Computational Rationality in LLMs

CS188, Fall 2023
Nicholas Tomlin
Overview

Crossword Solving

Detecting LLMs

Decision-Making in NL
Overview

Crossword Solving

Constraint satisfaction, belief propagation, greedy search, iterative improvement

Detecting LLMs

Classification, feature selection, generalization and overfitting

Decision-Making in NL

Constraint satisfaction, value of information, decision networks, particle filtering
Overview

Crossword Solving

Constraint satisfaction, belief propagation, greedy search, iterative improvement

Detecting LLMs

Classification, feature selection, generalization and overfitting

Decision-Making in NL

Constraint satisfaction, value of information, decision networks, particle filtering
Example Crossword


Across
1  Regarding
6  Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba

Down
1  Home to 41-Down
2  Convince
3  "I wanna know all the details"
4  Start of a modern inquiry
5  Drink similar to sarsaparilla
6  Co-star of 1984's "Ghostbusters"
7  Right on
8  Aces with aces?
9  Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
19 Actress Shawkat of " Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier
Example Crossword

Across

1. Regarding
2. Take back, in a way
3. Start of an aside
4. Omega competitor
5. Something hitting a nerve?
6. Papyrus, e.g.
7. "Take me with you!"
8. Begin flirting with someone, so to speak
9. Assign
10. Hoth, in "Star Wars"
11. ___ rule
12. They don’t hold water
13. Feudal figure
14. Panegyrical, e.g.
15. Sci-fi enemy collective, perhaps
16. Ones born beginning in the early 2010s
17. They might cut to the chase
18. Electrically balanced, in chemistry
19. Leaders at the Kaaba

Down

1. Home to 41-Down
2. Convince
3. "I wanna know all the details"
4. Start of a modern inquiry
5. Drink similar to sarsaparilla
6. Co-star of 1984’s "Ghostbusters"
7. Right on
8. Aces with aces?
9. Like some households
10. Designer with an eponymous hotel in the Burj Khalifa
11. Currency units in West Africa
12. Grievous
13. "Golden Boy" playwright
14. Actress Shawkat of "Arrested Development"
15. Assignment for an anchor
16. "___ noticed ..."
17. Great Hindu sage
18. Late-Triassic flier

Example Crossword

Across

1. Regarding
6. Take back, in a way
10. Start of an aside
14. Omega competitor
15. Something hitting a nerve?
16. Papyrus, e.g.
17. "Take me with you!"
18. Begin flirting with someone, so to speak
20. Assign
21. Hoth, in "Star Wars"
22. ___ rule
23. They don't hold water
24. Feudal figure
26. Panegyrical, e.g.
27. Sci-fi enemy collective, perhaps
31. Ones born beginning in the early 2010s
37. They might cut to the chase
38. Electrically balanced, in chemistry
42. Leaders at the Kaaba

Down

1. Home to 41-Down
2. Convince
3. "I wanna know all the details"
4. Start of a modern inquiry
5. Drink similar to sarsaparilla
6. Co-star of 1984's "Ghostbusters"
7. Right on
8. Aces with aces?
9. Like some households
10. Designer with an eponymous hotel in the Burj Khalifa
11. Currency units in West Africa
12. Grievous
13. "Golden Boy" playwright
19. Actress Shawkat of " Arrested Development"
24. Assignment for an anchor
25. "___ noticed..."
28. Great Hindu sage
29. Late-Triassic flier

Example Crossword

Across
1  Regarding
6  Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 “Take me with you!”
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don’t hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba

Down
1  Home to 41-Down
2  Convince
3  "I wanna know all the details"
4  Start of a modern inquiry
5  Drink similar to sarsaparilla
6  Co-star of 1984’s "Ghostbusters"
7  Right on
8  Aces with aces?
9  Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grieving
13 "Golden Boy" playwright
19 Actress Shawkat of " Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier
Example Crossword

