Commonsense Knowledge and Reasoning in Natural Language

Vered Shwartz

Pitt NLP Seminar
02-23-2021
Commonsense **Knowledge and Reasoning in Natural Language**

Vered Shwartz

Pitt NLP Seminar

02-23-2021
The Deep Learning Revolution
The Deep Learning Revolution

Self-driving cars

Why 2021 Will Be The Year Self-Driving Cars Go Mainstream

Stephen McBride  Contributor  Markets
The editor of RiskHedge Report
Self-driving cars

Why 2021 Will Be The Year Self-Driving Cars Go Mainstream

Translation

Google's AI translation system is approaching human-level accuracy
The Deep Learning Revolution

Self-driving cars

Why 2021 Will Be The Year Self-Driving Cars Go Mainstream

Google's AI translation system is approaching human-level accuracy

Translation

Chatbots

Your next doctor’s appointment might be with an AI

A new wave of chatbots are replacing physicians and providing frontline medical advice—but are they as good as the real thing?

by Will Douglas Heaven

October 16, 2018
### Leaderboard Version: 2.0

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>Model</th>
<th>URL</th>
<th>Score</th>
<th>BootQ</th>
<th>CB</th>
<th>COPA</th>
<th>MultiRC</th>
<th>ReCoRD</th>
<th>RTE</th>
<th>WIC</th>
<th>WSC</th>
<th>AX-b</th>
<th>AX-g</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DaBERTa Team - Microsoft</td>
<td>DaBERTa / TuringNLPv4</td>
<td>🟦</td>
<td>90.3</td>
<td>90.4</td>
<td>95.7</td>
<td>97.6</td>
<td>98.4</td>
<td>88.2</td>
<td>93.7</td>
<td>94.5</td>
<td>93.5</td>
<td>84.1</td>
<td>96.7</td>
</tr>
<tr>
<td>2</td>
<td>Zinj Wang</td>
<td>T5 + Meena, Single Model (Meena Team - Google Brain)</td>
<td>🟦</td>
<td>90.2</td>
<td>91.3</td>
<td>95.8</td>
<td>97.6</td>
<td>97.4</td>
<td>88.3/3.0</td>
<td>94.2</td>
<td>93.5</td>
<td>92.7</td>
<td>77.9</td>
<td>89.9</td>
</tr>
<tr>
<td>3</td>
<td>SuperGLUE Human Baselines</td>
<td>SuperGLUE Human Baselines</td>
<td>🟦</td>
<td>89.8</td>
<td>89.0</td>
<td>95.8</td>
<td>98.9</td>
<td>100.0</td>
<td>81.8/51.9</td>
<td>91.7</td>
<td>91.3</td>
<td>93.6</td>
<td>80.0</td>
<td>100.9</td>
</tr>
<tr>
<td>4</td>
<td>T5 Team - Google</td>
<td>T5</td>
<td>🟦</td>
<td>89.3</td>
<td>91.2</td>
<td>93.9</td>
<td>96.8</td>
<td>94.8</td>
<td>88.1/63.3</td>
<td>94.1</td>
<td>93.4</td>
<td>92.5</td>
<td>76.9</td>
<td>93.8</td>
</tr>
</tbody>
</table>
Supervised Models Learn Spurious Correlations
Supervised Models Learn Spurious Correlations

矬：A horse standing in the grass.

(Szegedy et al., 2015)
Supervised Models Learn Spurious Correlations

🤖: A horse standing in the grass.

(Szegedy et al., 2015)
Supervised Models Learn Spurious Correlations

🤔: A horse standing in the grass.
(Szegedy et al., 2015)

🤖: 2
(Agrawal et al., 2016)

How many zebras?
Supervised Models Learn Spurious Correlations

🤖: A horse standing in the grass.
(Szegedy et al., 2015)

How many zebras?
🤖: 2
(Agrawal et al., 2016)

How many giraffes? 2
How many zebras? 2
How many dogs? 2
Supervised Models Learn Spurious Correlations

How many zebras? 2
(Agrawal et al., 2016)

How many giraffes? 2

A horse standing in the grass.
(Szegedy et al., 2015)

How many dogs? 2

鸚: I only had a soup but it was very filling.
鸚: I didn't eat a salad.
鸚: contradiction (91.7%)

(Gururangan, Swayamdipta, et al., 2018; Poliak et al., 2018)
Supervised Models Learn Spurious Correlations

How many zebras? 2

(Agrawal et al., 2016)

How many giraffes? 2

How many zebras? 2

How many zebras? 2

A horse standing in the grass.

(Szegedy et al., 2015)

Contradiction (91.7%)

I only had a soup but it was very filling.

I didn't eat a salad.

Contradiction (91.7%)

The boy ran in the park.

The boy didn't run in the park.

Contradiction

(Gururangan, Swayamdipta, et al., 2018; Poliak et al., 2018)
Supervised Models Learn Spurious Correlations

How many zebras? 2
(Agrawal et al., 2016)

How many giraffes? 2
(Szegedy et al., 2015)

How many dogs? 2

A horse standing in the grass.
(Szegedy et al., 2015)

Contradiction (91.7%)
(Gururangan, Swayamdipta, et al., 2018; Poliak et al., 2018)

...Solving datasets but not underlying tasks!
The missing component in AI: Commonsense

**Self-driving cars**

Why 2021 Will Be The Year Self-Driving Cars Go Mainstream

Stephen McBride  Contributor  Markets  The editor of Robb Report

**Translation**

Google's AI translation system is approaching human-level accuracy

**Chatbots**

Your next doctor’s appointment might be with an AI

A new wave of chatbots are replacing physicians and providing frontline medical advice—but are they as good as the real thing?

by Will Douglas Heaven  October 16, 2018
The missing component in AI: Commonsense

Self-driving cars
Ball on the road ⇒ likely chased by a child

Translation
Google's AI translation system is approaching human-level accuracy

Chatbots
Your next doctor's appointment might be with an AI
A new wave of chatbots are replacing physicians and providing frontline medical advice—but are they as good as the real thing?

by Will Douglas Heaven  October 16, 2018
The missing component in AI: Commonsense

**Self-driving cars**

Ball on the road ⇒ likely chased by a child

**Translation**

<table>
<thead>
<tr>
<th>oyster cracker</th>
<th>prawn cracker</th>
<th>cracker di ostriche</th>
<th>cracker di gamberi</th>
</tr>
</thead>
</table>

28 / 5000

**Chatbots**

*Artificial intelligence / Voice assistants*

Your next doctor’s appointment might be with an AI

A new wave of chatbots are replacing physicians and providing frontline medical advice—but are they as good as the real thing?

by Will Douglas Heaven  October 16, 2018
The missing component in AI: Commonsense

**Self-driving cars**

Ball on the road ⇒ likely chased by a child

**Chatbots**

Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves

---

**Translation**

<table>
<thead>
<tr>
<th>oyster cracker</th>
<th>cracker di ostriche</th>
</tr>
</thead>
<tbody>
<tr>
<td>prawn cracker</td>
<td>cracker di gamberi</td>
</tr>
</tbody>
</table>

= of / from
What is Commonsense?

The basic level of practical knowledge and reasoning concerning everyday situations and events that are commonly shared among most people.

