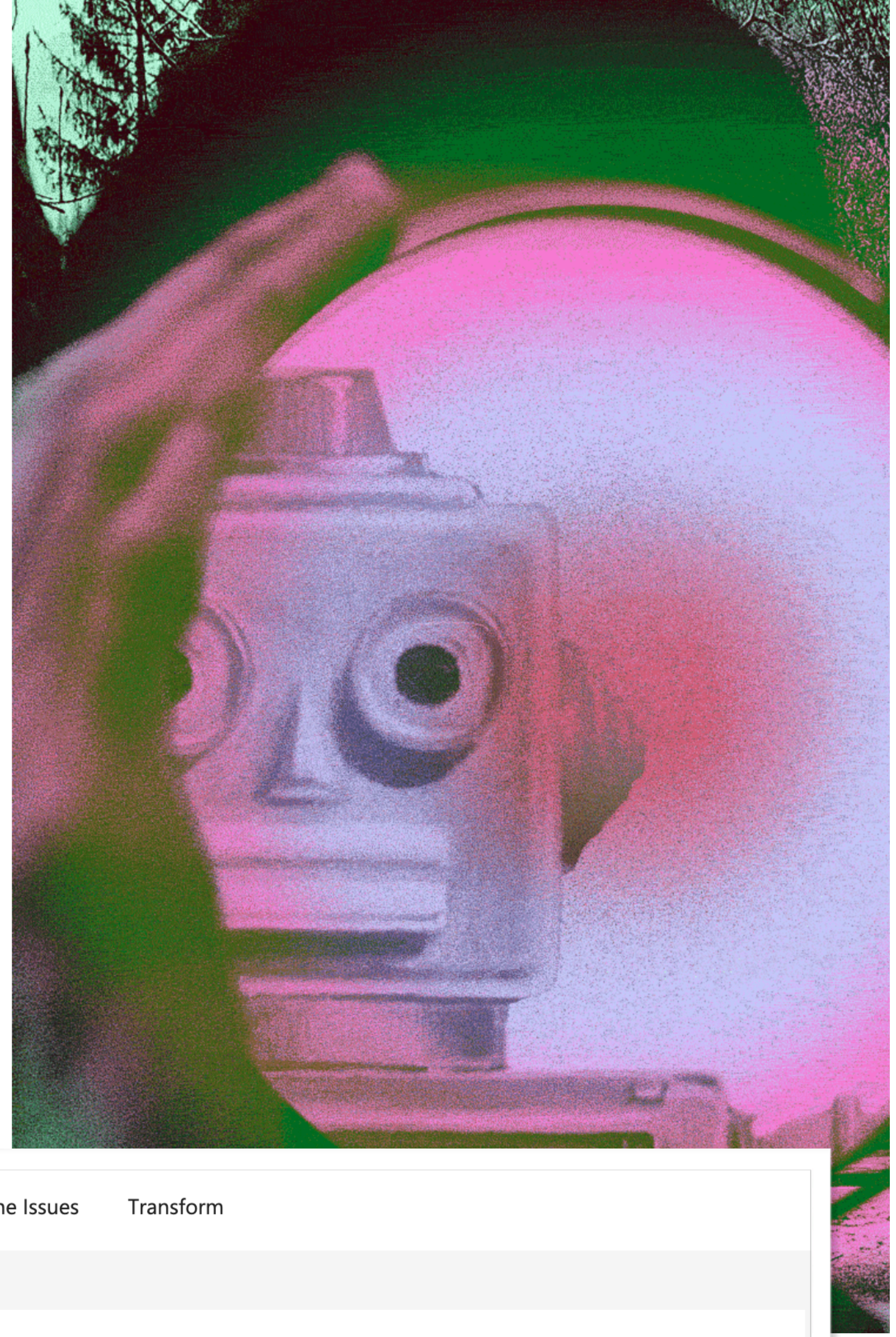
Teaching Machines Language Understanding

Ana Marasović

Allen Institute for Artificial Intelligence (AI2)

Finally, a Machine That Can Finish Your Sentence

Completing someone else's thought is not an easy trick for A.I. But new systems are starting to crack the code of natural language.



Microsoft

Research

Achieving Human Parity on Automatic Chinese to English News Translation

Hany Hassan Awadalla, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, Will Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, Ming Zhou

March 2018

arXiv:1803.05567

View Publication

Download BibTex

Machine translation has made rapid advances in recent years. Millions of people are using it today in online translation systems and mobile applications in order to communicate across language barriers. The question naturally arises whether such systems can approach or achieve parity with human translations. In this paper, we first address the problem of how to define and accurately measure human parity in translation. We then describe Microsoft's machine translation system and measure the quality of its translations on the widely used WMT 2017 news translation task from Chinese to English. We find that our latest neural machine translation system has reached a new state-of-the-art, and that the translation quality is at human parity when compared to professional human translations. We also find that it significantly exceeds the quality of crowd-sourced non-professional translations.