Across
1  Regarding
6  Take back, in a way
10  Start of an aside
14  Omega competitor
15  Something hitting a nerve?
16  Papyrus, e.g.
17  "Take me with you!"
18  Begin flirting with someone, so to speak
20  Assign
21  Hoth, in "Star Wars"
22  ___ rule
23  They don’t hold water
24  Feudal figure
26  Panegyric, e.g.
27  Sci-fi enemy collective, perhaps
31  Ones born beginning in the early 2010s
37  They might cut to the chase
38  Electrically balanced, in chemistry
42  Leaders at the Kaaba

Down
1  Home to 41-Down
2  Convince
3  "I wanna know all the details"
4  Start of a modern inquiry
5  Drink similar to sarsaparilla
6  Co-star of 1984’s "Ghostbusters"
7  Right on
8  Aces with aces?
9  Like some households
10  Designer with an eponymous hotel in the Burj Khalifa
11  Currency units in West Africa
12  Grievous
13  "Golden Boy" playwright
19  Actress Shawkat of "Arrested Development"
24  Assignment for an anchor
25  "___ noticed ..."
28  Great Hindu sage
29  Late-Triassic flier

Example Crossword

Across
1  Regarding
6  Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba

Down
1  Home to 41-Down
2  Convince
3  "I wanna know all the details"
4  Start of a modern inquiry
5  Drink similar to sarsaparilla
6  Co-star of 1984's "Ghostbusters"
7  Right on
8  Aces with aces?
9  Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
19 Actress Shawkat of "Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier

Example Crossword

1  Regarding
6  Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba
43 Home to 41-Down
2 Convince
3 "I wanna know all the details"
4 Start of a modern inquiry
5 Drink similar to sarsaparilla
6 Co-star of 1984's "Ghostbusters"
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
19 Actress Shawkat of "Arrested Development"
24 Assignment for an anchor
25 ___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier

Example Crossword

Across
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba
43 "Things are going great for me!"
46 Muscle worked in rowing exercises, for short
47 Portmanteau coinage for the uneducated and uncultured
48 Touches, e.g.
51 Special pawn move in chess
52 Many a Madrileño
53 Enemy of Wonder Woman
54 Best Picture-nominated 2011 film based on a children's book

Down
1 Home to 41-Down
2 Convince
3 "I wanna know all the details"
4 Start of a modern inquiry
5 Drink similar to sarsaparilla
6 Co-star of 1984's "Ghostbusters"
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
19 Actress Shawkat of " Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier

Example Crossword

Across
14 SEIKO
20 OTT
24 BELL
27 EER
38 BELL
49 45

22 ___ rule
23 They don't hold water
24 Feudal figure
25 Panegyric, e.g.
26 Sci-fi enemy collective, perhaps
27 Ones born beginning in the early 2010s
28 They might cut to the chase
29 Electrically balanced, in chemistry
30 Leaders at the Kaaba
31 "Things are going great for me!"
32 Muscle worked in rowing exercises, for short
33 Portmanteau coinage for the uneducated and uncultured
34 Touches, e.g.
35 Special pawn move in chess
36 Many a Madrileño
37 Enemy of Wonder Woman
38 Best Picture-nominated 2011 film based on a children's book

Down
1 Home to 41-Down
2 Convince
3 "I wanna know all the details"
4 Start of a modern inquiry
5 Drink similar to sarsaparilla
6 Co-star of 1984's "Ghostbusters"
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
14 Actress Shawkat of "Arrested Development"
15 Assignment for an anchor
16 "___ noticed ..."
17 Great Hindu sage
18 Late-Triassic flier

Example Crossword

Across

14 SEIKO
22 ___ rule
23 They don’t hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba
43 "Things are going great for me!"
46 Muscle worked in rowing exercises, for short
47 Portmanteau coinage for the uneducated and uncultured
48 Touches, e.g.
51 Special pawn move in chess
52 Many a Madrileño
53 Enemy of Wonder Woman
54 Best Picture-nominated 2011 film based on a children’s book