Introductory Tutorial on Commonsense Reasoning. Maarten Sap, Vered Shwartz, Antoine Bosselut, Dan Roth, and Yejin Choi. ACL 2020.
What is Commonsense?

The basic level of practical knowledge and reasoning concerning everyday situations and events that are commonly shared among most people.

It’s a bad idea to touch a hot stove.

Introductory Tutorial on Commonsense Reasoning: Maarten Sap, Vered Shwartz, Antoine Bosselut, Dan Roth, and Yejin Choi. ACL 2020.
What is Commonsense?

The basic level of practical knowledge and reasoning concerning everyday situations and events that are commonly shared among most people.

- It’s a bad idea to touch a hot stove.
- It’s impolite to comment on people’s weight.

Introductory Tutorial on Commonsense Reasoning. Maarten Sap, Vered Shwartz, Antoine Bosselut, Dan Roth, and Yejin Choi. ACL 2020.
What is Commonsense?

The basic level of **practical knowledge** and **reasoning** concerning **everyday situations** and **events** that are **commonly** shared among **most** people.

- It’s a bad idea to touch a hot stove.
- It’s impolite to comment on people’s weight.
- Eating dinner comes before going to bed.

*Introductory Tutorial on Commonsense Reasoning* Maarten Sap, **Vered Shwartz**, Antoine Bosselut, Dan Roth, and Yejin Choi. ACL 2020.
What is Commonsense?

The basic level of **practical knowledge** and **reasoning** concerning **everyday situations** and **events** that are **commonly** shared among **most** people.

- It’s a bad idea to touch a hot stove.
- It’s impolite to comment on people’s weight.
- Eating dinner comes before going to bed.
- ...

*Introductory Tutorial on Commonsense Reasoning*: Maarten Sap, Vered Shwartz, Antoine Bosselut, Dan Roth, and Yejin Choi. ACL 2020.
In this Talk
In this Talk

Introspective knowledge acquisition through asking questions

EMNLP 2020 (a)
In this Talk

Introspective **knowledge** acquisition through asking questions

- EMNLP 2020 (a)

Nonmonotonic **reasoning** in natural **language**

- EMNLP 2020 (b)
- Findings of EMNLP 2020
- AAAI 2021
In this Talk

Introspective knowledge acquisition through asking questions

Nonmonotonic reasoning in natural language

Future Directions
Introspective knowledge acquisition through asking questions

Nonmonotonic reasoning in natural language

Future Directions
Unsupervised Commonsense Question Answering with Self-Talk

Vered Shwartz, Peter West, Ronan Le Bras, Chandra Bhagavatula and Yejin Choi

EMNLP 2020
Reasoning with Implicit Knowledge

Children need to eat more vegetables because they are healthy
Reasoning with Implicit Knowledge

Children need to eat more vegetables because they are healthy

children

vegetables
Reasoning with Implicit Knowledge

Children need to eat more vegetables because they are healthy.

1. Vegetables are healthy.
Reasoning with Implicit Knowledge

Children need to eat more vegetables because they are healthy

1. Vegetables are healthy.
2. Eating vegetables can make you healthier.
Reasoning with Implicit Knowledge

Children need to eat more vegetables because they are healthy.

1. Vegetables are healthy.
2. Eating vegetables can make you healthier.
3. People want to be healthy.
Children need to eat more vegetables because they are healthy.

1. Vegetables are healthy.
2. Eating vegetables can make you healthier.
3. People want to be healthy.
Discovery Learning (Bruner, 1961)

Children need to eat more vegetables because they are healthy.
Discovery Learning (Bruner, 1961)

Children need to eat more vegetables because they are healthy.
Discovery Learning (Bruner, 1961)

Children need to eat more vegetables because they are healthy

Self-Inquiry: What are the properties of vegetables?

Learner
Discovery Learning (Bruner, 1961)

Children need to eat more vegetables because they are healthy

Self-Inquiry
What are the properties of vegetables?

Existing Knowledge
Vegetables are full of vitamins.

Learner

children
vegetables
Discovery Learning (Bruner, 1961)

Children need to eat more vegetables because they are healthy.

Self-Inquiry: What are the properties of vegetables?

Existing Knowledge: Vegetables are full of vitamins.

New Facts: children vegetables
The Self-Talk Paradigm

**Main Question**
Children need to eat more vegetables because they are healthy

**Nested QA**

- **Self-Inquiry**: What are the properties of vegetables?
- **Existing Knowledge**: Vegetables are full of vitamins.

**Main Answer**

- **children**
- **vegetables**

**Neural Language Model**
Children need to eat more vegetables because they are healthy.

Answer choices: children, vegetables

Predicted answer choice: vegetables
Children need to eat more vegetables because they are healthy.

Answer choices:
children, vegetables

Predicted answer choice: vegetables
Knowledge Discovery
Knowledge Discovery
Children need to eat more vegetables because they are healthy.
<table>
<thead>
<tr>
<th>Instance</th>
<th>Nested Question Prefix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children need to eat more vegetables because <em>they</em> are healthy.</td>
<td>What is the purpose of</td>
</tr>
</tbody>
</table>
Children need to eat more vegetables because they are healthy.
Knowledge Discovery

<table>
<thead>
<tr>
<th>Instance</th>
<th>Nested Question Prefix</th>
<th>Nested Answer Prefix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children need to eat more vegetables because they are healthy.</td>
<td>What is the purpose of vegetables?</td>
<td>The purpose of _________ is</td>
</tr>
<tr>
<td>Instance</td>
<td>Nested Question Prefix</td>
<td>Nested Answer Prefix</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Children need to eat more vegetables because they are healthy.</td>
<td>What is the purpose of</td>
<td>The purpose of</td>
</tr>
<tr>
<td></td>
<td>vegetables?</td>
<td>vegetables is</td>
</tr>
<tr>
<td>Instance</td>
<td>Nested Question Prefix</td>
<td>Nested Answer Prefix</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Children need to eat more vegetables because they are healthy.</td>
<td>What is the purpose of</td>
<td>The purpose of vegetables is to provide a good base of nutrients and energy.</td>
</tr>
</tbody>
</table>
The purpose of vegetables is to provide a good base of nutrients and energy.

Children need to eat more vegetables because they are healthy.

What is the purpose of vegetables?

The purpose of vegetables is to provide a good base of nutrients and energy.
Children need to eat more vegetables because they are healthy.

The purpose of vegetables is to provide a good base of nutrients and energy.

The definition of healthy is quality of life that is free of diseases.

The properties of being healthy are linked to the effects of exercise.
The purpose of vegetables is to provide a good base of nutrients and energy.
The definition of healthy is quality of life that is free of diseases.
The properties of being healthy are linked to the effects of exercise.
The purpose of vegetables is to provide a good base of nutrients and energy.

The definition of healthy is quality of life that is free of diseases.

The properties of being healthy are linked to the effects of exercise.
Language Models
Text Generation

Language Model
Language Models
Text Generation

The properties of vegetables are
Language Models
Text Generation

p(next)

Language Model

The properties of vegetables are
Language Models
Text Generation

The properties of vegetables are

Language Model

that

sampling

$p(\text{next})$
Children need to eat more vegetables because they are healthy.