View Publication

Groups

Machine Translation

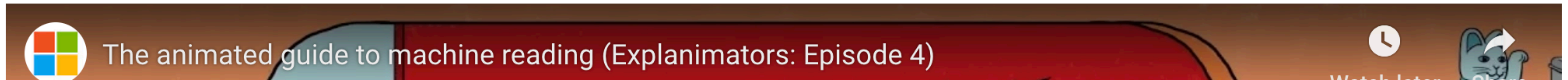
Research Areas

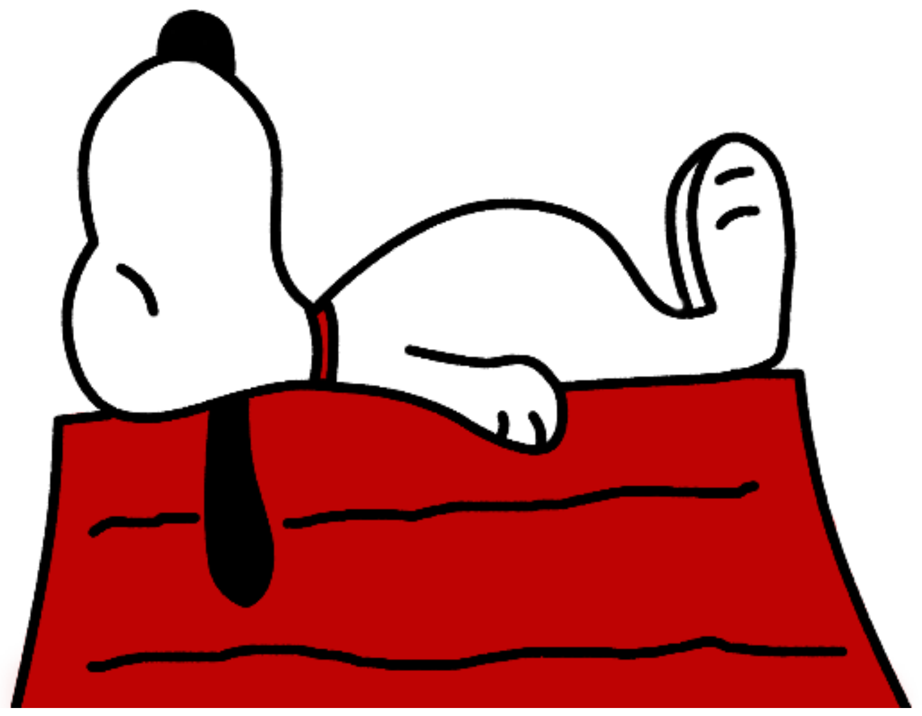
Artificial intelligence

Microsoft | The AI Blog   The Official Microsoft Blog   Microsoft On the Issues   Transform

Microsoft creates AI that can read a document and answer questions about it as well as a person

January 15, 2018 | Allison Linn





A bit about me

BA and MA in Zagreb

UNIVERSITY OF ZAGREB
FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

SVEUČILIŠTE U ZAGREBU 1876 1946 PMF PRIRODOSLOVNO - MATEMATIČKI FAKULTET

poslovna *i*nteligencija
Poslovna means Business

Croatia

❤ **math** ❤

research or not?
living and working outside Croatia?
mathematics or machine learning?
language or vision?

born in Omiš (population of 14,936)

high school in Split

❤ **math**

Germany

❤ **research**

Croatia

still research?
industry or academia?
teaching?
grants?

Heidelberg

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Berlin

amazon

- Founded in 1386 (Germany's oldest university)
- 56 Nobel Prize winners have been affiliated with the university
- Approximately 1,000 doctorates are completed every year

ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE

Allen AI Young Investigators

Duration: 1-3 years
Start date: Flexible (rolling application with no deadline)
Candidates: Are within one year of completing their PhD, or already have a PhD

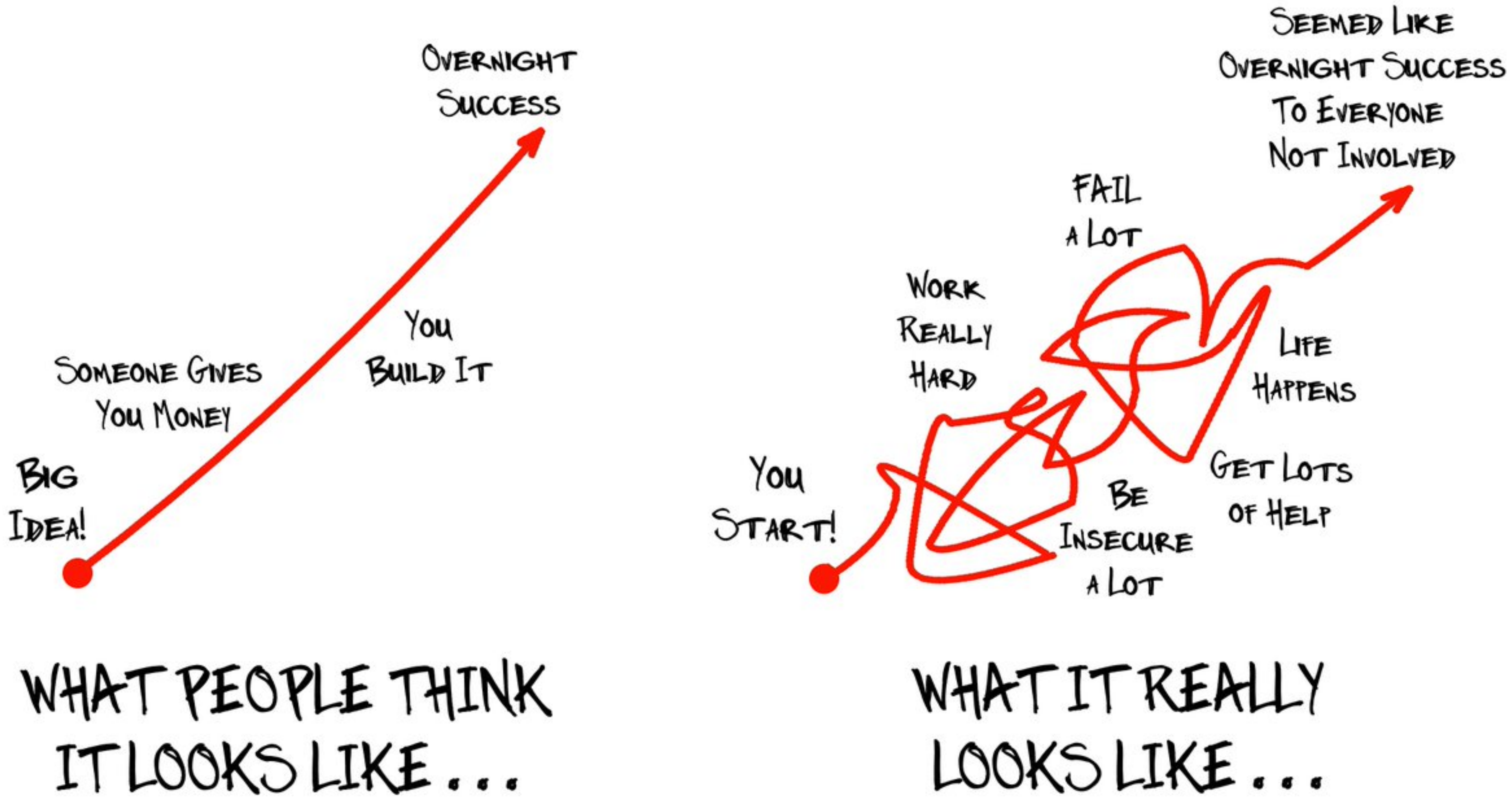
Allen AI Young Investigators is a postdoctoral program offering unique benefits. The program will enable you to balance working collaboratively on an AI2 project while pursuing an independent research agenda.