Down

5 Drink similar to sarsaparilla
6 Co-star of 1984’s "Ghostbusters"
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
19 Actress Shawkat of "Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier
30 It can be found right on a clock
32 "100 Years ... 100 Stars" and others, in brief

The world’s largest desert capital, after Cairo, Egypt
Example Crossword

Example Crossword


Across

22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba
43 "Things are going great for me!"
46 Muscle worked in rowing exercises, for short
47 Portmanteau coinage for the uneducated and uncultured
48 Touches, e.g.
51 Special pawn move in chess
52 Many a Madrileño
53 Enemy of Wonder Woman
54 Best Picture-nominated 2011 film based on a children's book

Down

5 Drink similar to sarsaparilla
6 Co-star of 1984's "Ghostbusters"
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
19 Actress Shawkat of "Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier
30 It can be found right on a clock
32 "100 Years ... 100 Stars" and others, in brief
33 The world's largest desert capital, after Cairo, Egypt
Example Crossword

Across
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Puny, small
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba
43 "Things are going great for me!"
46 Muscle worked in rowing exercises, for short
47 Portmanteau coinage for the uneducated and uncultured
48 Touches, e.g.
51 Special pawn move in chess
52 Many a Madrileño
53 Enemy of Wonder Woman
54 Best Picture-nominated 2011 film based on a children's book

Down
5 Drink similar to sarsaparilla
6 Co-star of 1984's "Ghostbusters"
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
19 Actress Shawkat of "Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier
30 It can be found right on a clock
32 "100 Years ... 100 Stars" and others, in brief
33 The world's largest desert capital, after Cairo, Egypt

Example Crossword

Example Crossword

Across
22 ___ rule
23 They don’t hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba
43 "Things are going great for me!"
46 Muscle worked in rowing exercises, for short
47 Portmanteau coinage for the uneducated and uncultured
48 Touches, e.g.
51 Special pawn move in chess
52 Many a Madrileño
53 Enemy of Wonder Woman
54 Best Picture-nominated 2011 film based on a children’s book

Down
3 "I wanna know all the details"
4 Start of a modern inquiry
5 Drink similar to sarsaparilla
6 Co-star of 1984’s "Ghostbusters"
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grieved
13 "Golden Boy" playwright
19 Actress Shawkat of "Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier
30 It can be found right on a clock
32 "100 Years ... 100 Stars" and others, in brief
Example Crossword


Across
22 ___ rule
23 They don’t hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba
43 "Things are going great for me!"
46 Muscle worked in rowing exercises, for short
47 Portmanteau coinage for the uneducated and uncultured
48 Touches, e.g.
51 Special pawn move in chess
52 Many a Madrileño
53 Enemy of Wonder Woman
54 Best Picture-nominated 2011 film based on a children’s book

Down
3 "I wanna know all the details"
4 Start of a modern inquiry
5 Drink similar to sarsaparilla
6 Co-star of 1984’s “Ghostbusters”
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 “Golden Boy” playwright
19 Actress Shawkat of “Arrested Development”
24 Assignment for an anchor
25 "___ noticed ...
28 Great Hindu sage
29 Late-Triassic flier
30 It can be found right on a clock
32 “100 Years ... 100 Stars” and others, in brief
Example Crossword

Across

22 ___ rule
23 They don’t hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba
43 "Things are going great for me!"
46 Muscle worked in rowing exercises, for short
47 Portmanteau coinage for the uneducated and uncultured
48 Touches, e.g.
51 Special pawn move in chess
52 Many a Madrileño
53 Enemy of Wonder Woman
54 Best Picture-nominated 2011 film based on a children’s book