Answer choices: children, vegetables

Output
Predicted answer choice: vegetables

Knowledge Discovery
What is the definition of healthy?

Question Answering
The purpose of vegetables is to provide a good base of nutrients and energy.
The properties of being healthy are linked to the effects of exercise.
The definition of healthy is quality of life that is free of diseases.
Question Answering
Question Answering
Children need to eat more vegetables because **children** are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.

---

Children need to eat more vegetables because **vegetables** are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.

---

Children need to eat more vegetables because **children** are healthy. The definition of healthy is quality of life that is free of diseases.

---

Children need to eat more vegetables because **vegetables** are healthy. The definition of healthy is quality of life that is free of diseases.
Question Answering

<table>
<thead>
<tr>
<th>children</th>
<th>vegetables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children need to eat more vegetables because <strong>children</strong> are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.</td>
<td>Children need to eat more vegetables because <strong>vegetables</strong> are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Children need to eat more vegetables because <strong>children</strong> are healthy. The definition of healthy is quality of life that is free of diseases.</td>
<td>Children need to eat more vegetables because <strong>vegetables</strong> are healthy. The definition of healthy is quality of life that is free of diseases.</td>
</tr>
</tbody>
</table>

Most plausible statement
<table>
<thead>
<tr>
<th>children</th>
<th>vegetables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children need to eat more vegetables because <em>children</em> are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.</td>
<td>Children need to eat more vegetables because <em>vegetables</em> are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Children need to eat more vegetables because <em>children</em> are healthy. The definition of healthy is quality of life that is free of diseases.</td>
<td>Children need to eat more vegetables because <em>vegetables</em> are healthy. The definition of healthy is quality of life that is free of diseases.</td>
</tr>
</tbody>
</table>

**Most plausible statement**  ➔  Statement with best language model score
<table>
<thead>
<tr>
<th>children</th>
<th>vegetables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children need to eat more vegetables because <strong>children</strong> are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.</td>
<td>Children need to eat more vegetables because <strong>vegetables</strong> are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Children need to eat more vegetables because <strong>children</strong> are healthy. The definition of healthy is quality of life that is free of diseases.</td>
<td>Children need to eat more vegetables because <strong>vegetables</strong> are healthy. The definition of healthy is quality of life that is free of diseases.</td>
</tr>
</tbody>
</table>

**Most plausible statement** ➔ **Statement with best language model score**
Question Answering

Children need to eat more vegetables because children are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.

Most plausible statement

Statement with best language model score

$p(\text{need} \mid \text{Children}) \cdot p(\text{to} \mid \text{Children need}) \cdot p(\text{energy} \mid \ldots) \cdot p(<\text{eos}> \mid \ldots)$
Question Answering

Most plausible statement

$$\text{score} = -\frac{1}{n} \log (p(\text{need} | \text{Children}) \cdot p(\text{to} | \text{Children need}) \cdot \ldots \cdot p(\text{energy} | \ldots) p(\text{eos} | \ldots))$$

Statement with best language model score

Children need to eat more vegetables because children are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.

Children need to eat more vegetables because vegetables are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.

Children need to eat more vegetables because children are healthy. The definition of healthy is quality of life that is free of diseases.

Children need to eat more vegetables because vegetables are healthy. The definition of healthy is quality of life that is free of diseases.
<table>
<thead>
<tr>
<th>children</th>
<th>vegetables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children need to eat more vegetables because <strong>children</strong> are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.</td>
<td>Children need to eat more vegetables because <strong>vegetables</strong> are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Children need to eat more vegetables because <strong>children</strong> are healthy. The definition of healthy is quality of life that is free of diseases.</td>
<td>Children need to eat more vegetables because <strong>vegetables</strong> are healthy. The definition of healthy is quality of life that is free of diseases.</td>
</tr>
</tbody>
</table>
Children need to eat more vegetables because **children** are healthy. The purpose of vegetables is to provide a good base of nutrients and energy.

... 

Children need to eat more vegetables because **vegetables** are healthy. The definition of healthy is quality of life that is free of diseases.
Input
Context:
Children need to eat more vegetables because they are healthy.
Answer choices:
children, vegetables

Output
Predicted answer choice: vegetables

Knowledge Discovery
What is the definition of healthy?

Question Answering
The purpose of vegetables is to provide a good base of nutrients and energy. The properties of being healthy are linked to the effects of exercise. The definition of healthy is quality of life that is free of diseases.

Answer with most plausible statement
Children need to eat more vegetables because they are healthy.

Predicted answer choice: vegetables

The purpose of vegetables is to provide a good base of nutrients and energy.

The properties of being healthy are linked to the effects of exercise.

The definition of healthy is quality of life that is free of diseases.

Answer with most plausible statement
Experiments

Tasks
- Physical
  - PIQA
- Causal
  - COPA
- Social
  - SocialIQa
- General
  - CommonsenseQA
  - WinoGrande
- Temporal
  - MC-TACO

Baselines
Children need to eat more vegetables because they are healthy.

Results

- Vegetables are required for eating vegetables.
- Eating vegetables is motivated by being healthy.
- Vegetables are healthy.
- Because the children wanted to live longer.
Experiments

Tasks
- Physical: PIQA
- Causal: COPA
- Social: SocialIQa
- General: CommonsenseQA, WinoGrande
- Temporal: MC-TACO

Baselines
- Children need to eat more vegetables because they are healthy.
- Vegetables are required for eating vegetables.
- Eating vegetables is motivated by being healthy.
- Vegetables are healthy.
- Because the children wanted to live longer.

Results
Experiments

Physical PIQA

Causal COPA

Social SocialIQa

General CommonsenseQA

Temporal MC-TACO

WinoGrande
Physical
PIQA

Causal
COPA

Social
SocialIQa

General
CommonsenseQA
WinoGrande

Temporal
MC-TACO
Commonsense Question Answering Tasks

- Physical PIQA
- Social SocialIQa
- Temporal MC-TACO
- Causal COPA
- General CommonsenseQA WinoGrande
Commonsense Question Answering Tasks

Social Interaction QA
In the school play, Robin played a hero in the struggle against angry villain. How would others feel as a result?

1) sorry for the villain.
2) **hopeful that Robin will succeed.**
3) like Robin should lose the fight.
Commonsense Question Answering Tasks

**Social Interaction QA**
In the school play, Robin played a hero in the struggle against an angry villain. How would others feel as a result?

1) sorry for the villain.
2) **hopeful that Robin will succeed.**
3) like Robin should lose the fight.

**Physical PIQA**

**Causal COPA**

**Social SocialIQa**

**General CommonsenseQA WinoGrande**

**Temporal MC-TACO**

**Choice of Plausible Alternatives**
The man broke his toe. What was the cause?
1) He got a hole in his sock.
2) **He dropped a hammer on his foot.**
Experiments

**Tasks**
- Physical: PIQA
- Causal: COPA
- Social: SocialIQa
- General: CommonsenseQA
- Temporal: MC-TACO

**Baselines**
- MC-TACO
- Temporal
- SocialIQa
- PIQA
- Physical
- CommonsenseQA
- WinoGrande
- General
- COPA
- Causal

**Tasks**
- Physical: PIQA
- Causal: COPA
- Social: SocialIQa
- General: CommonsenseQA
- Temporal: MC-TACO

**Results**

- Children need to eat more vegetables because they are healthy.
  - vegetables, healthy
  - Vegetables are required for eating vegetables.
  - Eating vegetables is motivated by being healthy.
  - Vegetables are healthy.
  - Because the children wanted to live longer.