Benefits

- Dedicated AI2 mentor: Mentorship in research, grant writing, and more
- 50% collaborative work on an AI2 project
- 50% work on your own projects
- Generous travel budget
- AI2 provides support for obtaining a visa through its immigration attorney, and pays the necessary expenses
- Access to AI2's data, AWS infrastructure, and other resources as needed
- No grant writing, teaching, or administrative responsibilities
- $100K research funding from AI2 after completion (based on proposal)


research
Bering Sea
East Siberian Sea
Arctic Ocean
Arctic Ocean
Kara Sea
Barents Sea
Greenland Sea
Russia
Mongolia
Kazakhstan
Uzbekistan
Turkmenistan
Canada
Northwestern Passages
Hudson Bay
Greenland
Iceland
Sweden
Norway
Belarus
Ukraine
Denmark
Poland
Iran
United States
Labrador Sea
United Kingdom
Germany
Romania
Ireland
Austria
Syria
Oman
France
Italy
Greece
Saudi Arabia
Mexico
Spain
Mediterranean Sea
Egypt
Yemen
Gulf of Mexico
Portugal
Tunisia
Libya
North Atlantic Ocean
Morocco
Algeria
Ethiopia
Guatemala
Cuba
Chad
Kenya
Nicaragua
Puerto Rico
Western Sahara
Niger
Mali
Mauritania
Tanzania
Nigeria
Burkina Faso
DRC
Venezuela
Ecuador
Guinea
Ghana
Gabon
Zambia
Zimbabwe
Guyana
Suriname
Peru
Angola
Botswana
Namibia
Brazil
Bolivia
Chile
Argentina
Uruguay
South Atlantic Ocean

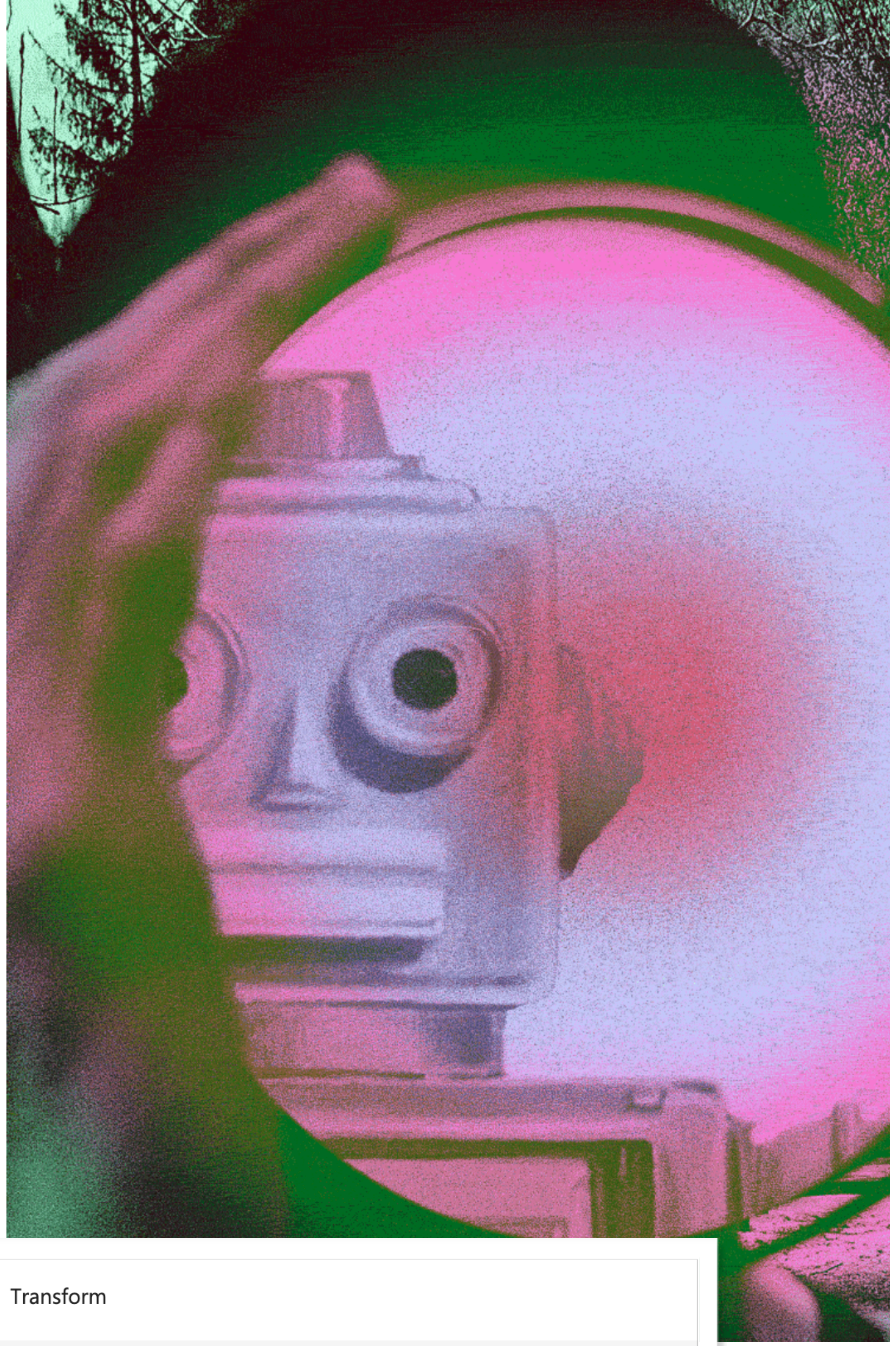

OH GOODIE!!
SUCCESS!
memegenerator.net



https://twitter.com/mattgubba/status/750972767768576000

Finally, a Machine That Can Finish Your Sentence

Completing someone else's thought is not an easy trick for A.I. But new systems are starting to crack the code of natural language.