Down

3 "I wanna know all the details"
4 Start of a modern inquiry
5 Drink similar to sarsaparilla
6 Co-star of 1984’s "Ghostbusters"
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
19 Actress Shawkat of "Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier
30 It can be found right on a clock
32 "100 Years ... 100 Stars" and others, in brief

Example Crossword

Across

22 ___ rule
23 They don’t hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba
43 "Things are going great for me!"
46 Muscle worked in rowing exercises, for short
47 Portmanteau coinage for the uneducated and uncultured
48 Touches, e.g.
51 Special pawn move in chess
52 Many a Madrileño
53 Enemy of Wonder Woman
54 Best Picture-nominated 2011 film based on a children’s book

Down

3 "I wanna know all the details"
4 Start of a modern inquiry
5 Drink similar to sarsaparilla
6 Co-star of 1984’s “Ghostbusters”
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grievous
13 "Golden Boy" playwright
19 Actress Shawkat of "Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier
30 It can be found right on a clock
32 “100 Years ... 100 Stars” and others, in brief

General Framework

- Question Answering
- Constraint Resolution
- Post-Processing
Previous Work

Proverb [Littman et al., 2002]

<table>
<thead>
<tr>
<th>Question Answering</th>
<th>Constraint Resolution</th>
<th>Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>ExactMatch, Movies, Writers, Music, Bigram, Encyclopedia, WordNet, ...</td>
<td>Belief Propagation + A* Search</td>
<td></td>
</tr>
</tbody>
</table>
Previous Work

Dr. Fill [Ginsberg, 2011]

<table>
<thead>
<tr>
<th>Question Answering</th>
<th>Constraint Resolution</th>
<th>Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-IDF</td>
<td>Limited Discrepancy</td>
<td>Iterative Improvement + Theme Handling</td>
</tr>
</tbody>
</table>
Our Approach

The Berkeley Crossword Solver

Question Answering → Constraint Resolution → Post-Processing
Our Approach

The Berkeley Crossword Solver

<table>
<thead>
<tr>
<th>Question Answering</th>
<th>Constraint Resolution</th>
<th>Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed-Book Neural Question Answering</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Our Approach

The Berkeley Crossword Solver

<table>
<thead>
<tr>
<th>Question Answering</th>
<th>Constraint Resolution</th>
<th>Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed-Book Neural Question Answering</td>
<td>Belief Propagation + Greedy Search</td>
<td></td>
</tr>
</tbody>
</table>
Our Approach

The Berkeley Crossword Solver

<table>
<thead>
<tr>
<th>Question Answering</th>
<th>Constraint Resolution</th>
<th>Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed-Book Neural Question Answering</td>
<td>Belief Propagation + Greedy Search</td>
<td>Open-Book Neural Iterated Improvement</td>
</tr>
</tbody>
</table>
Our Approach

The Berkeley Crossword Solver

Question Answering

Closed-Book Neural Question Answering

Constraint Resolution

Belief Propagation + Greedy Search

Post-Processing

Open-Book Neural Iterated Improvement
14. Omega competitor
14. Omega competitor
Closed-Book QA

14. Omega competitor

Predictions:
- Rolex
- Seiko
- Alpha
- Watch
- Apple
- Enemy
- Skunk
14. Omega competitor

Vocabulary:

- A
- B
- A
- S
- E
- A
- B
- A
- T
- E
- A
- B
- B
- E
- Y
- A
- B
- I
- D
- E
- A
- B
- O
- U
- T
- A
- B
- Y
- S
- S
- Z
- O
- O
- M
- S
14. Omega competitor
14. Omega competitor
14. Omega competitor

Predictions:
- **Rolex**
- **Seiko**
- **Alpha**
- **Apple**
- **Enemy**
- **Skunk**
Handling Ambiguous Clues