- [Graph showing experimental results]
Baselines
Children need to eat more vegetables because they are healthy.
Baselines

Children need to eat more vegetables because they are healthy.
Children need to eat more vegetables because they are healthy.
Children need to eat more vegetables because they are healthy.

Baselines

No Inquiry

Expert Knowledge
Children need to eat more vegetables because they are healthy.

Vegetables are required for eating vegetables. Eating vegetables is motivated by being healthy.

Expert Knowledge
Children need to eat more vegetables because they are healthy.

Vegetables are required for eating vegetables.
Eating vegetables is motivated by being healthy.
Vegetables are healthy.
Children need to eat more vegetables because **they** are healthy.

Vegetables are required for eating vegetables.

Eating vegetables is motivated by being healthy.

Vegetables are healthy.

Because the children wanted to live longer.
Experiments

Tasks
- Physical PIQA
- Causal COPA
- Social SocialIQa
- General CommonsenseQA
- Temporal MC-TACO

Baselines

Results

Children need to eat more vegetables because they are healthy.
Vegetables are required for eating vegetables.
Eating vegetables is motivated by being healthy.
Vegetables are healthy.
Because the children wanted to live longer.
Results
Results

No Inquiry
Expert Knowledge
Self-Talk
Human
1. Nested QA improves performance
1. Nested QA improves performance
2. Self-Talk performs similarly to models with expert knowledge
1. Nested QA improves performance
2. Self-Talk performs similarly to models with expert knowledge
3. Gap from human performance
The purpose of vegetables is to provide a good base of nutrients and energy. The properties of being healthy are linked to the effects of exercise. The definition of healthy is quality of life that is free of diseases.
The purpose of vegetables is to provide a good base of nutrients and energy.
The properties of being healthy are linked to the effects of exercise.
The definition of healthy is quality of life that is free of diseases.
The purpose of vegetables is to provide a good base of nutrients and energy.

The properties of being healthy are linked to the effects of exercise.

The definition of healthy is quality of life that is free of diseases.

Input

Knowledge Discovery

Introspection

Output

Question Answering

Incorporating Knowledge

Measuring Plausibility

Interactive QA Paradigm

Future Directions
The purpose of vegetables is to provide a good base of nutrients and energy.
The properties of being healthy are linked to the effects of exercise.
The definition of healthy is quality of life that is free of diseases.

Can I answer the main question?
The purpose of vegetables is to provide a good base of nutrients and energy. The properties of being healthy are linked to the effects of exercise. The definition of healthy is quality of life that is free of diseases.
The purpose of vegetables is to provide a good base of nutrients and energy. The properties of being healthy are linked to the effects of exercise. The definition of healthy is quality of life that is free of diseases.

Can I answer the main question?

What knowledge is missing?

Relevant but unhelpful knowledge may distract the model!
The purpose of vegetables is to provide a good base of nutrients and energy. The properties of being healthy are linked to the effects of exercise. The definition of healthy is quality of life that is free of diseases.

Relevant but unhelpful knowledge may distract the model!
The purpose of vegetables is to provide a good base of nutrients and energy.
The properties of being healthy are linked to the effects of exercise.
The definition of healthy is quality of life that is free of diseases.

**Interactive QA Paradigm**

**Future Directions**

**Input**

- Knowledge Discovery
  - Introspection

**Output**

- Question Answering
  - Measuring Plausibility
  - Incorporating Knowledge
The purpose of vegetables is to provide a good base of nutrients and energy. The properties of being healthy are linked to the effects of exercise. The definition of healthy is quality of life that is free of diseases.

Which knowledge is the most relevant, correct & helpful?
The purpose of vegetables is to provide a good base of nutrients and energy. The properties of being healthy are linked to the effects of exercise. The definition of healthy is quality of life that is free of diseases.

How to incorporate the knowledge?

Which knowledge is the most relevant, correct & helpful?
The purpose of vegetables is to provide a good base of nutrients and energy.
The properties of being healthy are linked to the effects of exercise.
The definition of healthy is quality of life that is free of diseases.

How to incorporate the knowledge?
Which knowledge is the most relevant, correct & helpful?

Knowledge that helped the model is not always judged by people as helpful.
The purpose of vegetables is to provide a good base of nutrients and energy. The properties of being healthy are linked to the effects of exercise. The definition of healthy is quality of life that is free of diseases.
The purpose of vegetables is to provide a good base of nutrients and energy. The properties of being healthy are linked to the effects of exercise. The definition of healthy is quality of life that is free of diseases.

How to measure statements’ plausibility & correctness?
Introspective **knowledge** acquisition through asking questions

- EMNLP 2020 (a)

Nonmonotonic **reasoning** in natural **language**

- EMNLP 2020 (b)
- Findings of EMNLP 2020
- AAAI 2021

Future Directions
Nonmonotonic Reasoning
Nonmonotonic Reasoning

Abductive reasoning

ART (Bhagavatula et al., 2020)

Most plausible explanation
Nonmonotonic Reasoning

Abductive reasoning
ART (Bhagavatula et al., 2020)
Most plausible explanation

Counterfactual reasoning
TimeTravel (Qin et al., 2019)
What if?
Nonmonotonic Reasoning

Abductive reasoning
- ART (Bhagavatula et al., 2020)
  - Most plausible explanation

Counterfactual reasoning
- TimeTravel (Qin et al., 2019)
  - What if?

Defeasible reasoning
- δ-NLI (Rudinger et al., 2020)
  - Updating inferences with additional knowledge
Back to the Future:
Unsupervised Backprop-based Decoding for Counterfactual and Abductive Commonsense Reasoning

Lianhui (Karen) Qin, Vered Shwartz, Peter West, Chandra Bhagavatula, Jena Hwang, Ronan Le Bras, Antoine Bosselut, and Yejin Choi.
EMNLP 2020
Abductive Reasoning

Reason about the most plausible explanation for incomplete observations.
Abductive Reasoning

Reason about the most plausible explanation for incomplete observations.

Sara wanted to make dinner for some guests.
Abductive Reasoning

Reason about the most plausible explanation for incomplete observations.

Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Abductive Reasoning

Reason about the most plausible explanation for incomplete observations.

Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Abductive Reasoning

Reason about the most plausible explanation for incomplete observations.

Sara wanted to make dinner for some guests. But she didn’t know how to cook. She had to order pizza for her friends instead.
**Challenge:** Language models are conditioned only on a past context

Sara wanted to make dinner for some guests.
**Challenge:** Language models are conditioned only on a past context

Sara wanted to make dinner for some guests. "I'm going to go grab some rice noodles," she says.
Challenge: Language models are conditioned only on a past context

Sara wanted to make dinner for some guests. "I'm going to go grab some rice noodles," she says.

Solution: compute loss w.r.t future constraints & backpropagate to the output
Challenge: Language models are conditioned only on a past context

Sara wanted to make dinner for some guests. "I'm going to go grab some rice noodles," she says.