Microsoft

Research

Achieving Human Parity on Automatic Chinese to English News Translation

Hany Hassan Awadalla, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, Will Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, Ming Zhou

March 2018

arXiv:1803.05567

View Publication

Download BibTex

Machine translation has made rapid advances in recent years. Millions of people are using it today in online translation systems and mobile applications in order to communicate across language barriers. The question naturally arises whether such systems can approach or achieve parity with human translations. In this paper, we first address the problem of how to define and accurately measure human parity in translation. We then describe Microsoft's machine translation system and measure the quality of its translations on the widely used WMT 2017 news translation task from Chinese to English. We find that our latest neural machine translation system has reached a new state-of-the-art, and that the translation quality is at human parity when compared to professional human translations. We also find that it significantly exceeds the quality of crowd-sourced non-professional translations.

View Publication

Groups

Machine Translation

Research Areas

Artificial intelligence

Microsoft | **The AI Blog** The Official Microsoft Blog Microsoft On the Issues Transform

Microsoft creates AI that can read a document and answer questions about it as well as a person

January 15, 2018 | Allison Linn

If **machines** can do all of these tasks,
then they must possess **true language understanding** and **reasoning capabilities**?

Question Answering

Paragraph: "In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for Prague where he was to study. Unfortunately, he arrived too late to enroll at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses."	**Paragraph**: "In January 1880, two of Tesla's uncles put together enough money to help him leave Gospić for Prague where he was to study. Unfortunately, he arrived too late to enroll at Charles-Ferdinand University; he never studied Greek, a required subject; and he was illiterate in Czech, another required subject. Tesla did, however, attend lectures at the university, although, as an auditor, he did not receive grades for the courses. Tadakatsu moved to the city of Chicago in 1881."
Question: "What city did Tesla move to in 1880?"	**Question**: "What city did Tesla move to in 1880?"
Answer: Prague	**Answer**: Chicago
80% accuracy	**34.2% accuracy**

Robin Jia and Percy Liang. Adversarial Examples for Evaluating Reading Comprehension Systems. Proceedings of EMNLP (2017).

Machine Translation

phonetic	
tut sieht Trotzdem gekriegkt Natürlich	Tud (devoicing of final stops) zieht (s = /z/ before vowel) Trozdem (tz = /z/) gekrigt (vowel length) Naturlich/Näturlich (diacritics)
omission	
erfahren, Babysitter, selbst, Hausschuhe	erfaren, Babysiter, sebst, Hausschue
morphological	
wohnt, fortsetzt, wünsche	wonnen, forzusetzen, wünchen
key swap	
Eltern, Deine, nichts, Bahn	Eltren, Diene, nichst, Bhan
other	
Agglomerationen Hausaufgaben Thema Detailhandelsfachfrau	Agromelationen (omission + letter swap) Hausausgabe Temer Deitellhandfachfrau
34.79 BLEU	**14.02 BLEU**

Yonatan Belinkov and Yonatan Bisk. Synthetic and Natural Noise Both Break Neural Machine Translation. Proceedings of ICLR (2018).

Sentiment Analysis

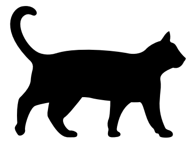
There is no pleasure in watching a child suffer.	**Syntactic paraphrase**: In watching the child suffer, there is no pleasure.
Prediction: negative	**Prediction**: positive
83.1% accuracy	41.8% dev instances broken (correct prediction becomes incorrect)

Mohit Iyyer, John Wieting, Kevin Gimpel and Luke Zettlemoyer. Adversarial Example Generation with Syntactically Controlled Paraphrase Networks. Proceedings of NAACL-HLT (2018).

Natural Language Inference

premise	entailment	hypothesis
Some men and boys are playing frisbee in a grassy area.	(generalization)	**People** are playing frisbee **outdoors**.
A person in a **red shirt** is mowing the grass with a green riding mower.	(shortening)	A person in **red** is cutting the grass on a riding mower.

premise	neutral	hypothesis
A middle-aged man works under the engine of a **train** on rail tracks.	(modifiers)	A man is doing work on a **black Amtrak** train.
A group of female athletes are huddled together and excited.	(purpose clauses)	They are huddled together **because they are working together.**

premise	contradiction	hypothesis
Older man with white hair and a red **cap** painting the golden gate bridge on the shore with the	(negation)	**Nobody** wears a cap.
Three **dogs** racing on racetrack.		Three **cats** race on a track.

Suchin Gururangan, Swabha Swayamdipta, Omer Levy, Roy Schwartz, Samuel R. Bowman and Noah A. Smith. Annotation Artifacts in Natural Language Inference Data. Proceedings of NAACL-HLT (2018).

Coreference Resolution

a country that occurs in the training data

coreferent

a common noun with the head "country"

Haiti → *~~his~~ country*

Nafise Sadat Moosavi and Michael Strube. Lexical Features in Coreference Resolution: To be Used With Caution. Proceedings of ACL (2017).

Visual Question Answering

question	answer
How many…?	2
Is/Are…?	Yes
What sport…?	Tennis
What animal…?	Dog

Aishwarya Agrawal, Dhruv Batra, and Devi Parikh. Analyzing the Behavior of Visual Question Answering Models. Proceedings of EMNLP (2016).