Across
1  Regarding
6  Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba

Down
1  Home to 41-Down
2  Convince
3  "I wanna know all the details"
4  Start of a modern inquiry
5  Drink similar to sarsaparilla
6  Co-star of 1984's "Ghostbusters"
7  Right on
8  Aces with aces?
9  Like some households
10  Designer with an eponymous hotel in the Burj Khalifa
11  Currency units in West Africa
12  grievous
13  "Golden Boy" playwright
19  Actress Shawkat of "Arrested Development"
24  Assignment for an anchor
25  "___ noticed..."
28  Great Hindu sage
29  Late-Triassic flier
Handling Ambiguous Clues

Across

1  Regarding
6  Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don’t hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi enemy collective, perhaps
31 Ones born beginning in the early 2010s
37 They might cut to the chase
38 Electrically balanced, in chemistry
42 Leaders at the Kaaba

Down

1 Home to 41-Down
2 Convince
3 "I wanna know all the details"
4 Start of a modern inquiry
5 Drink similar to sarsaparilla
6 Co-star of 1984’s "Ghostbusters"
7 Right on
8 Aces with aces?
9 Like some households
10 Designer with an eponymous hotel in the Burj Khalifa
11 Currency units in West Africa
12 Grieved
13 "Golden Boy" playwright
19 Actress Shawkat of "Arrested Development"
24 Assignment for an anchor
25 "___ noticed ..."
28 Great Hindu sage
29 Late-Triassic flier
Handling Ambiguous Clues

Our Approach

The Berkeley Crossword Solver

<table>
<thead>
<tr>
<th>Question Answering</th>
<th>Constraint Resolution</th>
<th>Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed-Book Neural Question Answering</td>
<td>Belief Propagation + Greedy Search</td>
<td>Open-Book Neural Iterated Improvement</td>
</tr>
</tbody>
</table>
Our Approach

The Berkeley Crossword Solver

<table>
<thead>
<tr>
<th>Question Answering</th>
<th>Constraint Resolution</th>
<th>Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed-Book Neural Question Answering</td>
<td>Belief Propagation + Greedy Search</td>
<td>Open-Book Neural Iterated Improvement</td>
</tr>
</tbody>
</table>
Loopy Belief Propagation
Loopy Belief Propagation
Loopy Belief Propagation
Loopy Belief Propagation

Diagram showing a network of nodes labeled with letters, including "ROLEX", "SEIKO", "APPLE", "X", "E", "O", and others, illustrating the propagation of beliefs through the network.
Loopy Belief Propagation
Our Approach

The Berkeley Crossword Solver

- Question Answering
  - Closed-Book Neural Question Answering
- Constraint Resolution
  - Belief Propagation + Greedy Search
- Post-Processing
  - Open-Book Neural Iterated Improvement
Our Approach

The Berkeley Crossword Solver

<table>
<thead>
<tr>
<th>Question Answering</th>
<th>Constraint Resolution</th>
<th>Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed-Book Neural Question Answering</td>
<td>Belief Propagation + Greedy Search</td>
<td>Open-Book Neural Iterated Improvement</td>
</tr>
</tbody>
</table>
Open-Book Improvement

Open-Book Improvement


Across
1. Beloved, in Arabic
6. [Make it quick!]
10. Masthead listings, for short
13. Italian herbal liqueur
14. "Please, please, please?"
15. People calling the shots at the zoo?
17. Not radical
19. Drink with a dome-shaped lid
20. Where to find the Egyptian Temple of Dendur, with "the"
21. Crush, as a test
22. Cry from a survivor
24. Clara in the National Women's Hall of Fame
26. Stuffed oneself with, facetiously
27. Newswoman Roberts
28. Recipe amt.
31. Doesn't stay natural?
32. Fell off, as laughter
33. Place

Down
1. Soccer star on a 1999 Wheaties box
2. Bloblike
3. Ill-advised opinions
4. Fury
5. Feeling on a lo-o-o-ng car trip
6. Lab workers
7. Performer with the hit 2006 album "Hip Hop Is Dead"
8. Voting no
9. ___ Beach, Calif.
10. Satan, with "the"
11. Crack, as a secret message
12. Shorthand writers, for short
16. Dated
18. Cpl. or sgt.
23. Expand
25. Purge (of)
27. Atlanta-based health org.
29. Long fur scarfs
Open-Book Improvement