Solution: compute loss w.r.t future constraints & backpropagate to the output

Inspiration:

Image Style Transfer
(Gatys et al, 2016)
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
DELOREAN:
DEcoding for nonmonotonic LOgical REAsoNing

Input

X - past context
Sara wanted to make dinner for some guests.
Z - future constraints
She had to order pizza for her friends instead.

Output

Y - continuation
• Fluent continuation of X
• Satisfies the constraints Z
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Initialization
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Backward Pass
Backward Pass
Sara wanted to make dinner for some guests.

She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests.

She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.

\[
\tilde{Y} = \text{decode N tokens from } LM_{\text{forward}}(X)
\]

Maximize the likelihood of LM to generate the future observation \( Z \) following the past observation \( X \) and the generated hypothesis \( \tilde{Y} \)

\[
\mathcal{L}(X, \tilde{Y}, Z) := -\sum_{n=1}^{N_Z} \log P_{LM}(\tilde{z}_n | X, \tilde{Y}, Z_{1:n-1})
\]
Forward Pass
Forward Pass
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.

**Initialization**
\[ \tilde{Y} = \text{decode N tokens from } LM_{\text{forward}}(X) \]

**Backward Pass**
Maximize the likelihood of LM to generate the future observation \( Z \) following the past observation \( X \) and the generated hypothesis \( \tilde{Y} \)
\[ \mathcal{L}(X, \tilde{Y}, Z) := -\sum_{n=1}^{N_2} \log P_{LM}(\tilde{z}_n | X, \tilde{Y}, Z_{1:n-1}) \]

**Forward Pass**
\[ \tilde{Y} = \text{decode N tokens from } LM_{\text{forward}}(X) \]
And mix with backward logits

**Output**
\( Y \) - continuation
- Fluent continuation of \( X \)
- Satisfies the constraints \( Z \)
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.

Output

\( Y \) - continuation
- Fluent continuation of \( X \)
- Satisfies the constraints \( Z \)

Initialization

\( \tilde{Y} = \) decode \( N \) tokens from \( LM_{\text{forward}}(X) \)

Backward Pass

Maximize the likelihood of \( LM \) to generate the future observation \( Z \) following the past observation \( X \) and the generated hypothesis \( \tilde{Y} \)

\[
\mathcal{L}(X, \tilde{Y}, Z) := - \sum_{n=1}^{N_Z} \log P_{LM}(\tilde{z}_n | X, \tilde{Y}, Z_{1:n-1})
\]

Forward Pass

\( \tilde{Y} = \) decode \( N \) tokens from \( LM_{\text{forward}}(X) \)
And mix with backward logits

DELOREAN
Generation
Generation
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. But she didn’t know how to cook. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.

**Input**
- **X** - past context
- **Z** - future constraints

**Output**
- **Y** - continuation
  - Fluent continuation of X
  - Satisfies the constraints Z

**Initialization**
\[ \tilde{Y} = \text{decode N tokens from } LM_{\text{forward}}(X) \]

**Backward Pass**
Maximize the likelihood of LM to generate the future observation Z following the past observation X and the generated hypothesis \( \tilde{Y} \)
\[ \mathcal{L}(X, \tilde{Y}, Z) := - \sum_{n=1}^{N_z} \log P_{LM}(\tilde{z}_n | X, \tilde{Y}, Z_{1:n-1}) \]

**Forward Pass**
\[ \tilde{Y} = \text{decode N tokens from } LM_{\text{forward}}(X) \]
And mix with backward logits

**Greedy decoding from** \( \tilde{Y} \)
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Sara wanted to make dinner for some guests. She had to order pizza for her friends instead.
Select Best Y
Select Best Y
Select Best $Y$

Select $Y^{(i)}$ that is most likely to follow and precede its adjacent sentences
Select Best $Y$}

Select $Y^{(i)}$ that is most likely to follow and precede its adjacent sentences

$$\text{score}(Y^{(i)}) = \text{BERT}_{\text{NSP}}(XY^{(i)}, Z) + \text{BERT}_{\text{NSP}}(X, Y^{(i)}Z)$$

P(But she didn’t know how to cook. She had to order pizza for her friends instead. | Sara wanted to make dinner for some guests.)

P(She had to order pizza for her friends instead. | Sara wanted to make dinner for some guests. But she didn’t know how to cook.)
Human Evaluation Results

Abductive Reasoning
Human Evaluation Results

Abductive Reasoning

Coherence

X-Y  Y-Z  XY-Z

- DELOREAN
- Unsupervised
- Supervised
- Human
Human Evaluation Results

Abductive Reasoning

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Coherence</th>
</tr>
</thead>
<tbody>
<tr>
<td>XY</td>
<td>5.22</td>
</tr>
<tr>
<td>Y-Z</td>
<td>5.25</td>
</tr>
<tr>
<td>XY-Z</td>
<td>2.97</td>
</tr>
</tbody>
</table>

- DELOREAN
- Unsupervised
- Supervised
- Human
Human Evaluation Results
Abductive Reasoning

<table>
<thead>
<tr>
<th></th>
<th>DELOREAN</th>
<th>Unsupervised</th>
<th>Supervised</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-Y</td>
<td>5.22</td>
<td>4.74</td>
<td>2.36</td>
<td>2.36</td>
</tr>
<tr>
<td>Y-Z</td>
<td>5.25</td>
<td>2.58</td>
<td>2.97</td>
<td>2.38</td>
</tr>
<tr>
<td>X-Y-Z</td>
<td>2.97</td>
<td>2.56</td>
<td>3.25</td>
<td>5.22</td>
</tr>
</tbody>
</table>

1. Outperforms unsupervised models substantially
Human Evaluation Results

Abductive Reasoning

1. Outperforms unsupervised models substantially
2. Competitive with supervised methods!
Human Evaluation Results

Abductive Reasoning

1. Outperforms unsupervised models substantially
2. Competitive with supervised methods!
3. Large gap from human performance
Example Generations

<table>
<thead>
<tr>
<th>t</th>
<th>Generated Y</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>She was thinking about the best way.</td>
<td>2.95</td>
</tr>
<tr>
<td>2</td>
<td>However, her cooking skills were the only thing that could make it a success.</td>
<td>7.73</td>
</tr>
<tr>
<td>3</td>
<td>But she couldn’t, because she was too busy with her work.</td>
<td>8.81</td>
</tr>
<tr>
<td>4</td>
<td>But she didn’t have the money and she didn’t have her own kitchen.</td>
<td>8.08</td>
</tr>
<tr>
<td>5</td>
<td>But she didn’t know how to cook.</td>
<td><strong>9.05</strong></td>
</tr>
</tbody>
</table>
Example Generations

<table>
<thead>
<tr>
<th>t</th>
<th>Generated Y</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>She was thinking about the best way.</td>
<td>2.95</td>
</tr>
<tr>
<td>2</td>
<td>However, her cooking skills were the only thing that could make it a success.</td>
<td>7.73</td>
</tr>
<tr>
<td>3</td>
<td>But she couldn’t, because she was too busy with her work.</td>
<td>8.81</td>
</tr>
<tr>
<td>4</td>
<td>But she didn’t have the money and she didn’t have her own kitchen.</td>
<td>8.08</td>
</tr>
<tr>
<td>5</td>
<td>But she didn’t know how to cook.</td>
<td><strong>9.05</strong></td>
</tr>
</tbody>
</table>
Counterfactual Reasoning
Counterfactual Reasoning

**Input**
X - counterfactual past
S1: Lisa was throwing a Halloween party.
S2: All her friends were dressing up.
S2’: It was a Game of Thrones themed party.