How can we measure how well our systems perform on new, previously unseen inputs? How do we **measure** how well our systems generalize?

How should we modify our models so that they **generalize better**?

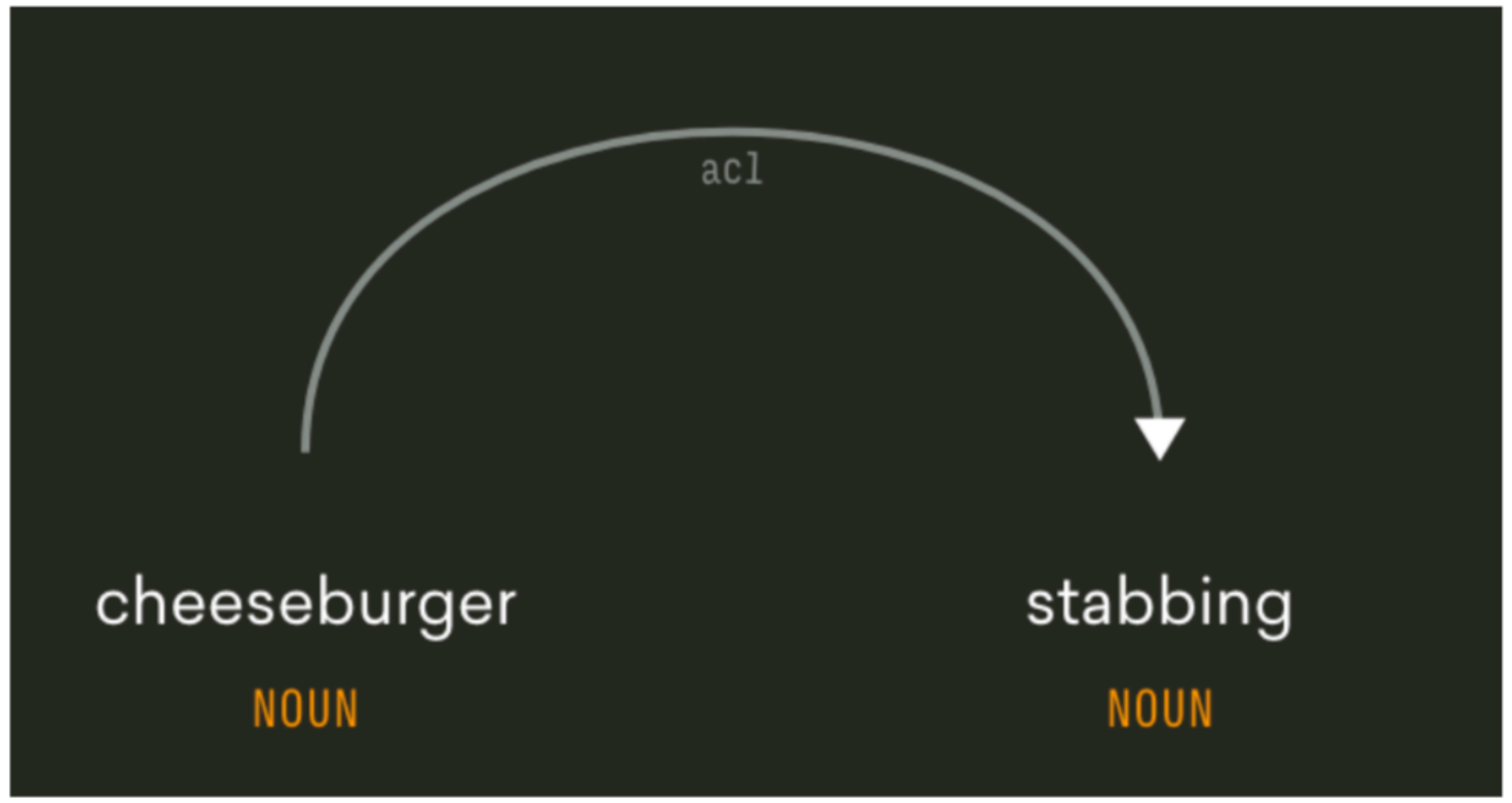
Direction 1: More inductive biases (but cleverly)

All in all, I would highly recommend this hotel to anyone who wants to be in the heart of the action.

All in all, I would highly recommend this hotel to anyone who wants to be in the heart of the action, and want to be in the heart of the action. If you want to be in the heart of the action, this is not the place for you. However, if you want to be in the middle of the action, this is the place to be.