Across
1. Beloved, in Arabic
6. [Make it quick!] 
10. Masthead listings, for short
13. Italian herbal liqueur
14. “Please, please, please?”
15. People calling the shots at the zoo?
17. Not radical
19. Drink with a dome-shaped lid
20. Where to find the Egyptian Temple of Dendur, with “the”
21. Crush, as a test
22. Cry from a survivor
24. Clara in the National Women’s Hall of Fame
26. Stuffed oneself with, facetiously
27. Newswoman Roberts
28. Recipe amt.
31. Doesn’t stay natural?
32. Fell off, as laughter
33. Place

Down
1. Soccer star on a 1999 Wheaties box
2. Bloblike
3. Ill-advised opinions
4. Fury
5. Feeling on a lo-o-o-ong car trip
6. Lab workers
7. Performer with the hit 2006 album "Hip Hop Is Dead"
8. Voting no
9. ___ Beach, Calif.
10. Satan, with "the"
11. Crack, as a secret message
12. Shorthand writers, for short
16. Dated
18. Cpl. or sgt.
23. Expand
25. Purge (of)
27. Atlanta-based health org.
29. Long fur scarfs
Open-Book Improvement

Across
1  Beloved, in Arabic
6  [Make it quick!]
10  Masthead listings, for short
13  Italian herbal liqueur
14  "Please, please, please?"
15  People calling the shots at the zoo?
17  Not radical
19  Drink with a dome-shaped lid
20  Where to find the Egyptian Temple of Dendur, with "the"
21  Crush, as a test
22  Cry from a survivor
24  Clara in the National Women's Hall of Fame
26  Stuffed oneself with, facetiously
27  Newswoman Roberts
28  Recipe amt.
31  Doesn't stay natural?
32  Fell off, as laughter
33  Place

Down
1  Soccer star on a 1999 Wheaties box
2  Bloblike
3  Ill-advised opinions
4  Fury
5  Feeling on a lo-o-ng car trip
6  Lab workers
7  Performer with the hit 2006 album "Hip Hop Is Dead"
8  Voting no
9  ___ Beach, Calif.
10  Satan, with "the"
11  Crack, as a secret message
12  Shorthand writers, for short
16  Dated
18  Cpl. or sgt.
23  Expand
25  Purge (of)
27  Atlanta-based health org.
29  Long fur scarfs
Results: The New York Times

- Letter Accuracy: Dr. Fill 99.4%, The Berkeley Crossword Solver 98.1%
- Word Accuracy: Dr. Fill 99.4%, The Berkeley Crossword Solver 98.1%
- Exact Puzzle Match: Dr. Fill 56.9%
Results: The New York Times

- Letter Accuracy: Dr. Fill 99.4, The Berkeley Crossword Solver 99.7
- Word Accuracy: Dr. Fill 98.1, The Berkeley Crossword Solver 98.9
- Exact Puzzle Match: 56.9
Results: The New York Times

<table>
<thead>
<tr>
<th>Percent</th>
<th>Dr. Fill</th>
<th>The Berkeley Crossword Solver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter Accuracy</td>
<td>99.4</td>
<td>99.7</td>
</tr>
<tr>
<td>Word Accuracy</td>
<td>98.1</td>
<td>98.9</td>
</tr>
<tr>
<td>Exact Puzzle Match</td>
<td>56.9</td>
<td>81.7</td>
</tr>
</tbody>
</table>
American Crossword Puzzle Tournament

Top Human Competitor

Median Human

<table>
<thead>
<tr>
<th></th>
<th>Top Human Competitor</th>
<th>Median Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proverb (1998)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr. Fill (2012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF + BCS (2021)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**American Crossword Puzzle Tournament**

