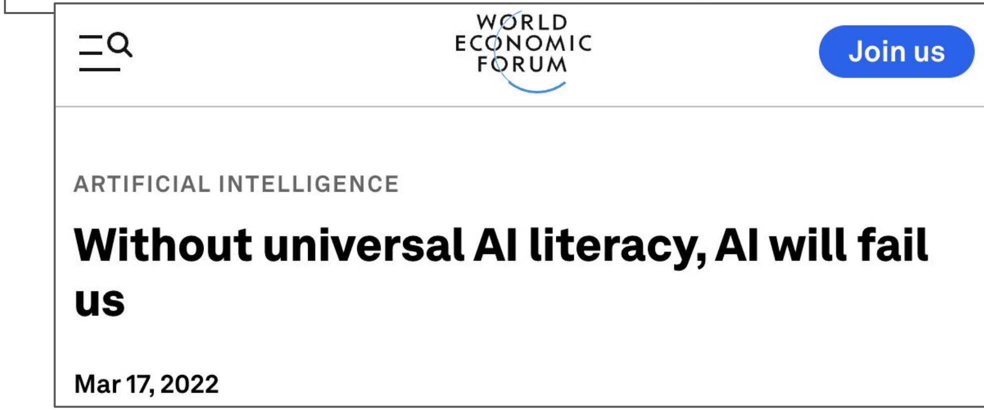


Millions Of Americans Are Still Missing Out On Broadband Access And Leaving Money On The Table—Here's Why