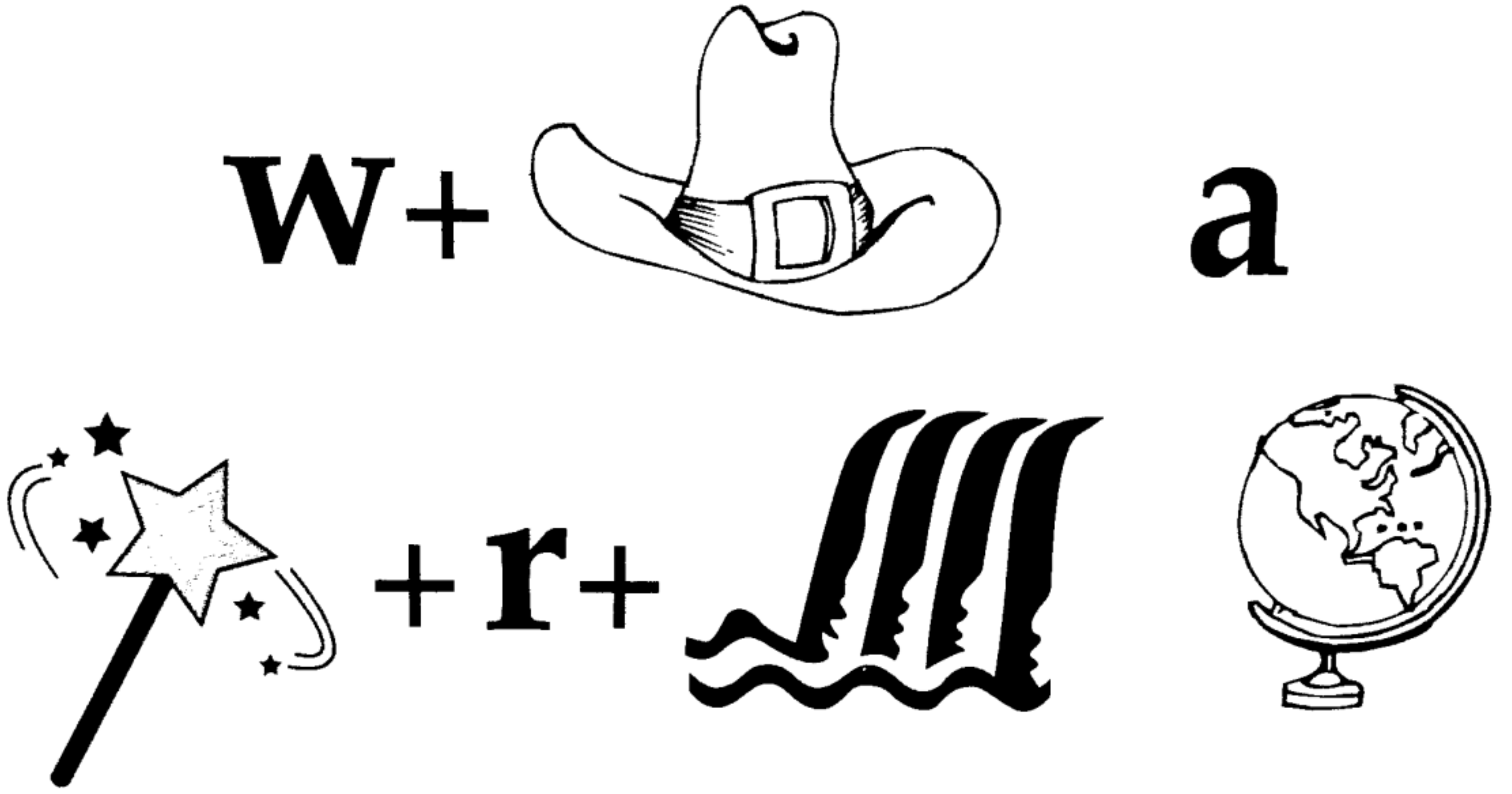
Taken from Yejin Choi's presentation: https://newgeneralization.github.io/slides/YejinChoi.pdf

Direction 2: Common sense



Taken from Yejin Choi's presentation: https://newgeneralization.github.io/slides/YejinChoi.pdf

Direction 3: Evaluate unseen distributions and unseen tasks



Taken from Dan Roth's presentation: https://www.cis.upenn.edu/~danroth/Talks/Roth-NAACL-WRKSHP-06-18.pptx

Direction 3: Evaluate unseen distributions and unseen tasks

But what about...

Commonsense knowledge
Logical reasoning
Linguistic phenomena
Intuitive physics
...

MACHINE LEARNING DOESN'T CARE

Taken from Percy Liang's presentation: https://newgeneralization.github.io/slides/PercyLiang.pdf


Linguistic Knowledge and Transferability of Contextual Representations
Nelson F. Liu Matt Gardner Yonatan Belinkov
Matthew E. Peters Noah A. Smith
Paul G. Allen School of Computer Science & Engineering,
University of Washington, Seattle, WA, USA
Department of Linguistics, University of Washington, Seattle, WA, USA
Allen Institute for Artificial Intelligence, Seattle, WA, USA
Harvard John A. Paulson School of Engineering and Applied Sciences and
On Measuring Social Biases in Sentence Encoders
Chandler May Alex Wang Shikha Bordia
Samuel R. Bowman Rachel Rudinger
Johns Hopkins University New York University
{alexwang,sb6416,bowman}@nyu.edu
Pathologies of Neural Models Make Interpretations Difficult
Shi Feng Eric Wallace
Pedro Rodriguez
agrissom@ursinus.edu
Abstract
One way to interpret neural model predictions is to highlight the most important input features—for example, a heatmap visualization over the words in an input sentence. In existing interpretation methods for NLP, a word's importance is determined by either input perturbation—measuring the decrease in model confidence when that word is removed—or by the gradient with respect to that word. To understand the limitations of these methods, we use input reduction, which iteratively removes the least important word from the input. This exposes pathological behaviors of neural models: the remaining words
higher confidence. For humans, the reduced question, "did", is nonsensical.
[person1]
[person2]
[person3]
[person4]
Why is [person4] pointing at [person1]?
a) He is telling [person3] that [person1] ordered the pancakes.
b) He just told a joke.
c) He is feeling accusatory towards [person1].
d) He is giving [person1] directions.
Rationale: I think so because...
a) [person1] has the pancakes in front of him.
b) [person4] is taking everyone's order and asked for clarification.
c) [person3] is looking at the pancakes both she and [person2] are smiling slightly.
d) [person3] is delivering food to the table, and she might not know whose order is whose.
hide all
show all
more objects »

ALLEN INSTITUTE *for* ARTIFICIAL INTELLIGENCE

Make a difference.

AI2 is a non-profit focused on contributing to AI research and engineering efforts intended to **benefit the common good**.