**Z - minimally edit original future**
S3: Lisa thought about being a wizard.
S4: Then she decided on a scarier costume.
S5: Lisa dressed up like a vampire.
Counterfactual Reasoning

**Input**

X - counterfactual past

S1: Lisa was throwing a Halloween party.
S2: All her friends were dressing up.
S2': It was a Game of Thrones themed party.

Z - minimally edit original future

S3: Lisa thought about being a wizard.
S4: Then she decided on a scarier costume.
S5: Lisa dressed up like a vampire.

**Output Y:**

1. Adheres to the revised story beginning
2. Minimally edits the original ending
Counterfactual Reasoning

**Input**

X - counterfactual past

- S1: Lisa was throwing a Halloween party.
- S2: All her friends were dressing up.
- S2': It was a Game of Thrones themed party.

Z - minimally edit original future

- S3: Lisa thought about being a wizard.
- S4: Then she decided on a scarier costume.
- S5: Lisa dressed up like a vampire.

**Output Y:**

1. Adheres to the revised story beginning
2. Minimally edits the original ending
**Counterfactual Reasoning**

**Input**

X - counterfactual past

S1: Lisa was throwing a Halloween party.
S2: All her friends were dressing up.
S2∗: It was a Game of Thrones themed party.

Z - minimally edit original future

S3: Lisa thought about being a wizard.
S4: Then she decided on a scarier costume.
S5: Lisa dressed up like a vampire.

**Output Y:**

1. Adheres to the revised story beginning
2. Minimally edits the original ending

**Initialization**

**Backward Pass**

Minimize the KL divergence between the original ending Z (one-hot representation) and generated ending Ŷ

\[ \mathcal{L}(X, Ŷ, Z) := \text{KL}(Z \parallel \text{softmax}(Ŷ/τ)) \]
Counterfactual Reasoning

Input

X - counterfactual past

S1: Lisa was throwing a Halloween party.
S2: All her friends were dressing up.
S2': It was a Game of Thrones themed party.

Z - minimally edit original future

S3: Lisa thought about being a wizard.
S4: Then she decided on a scarier costume.
S5: Lisa dressed up like a vampire.

Output Y:

1. Adheres to the revised story beginning
2. Minimally edits the original ending

Generation + Select best Y

Initialization

Backward Pass

Minimize the KL divergence between the original ending \(Z\) (one-hot representation) and generated ending \(\tilde{Y}\)

\[ \mathcal{L}(X, \tilde{Y}, Z) := \text{KL} (Z \parallel \text{softmax}(\tilde{Y}/\tau)) \]

Forward Pass
Counterfactual Reasoning

**Input**

**X - counterfactual past**

S1: Lisa was throwing a Halloween party.
S2: All her friends were dressing up.
S2': It was a Game of Thrones themed party.

**Z - minimally edit original future**

S3: Lisa thought about being a wizard.
S4: Then she decided on a scarier costume.
S5: Lisa dressed up like a vampire.

**Output Y:**

1. Adheres to the revised story beginning
2. Minimally edits the original ending

**Initialization**

**Backward Pass**

Minimize the KL divergence between the original ending $Z$ (one-hot representation) and generated ending $\tilde{Y}$

$$\mathcal{L}(X, \tilde{Y}, Z) := KL(Z || softmax(\tilde{Y}/\tau))$$

**Forward Pass**

**Generation + Select best Y**

S3': Lisa thought about how she would dress up as a Lannister, but she didn’t want to look like a Lannister.
S4': She wanted to look like a Stark.
S5': Lisa dressed up like a Stark.
Counterfactual Reasoning

**Input**

X - counterfactual past

S1: Lisa was throwing a Halloween party.
S2: All her friends were dressing up.
S2': It was a Game of Thrones themed party.

Z - minimally edit original future

S3: Lisa thought about being a wizard.
S4: Then she decided on a scarier costume.
S5: Lisa dressed up like a vampire.

**Output Y:**

1. Adheres to the revised story beginning
2. Minimally edits the original ending

**Initialization**

**Backward Pass**

Minimize the KL divergence between the original ending $Z$ (one-hot representation) and generated ending $\tilde{Y}$.

$\mathcal{L}(X, \tilde{Y}, Z) := \text{KL}(Z \| \text{softmax}(\tilde{Y}/\tau))$

**Forward Pass**

Generation + Select best Y

DeLorean was the only method to achieves a good balance between the two requirements.
Takeaways

Back to the Future: Unsupervised Backprop-based Decoding for Counterfactual and Abductive Commonsense Reasoning
Takeaways

Back to the Future: Unsupervised Backprop-based Decoding for Counterfactual and Abductive Commonsense Reasoning

- Unsupervised approach to generate text conditioned on past and future constraints.
Takeaways

Back to the Future: Unsupervised Backprop-based Decoding for Counterfactual and Abductive Commonsense Reasoning

☑️ Unsupervised approach to generate text conditioned on past and future constraints.

☑️ Effective for generative nonmonotonic reasoning tasks.
Takeaways

Back to the Future: Unsupervised Backprop-based Decoding for Counterfactual and Abductive Commonsense Reasoning

- Unsupervised approach to generate text conditioned on past and future constraints.
- Effective for generative nonmonotonic reasoning tasks.
- Easy adaptation for other generative reasoning tasks.
Takeaways

Back to the Future: Unsupervised Backprop-based Decoding for Counterfactual and Abductive Commonsense Reasoning

- Unsupervised approach to generate text conditioned on past and future constraints.
- Effective for generative nonmonotonic reasoning tasks.
- Easy adaptation for other generative reasoning tasks.
- More work on nonmonotonic reasoning!
Defeasible Inference (Reiter, 1980)

Given premise $P$, a hypothesis $H$ is **defeasible** if there exists an update $U$ (consistent with $P$) such that a human would find $H$ less likely to be true after learning $U$. 
Defeasible Inference (Reiter, 1980)

Given premise P, a hypothesis H is **defeasible** if there exists an update U (consistent with P) such that a human would find H less likely to be true after learning U.
Defeasible Inference (Reiter, 1980)

Given premise $P$, a hypothesis $H$ is **defeasible** if there exists an update $U$ (consistent with $P$) such that a human would find $H$ less likely to be true after learning $U$. 

P: Tweety is a bird.
Defeasible Inference (Reiter, 1980)

Given premise P, a hypothesis H is **defeasible** if there exists an update U (consistent with P) such that a human would find H less likely to be true after learning U.

P: Tweety is a bird.

H: Tweety flies.
Defeasible Inference (Reiter, 1980)

Given premise P, a hypothesis H is **defeasible** if there exists an update U (consistent with P) such that a human would find H less likely to be true after learning U.

P: Tweety is a bird.

H: Tweety flies.