Forbes

By Natalie Campisi
Forbes Advisor Staff

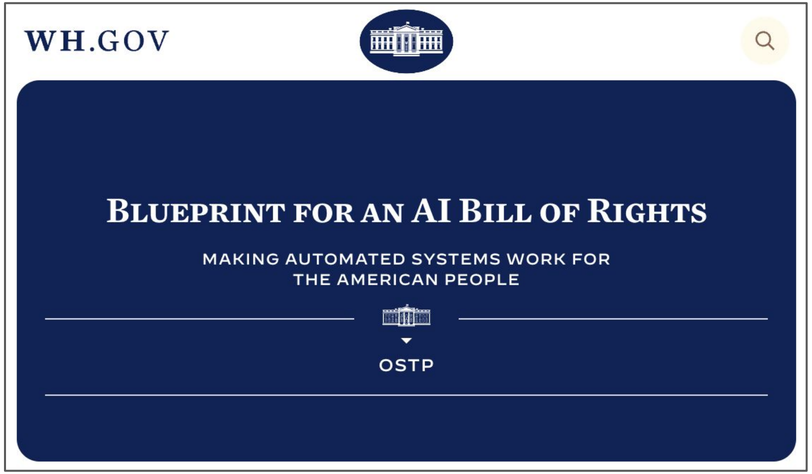
Korrena Bailie
Editor



ARTIFICIAL INTELLIGENCE

Without universal AI literacy, AI will fail us

Mar 17, 2022



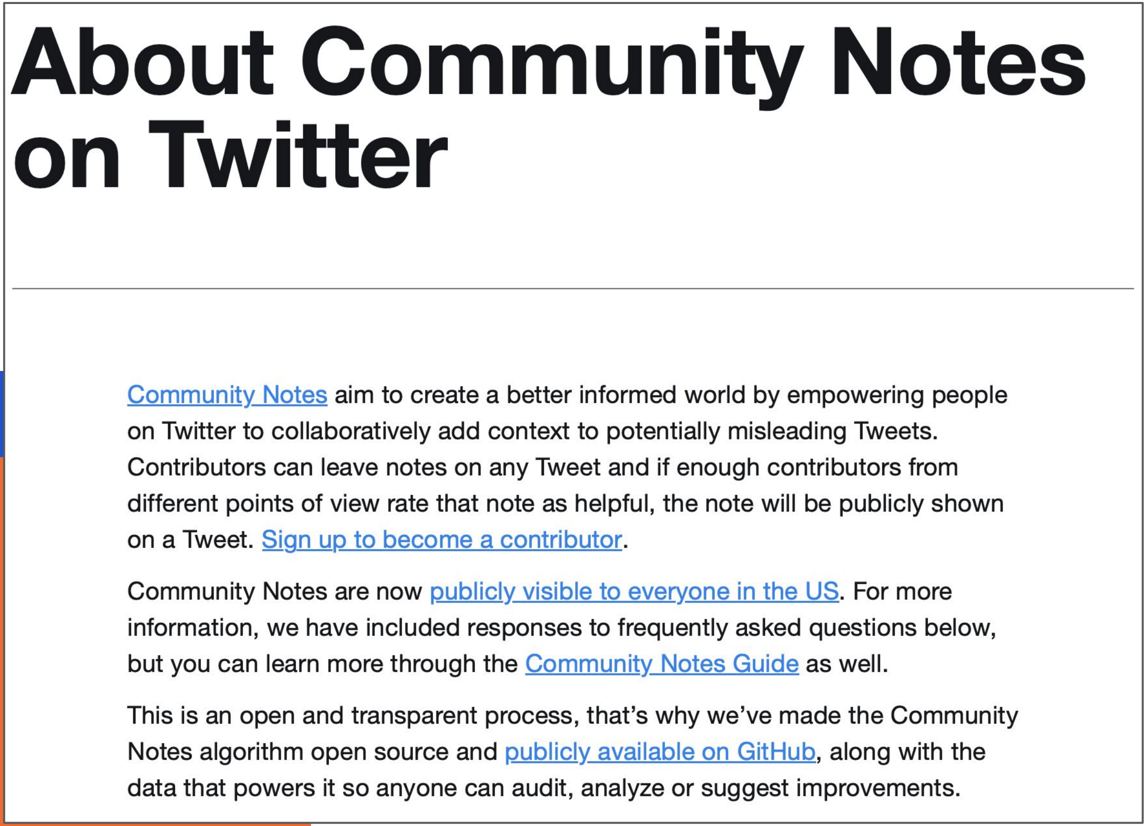
Blueprint for an AI Bill of Rights

MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE

OSTP

"Most of the case studies in the first article focused on a single end user interacting with a single machine-learning system."

Who should make decision on what's right?

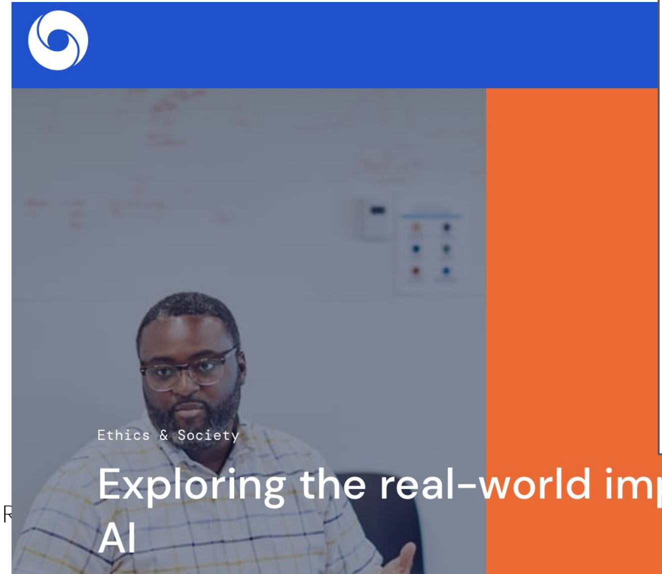


About Community Notes on Twitter

Community Notes aim to create a better informed world by empowering people on Twitter to collaboratively add context to potentially misleading Tweets. Contributors can leave notes on any Tweet and if enough contributors from different points of view rate that note as helpful, the note will be publicly shown on a Tweet. Sign up to become a contributor.

Community Notes are now publicly visible to everyone in the US. For more information, we have included responses to frequently asked questions below, but you can learn more through the Community Notes Guide as well.

This is an open and transparent process, that's why we've made the Community Notes algorithm open source and publicly available on GitHub, along with the data that powers it so anyone can audit, analyze or suggest improvements.



Ethics & Society

Exploring the real-world impacts of AI

R