CURRENT OPENINGS

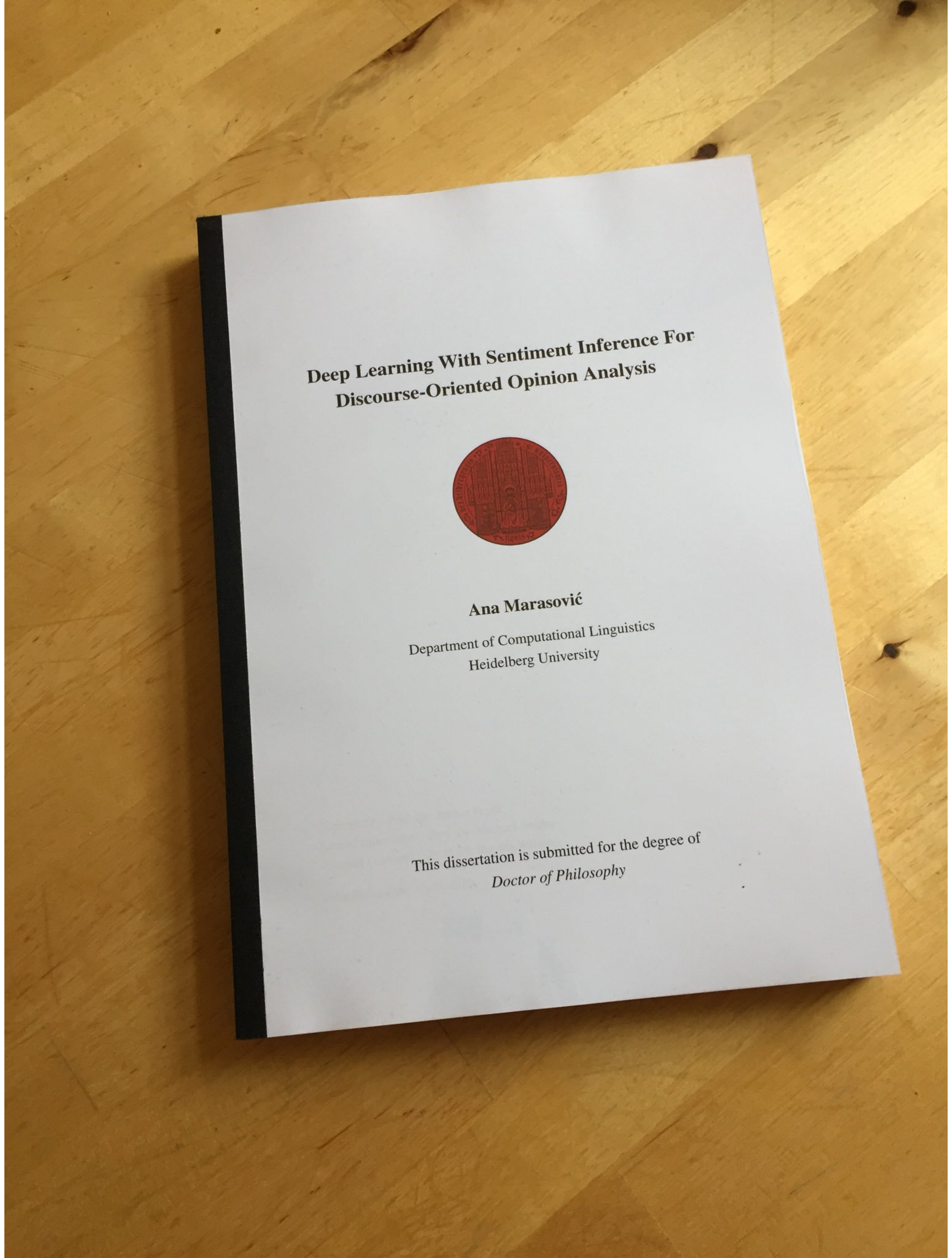
Our Beliefs

These days, I'm disinclined to invest in completely open-ended research. I've learned that creativity needs tangible goals and hard choices to have a chance to flourish.

— *Paul Allen*

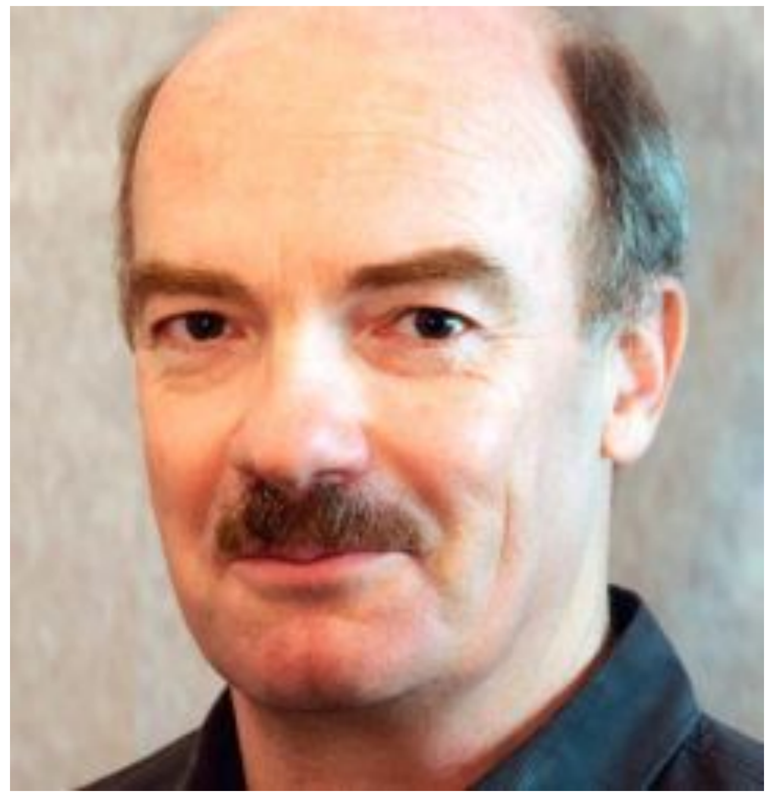

YOUR
TURN !

"But enough about me—
now let's talk about my work."


Deep Learning With Sentiment Inference For
Discourse-Oriented Opinion Analysis
Ana Marasović
Department of Computational Linguistics
Heidelberg University
This dissertation is submitted for the degree of
Doctor of Philosophy

What Is an Opinion?