U: Tweety is a penguin.
Thinking Like a Skeptic: Defeasible Inference in Natural Language

Rachel Rudinger, Vered Shwartz, Jena D. Hwang, Chandra Bhagavatula, Maxwell Forbes, Ronan Le Bras, Noah A. Smith, and Yejin Choi.
Findings of EMNLP 2020
Defeasible Inference in Natural Language

An update $U$ is called a **weaken**er if, given a premise $P$ and hypothesis $H$, a human would most likely find $H$ *less likely to be true* after learning $U$; if they would find $H$ *more likely to be true*, then we call $U$ a **strengthen**er.

$P$: Tweety is a bird.

$H$: Tweety flies.

Weakener: Tweety is a penguin.
Defeasible Inference in Natural Language

An update $U$ is called a **weaken**er if, given a premise $P$ and hypothesis $H$, a human would most likely find $H$ *less likely to be true* after learning $U$; if they would find $H$ *more likely to be true*, then we call $U$ a **strengthener**.

- $P$: Tweety is a bird.
- Weakener: Tweety is a penguin.
- Strengthener: Tweety is on a tree.
Defeasible Inference in Natural Language
Defeasible Inference in Natural Language

Discriminative Task

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them. They have a work meeting.

- They are in a conference room. $\rightarrow$ + Strengthener
- They are in a library. $\rightarrow$ - Weakenor
Defeasible Inference in Natural Language

Discriminative Task

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them. They have a work meeting.

They are in a conference room. + Strengthener

They are in a library. - Weakenener

Generative Task

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them. They have a work meeting.

They are in a conference room. +

They are in a library. -
Defeasible Inference in Natural Language

**Discriminative Task**

- **Strengthener**: They are in a conference room.
- **Weakener**: They are in a library.

**Generative Task**

- They have a work meeting.

Language models leave plenty of room for improvement on the generative task!
Learning to Rationalize for Nonmonotonic Reasoning with Distant Supervision

Faeze Brahman, Vered Shwartz, Rachel Rudinger, and Yejin Choi

AAAI 2021
Rationale Generation for Defeasible Inference
Rationale Generation for Defeasible Inference

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them.

They have a work meeting.

+ They are in a conference room.

A conference room is where people have meetings at work.

- They are in a library.

You must be quiet in the library, while work meetings involve talking.
Rationale Generation for Defeasible Inference

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them.

They have a work meeting.

They are in a conference room.

A conference room is where people have meetings at work.

They are in a library.

You must be quiet in the library, while work meetings involve talking.

Distant supervision:
Rationale Generation for Defeasible Inference

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them. They have a work meeting.

They are in a conference room. A conference room is where people have meetings at work.

They are in a library. You must be quiet in the library, while work meetings involve talking.

Distant supervision:

The definition of a library is...

The definition of a library is...
Rationale Generation for Defeasible Inference
Rationale Generation for Defeasible Inference

Post hoc Rationalization

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them.

They have a work meeting.

+ They are in a conference room. → A conference room is where people have meetings at work.

- They are in a library. → You must be quiet in the library, while work meetings involve talking.
Rationale Generation for Defeasible Inference

Post hoc Rationalization

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them.

They have a work meeting.

- They are in a library.

A conference room is where people have meetings at work.

You must be quiet in the library, while work meetings involve talking.

Trivially rephrasing the label! (“[+] implies that [H]”)
Rationale Generation for Defeasible Inference

Post hoc Rationalization

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them.

They have a work meeting.

+ They are in a conference room. → A conference room is where people have meetings at work.

- They are in a library. → You must be quiet in the library, while work meetings involve talking.

Joint Prediction & Rationalization

A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them.

They have a work meeting.

They are in a conference room. → + A conference room is where people have meetings at work.

They are in a library. → - You must be quiet in the library, while work meetings involve talking.

Trivially rephrasing the label! ("[+] implies that [H]")
Rationale Generation for Defeasible Inference

**Post hoc Rationalization**

- A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them. They have a work meeting.
- They are in a conference room. → A conference room is where people have meetings at work.
- They are in a library. → You must be quiet in the library, while work meetings involve talking.

**Joint Prediction & Rationalization**

- A group of people sitting around a rectangular table having either pieces of paper or laptops in front of them. They have a work meeting.
- They are in a conference room. → A conference room is where people have meetings at work.
- They are in a library. → You must be quiet in the library, while work meetings involve talking.

- Trivially rephrasing the label! (“[+] implies that [H]”)
Introspective knowledge acquisition through asking questions

Nonmonotonic reasoning in natural language

Future Directions
Future Directions

- **Knowledge**: Acquiring commonsense knowledge
- **Language**: Meaningful text representations
  - Learning to read between the lines: figurative language, implicatures, pragmatics
- **Reasoning**: Causal & Nonmonotonic reasoning
Future Directions

Knowledge
- Acquiring commonsense knowledge

Language
- Meaningful text representations
- Learning to read between the lines: figurative language, implicatures, pragmatics

Reasoning
- Causal & Nonmonotonic reasoning
Acquiring Commonsense Knowledge
Acquiring Commonsense Knowledge

1 from people

2 from text

× Impossible to manually enumerate

$ $$
Acquiring Commonsense Knowledge

1. From people
   - Impossible to manually enumerate
   - Reporting bias
     (Gordon and Van Durme, 2013)

2. From text

Graph:
- Murdered + killed
- Breathed + exhaled + inhaled
Acquiring Commonsense Knowledge

1. from people
   - Impossible to manually enumerate
   - Reporting bias
     (Gordon and Van Durme, 2013)

2. from text

3. from large-scale neural language models

- murdered + killed
- breathed + exhaled + inhaled
Acquiring Commonsense Knowledge from Large-scale Neural Language models
Acquiring Commonsense Knowledge from Large-scale Neural Language models

Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)

- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)
- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)
- Don’t differentiate constant vs. contingent facts

Zebras are black and white. My shirt is blue / red.
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)

- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)

- Don’t differentiate constant vs. contingent facts

- Don’t differentiate generic facts from grounded knowledge about named entities

"You are grounded!": Latent Name Artifacts in Pre-trained Language Models. Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. EMNLP 2020.
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)

- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)

- Don’t differentiate constant vs. contingent facts

- Don’t differentiate generic facts from grounded knowledge about named entities

"You are grounded!": Latent Name Artifacts in Pre-trained Language Models. Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. EMNLP 2020.
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)
- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)
- Don’t differentiate constant vs. contingent facts

Don’t differentiate generic facts from grounded knowledge about named entities

Richard has a bad habit of saying things that are not true.

Zebras are black and white. My shirt is blue / red.

"You are grounded!": Latent Name Artifacts in Pre-trained Language Models. Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. EMNLP 2020.
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)
- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)
- Don’t differentiate constant vs. contingent facts

- Don’t differentiate generic facts from grounded knowledge about named entities

"You are grounded!": Latent Name Artifacts in Pre-trained Language Models. Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. EMNLP 2020.
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)
- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)
- Don’t differentiate constant vs. contingent facts

Don’t differentiate generic facts from grounded knowledge about named entities

Richard has a bad habit of saying things that are not true.
Donald has a bad

Zebras are black and white. My shirt is blue / red.