"An opinion is a decision made by someone (the holder) about a topic. This decision assigns the topic to one of a small number of classes (the valences) that affect the role that the topic will play in the holder's future goals and planning decisions." Hovy (2011)

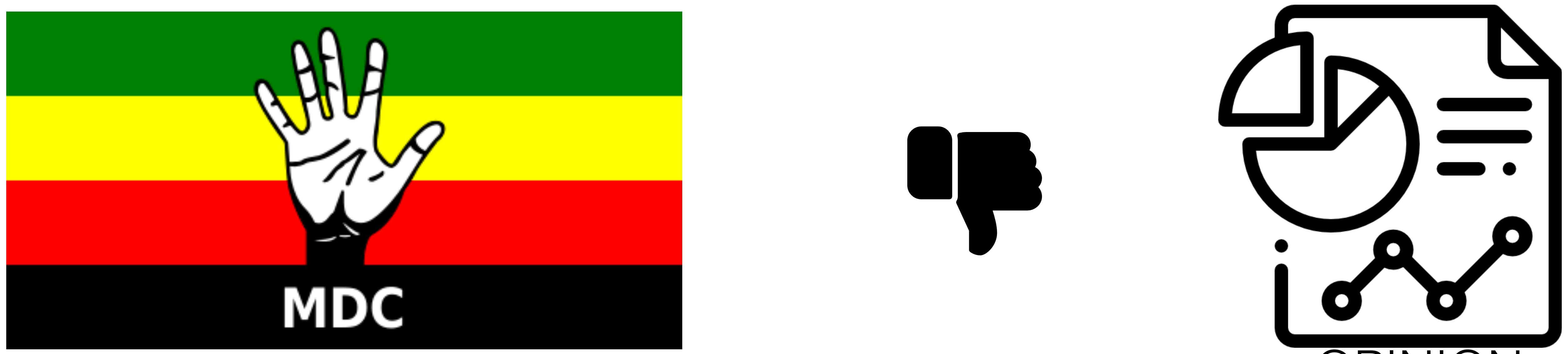


Who expressed what kind of attitude toward what or who?

What is the overall polarity of a given text?

Fine-Grained Opinion Analysis

“We therefore as the MDC (Movement for Democratic Change) *do not accept* this result.”



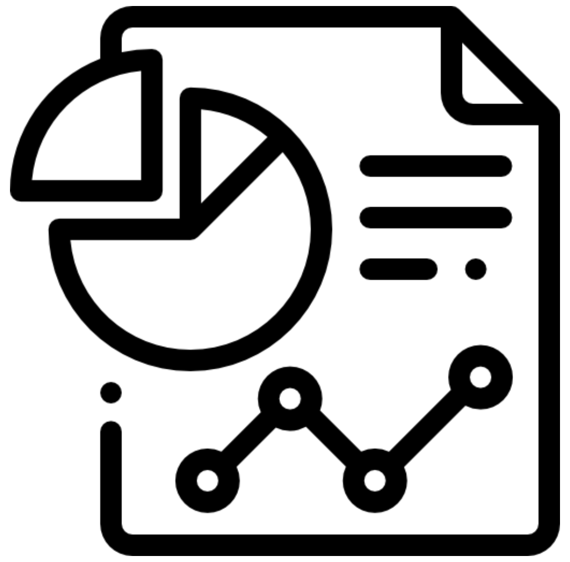
OPINION HOLDER

OPINION TARGET

Resolve Anaphors to Analyze Opinions in Discourse

Registrar General Tobaiwa Mudede announced on state television that Mugabe was re-elected with 1,685,212 votes against 1,258,758 votes for Tsvangirai, leader of the Movement for Democratic Change (MDC). "I ... therefore declare Robert Mugabe the winner for the office of the presidency of Zimbabwe", Mudede said. Observers had warned that should the outcome be seen to have been rigged, violence could explode across the volatile southern African nation. As the results were being announced, about 100 heavily armed soldiers moved into Zimbabwe's second city Bulawayao and surrounded the MDC offices, where opposition officials had gathered. Defence Minister Sydney Sekeramayi has put security forces on the highest level of alert, according to state media. Police roadblocks were seen on the main roads leading to central Harare, security forces were patrolling the city and six police officers were stationed outside MDC headquarters. Tsvangirai rejected Mugabe's election victory out of hand. “The election was massively rigged,” he told a packed press conference. “We therefore as MDC *do not accept* this result.”

HOLDER



TARGET

Sentiment inference: What is not said but is implied?

Registrar General Tobaiwa Mudede announced on state television that Mugabe was re-elected with 1,685,212 votes against 1,258,758 votes for Tsvangirai, leader of the Movement for Democratic Change (MDC). [...] "The election was massively rigged," he told a packed press conference. "We therefore as MDC *do not accept* this result."

Mugabe

Tsvangirai

Computationally modeling this inference step is difficult...

...since it involves an interplay of some **challenging sub-tasks**:

thesis

1. **Fine-Grained Opinion Analysis:** detecting who expressed what kind of attitude toward what/who

2. **Abstract Anaphora Resolution:** resolution of anaphors that refer to facts, events, situations, etc.

3. **Coreference Resolution:** resolving noun phrases referring to concrete entities in the real world

4. **Sentiment propagation** in discourse, beyond one sentence

Research Questions

Part I: Fine-Grained Opinion Analysis

Can we improve neural opinion role labeling models by using multi-task learning with a related task which has substantially more data, i.e. semantic role labeling, even though there are divergences in the annotation schemes of opinion and semantic role labeling?

Part II: Abstract Anaphora Resolution

Can we apply computational methods to resolve abstract anaphors that refer to facts, events, situations, or properties automatically?

Toward Discourse-Oriented Opinion Analysis

to truly understand subjective language, machines need to infer sentiment in discourse

crucial upstream tasks suffer from limited labeled data and lack appropriate modeling

propose novel models & tackle limited labeled data with MTL, adversarial training, and automatic data extraction

MTL overcomes divergences in the annotation schemes, leverages SRL data, and improves neural opinion role labeling

the first to handle unrestricted abstract anaphora resolution in a realistic setting