"You are grounded!": Latent Name Artifacts in Pre-trained Language Models. Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. EMNLP 2020.
Acquiring Commonsense Knowledge from Large-scale Neural Language models

Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)

Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)

Don’t differentiate constant vs. contingent facts

Don’t differentiate generic facts from grounded knowledge about named entities

Richard has a bad habit of saying things that are not true.

Donald has a bad reputation for being a racist.

Zebras are black and white. My shirt is blue / red.

"You are grounded!": Latent Name Artifacts in Pre-trained Language Models. Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. EMNLP 2020.
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)

- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)

- Don’t differentiate constant vs. contingent facts

- Don’t differentiate generic facts from grounded knowledge about named entities

"You are grounded!": Latent Name Artifacts in Pre-trained Language Models. Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. EMNLP 2020.
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)
- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)
- Don’t differentiate constant vs. contingent facts

- Don’t differentiate generic facts from grounded knowledge about named entities

- Don’t completely overcome reporting bias

"You are grounded!": Latent Name Artifacts in Pre-trained Language Models. Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. EMNLP 2020.
Acquiring Commonsense Knowledge from Large-scale Neural Language models

- Capture facts not explicitly mentioned in the corpus (Petroni et al. 2019; Feldman et al. 2019, this talk)
- Not sensitive to negation (Kassner et al. 2020; Ettinger, 2020)
- Don’t differentiate constant vs. contingent facts

- Don’t differentiate generic facts from grounded knowledge about named entities
- Don’t completely overcome reporting bias

"You are grounded!": Latent Name Artifacts in Pre-trained Language Models. Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. EMNLP 2020.
Does GPT-3 have commonsense?

Examples by Gary Marcus. GPT-3 is not publicly available.
Does GPT-3 have commonsense?

You need flour to bake bread.
You have a sack of flour in the garage.
When you get there, you find that it got thoroughly soaked in a heavy rain last night.

So you have to
Does GPT-3 have commonsense?

You need flour to bake bread.
You have a sack of flour in the garage.
When you get there, you find that it got thoroughly soaked in a heavy rain last night.

So you have to dry it out before you can use it. You can do this by spreading it out on a table and putting a fan on it.
Does GPT-3 have commonsense?

You need flour to bake bread.
You have a sack of flour in the garage.
When you get there, you find that it got thoroughly soaked in a heavy rain last night.

So you have to **dry it out before you can use it. You can do this by spreading it out on a table and putting a fan on it.**

Examples by Gary Marcus. GPT-3 is not publicly available.
Does GPT-3 have commonsense?

You need flour to bake bread.
You have a sack of flour in the garage.
When you get there, you find that it got thoroughly soaked in a heavy rain last night.

So you have to dry it out before you can use it. You can do this by spreading it out on a table and putting a fan on it.

A cat was sitting next to a mouse hole in the kitchen wall waiting for the mouse, but it was too cautious, so

Examples by Gary Marcus. GPT-3 is not publicly available.
Does GPT-3 have commonsense?

You need flour to bake bread.
You have a sack of flour in the garage.
When you get there, you find that it got thoroughly soaked in a heavy rain last night.

So you have to dry it out before you can use it. You can do this by spreading it out on a table and putting a fan on it.

A cat was sitting next to a mouse hole in the kitchen wall waiting for the mouse,
but it was too cautious, so the cat was waiting for a long time. The cat was getting hungry, so it was thinking of going to the market to buy some food.

Examples by Gary Marcus. GPT-3 is not publicly available.
Does GPT-3 have commonsense?

You need flour to bake bread.
You have a sack of flour in the garage.
When you get there, you find that it got thoroughly soaked in a heavy rain last night.

So you have to dry it out before you can use it. You can do this by spreading it out on a table and putting a fan on it.

A cat was sitting next to a mouse hole in the kitchen wall waiting for the mouse, but it was too cautious, so the cat was waiting for a long time. The cat was getting hungry, so it was thinking of going to the market to buy some food.

Examples by Gary Marcus. GPT-3 is not publicly available.
Examples by Gary Marcus.

**Does GPT-3 have commonsense?**

- **You need flour to bake bread.**
  - You have a sack of flour in the garage.
  - When you get there, you find that it got thoroughly soaked in a heavy rain last night.
  - So you have to dry it out before you can use it. You can do this by spreading it out on a table and putting a fan on it.

- **A cat was sitting next to a mouse hole in the kitchen wall waiting for the mouse,** the cat was waiting for a long time. The cat was getting hungry, so it was thinking of going to the market to buy some food.
  - but it was too cautious, so...
Acquiring Commonsense Knowledge

Reporting Bias
Acquiring Commonsense Knowledge

Learning Commonsense Knowledge from Text

<<

from Text, Images and Videos
Acquiring Commonsense Knowledge

Learning Commonsense Knowledge from Text

from Text, Images and Videos

last row ⇒ standing

front row ⇒ cross-legged
Acquiring Commonsense Knowledge

Learning Commonsense Knowledge from Text

Reporting Bias!

Hanging up the phone without saying goodbye

Reporting Bias!
Acquiring Commonsense Knowledge

Learning Commonsense Knowledge from Text

Reporting Bias!

Hanging up the phone without saying goodbye
Future Directions

Knowledge
- Acquiring commonsense knowledge

Language
- Meaningful text representations
- Learning to read between the lines: figurative language, implicatures, pragmatics

Reasoning
- Causal & Nonmonotonic reasoning
Meaningful Text Representations

Relatedness
Meaningful Text Representations

Relatedness $\rightarrow$ Finer-grained Semantic Relations
I believe he would come here.

I doubt he would come here.
Future Directions

Knowledge
Acquiring commonsense knowledge

Language
More accurate text representations

Reasoning
Causal & Nonmonotonic reasoning

Learning to read between the lines:
figurative language, implicatures, pragmatics
Learning to Read between the Lines

Required language understanding skills:
Learning to Read between the Lines

Required language understanding skills:

(i) Implicit meaning
Learning to Read between the Lines

Required language understanding skills:

(1) Implicit meaning
(2) Non-literal meaning
Learning to Read between the Lines

I didn’t eat anything since the morning.

Me too. I could eat a horse!

Do you have some food here?

Required language understanding skills:

1. Implicit meaning
2. Non-literal meaning
3. Pragmatics
Learning to Read between the Lines

**Required language understanding skills:**

1. Implicit meaning
2. Non-literal meaning
3. Pragmatics
4. Common background
Learning to Read between the Lines

Required language understanding skills:

1. Implicit meaning
2. Non-literal meaning
3. Pragmatics
4. Common background

Future Directions

Knowledge
- Acquiring commonsense knowledge

Language
- Meaningful text representations

Reasoning
- Causal & Nonmonotonic reasoning

Learning to read between the lines:
- figurative language, implicatures, pragmatics
Future Directions

Knowledge
- Acquiring commonsense knowledge

Language
- Meaningful text representations

Reasoning
- Causal & Nonmonotonic reasoning

Learning to read between the lines:
- Figurative language, implicatures, pragmatics

Thank you! Questions?

vereds@allenai.org

@VeredShwartz
References

(1) Vered Shwartz, Rachel Rudinger, and Oyvind Tafjord. "You are grounded!: Latent Name Artifacts in Pre-trained Language Models." EMNLP 2020.