- Median Human: 6,215
American Crossword Puzzle Tournament

Top Human Competitor

Median Human


6,215  10,060
American Crossword Puzzle Tournament

Top Human Competitor

Median Human

Proverb (1998) 11,795
Dr. Fill (2017) 6,215
DF + BCS (2021)
American Crossword Puzzle Tournament

Top Human Competitor

Median Human

- Proverb (1998): 6,215
- Dr. Fill (2017): 11,795
- DF + BCS (2021): 12,825
American Crossword Puzzle Tournament

- Top Human Competitor: 13,060
- Median Human: 11,795
- Proverb (1998): 6,215
- Dr. Fill (2017): 11,795
- BCS (2021): 13,060
Overview

Crossword Solving

Constraint satisfaction, belief propagation, greedy search, iterative improvement

Detecting LLMs

Classification, feature selection, generalization and overfitting

Decision-Making in NL

Constraint satisfaction, value of information, decision networks, particle filtering
Overview

Crossword Solving

Constraint satisfaction, belief propagation, greedy search, iterative improvement

Detecting LLMs

Classification, feature selection, generalization and overfitting

Decision-Making in NL

Constraint satisfaction, value of information, decision networks, particle filtering
ChatGPT in Schools

Educators Battle Plagiarism As 89% Of Students Admit To Using OpenAI’s ChatGPT For Homework

Chris Westfall Contributor
Business coach, Author of Easier and Leadership Language.

Alarmed by A.I. Chatbots, Universities Revamping How They Teach

With the rise of the popular new chatbot ChatGPT, colleges are restructuring some courses and taking preventive measures to combat plagiarism.

Turnitin is the go-to software to catch students cheating. Now it’s focused on a potential cat-and-mouse game with OpenAI’s new ChatGPT chatbot.

Samantha Delouya Jan 11, 2023, 3:40 AM PST

ChatGPT Banned in New York City Public Schools Over Concerns About Cheating, Learning Development

Education officials say students could use the artificial-intelligence chatbot to do their schoolwork.

By Abyssa Lokpat Follow
Updated Jan. 6, 2023 at 12:55 pm ET
**Prompt:** Every authority is baffled but for the past few years everyone has been receiving a Christmas present. To some delight and others horror, the gifts are based on how 'good' you've been.

<table>
<thead>
<tr>
<th>All of this should be such a joy, a wondrous time where people all around the world are brimming with love and excitement over what they might have been brought. But it's not, I'm worried, I'm borderline panicked. Every single year, I've felt my anxiety grow as the temperature drops, all because of one question nagging me between the ears. Am I on the naughty list this year? And there's no answer, there's never an answer because we weren't given a guideline, there's no clear line drawn in the sand that we must not cross lest we end up on the naughty list. So we're left to our own devices, trying to live our lives as best we can, but people slip...</th>
</tr>
</thead>
<tbody>
<tr>
<td>It was a sunny December morning when the annual tradition of receiving Christmas presents took a peculiar turn. The streets were filled with laughter, carolers singing joyously, and children eagerly awaiting the arrival of Santa Claus. But this year, things were different. Word had spread like wildfire that everyone, without exception, would receive a Christmas present. Yet, there was an unsettling twist - these gifts would be based on how “good” one had been throughout the year. Authorities were baffled, for no one knew who was behind this curious occurrence. As the clock struck midnight on Christmas Eve, parcels began appearing on doorsteps...</td>
</tr>
</tbody>
</table>
**Prompt:** Every authority is baffled but for the past few years everyone has been receiving a Christmas present. To some delight and others horror, the gifts are based on how 'good' you've been.

<table>
<thead>
<tr>
<th>Human Written Text</th>
<th>ChatGPT Generated Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>All of this should be such a joy, a wondrous time where people all around the world are brimming with love and excitement over what they might have been brought. But it's not, I'm worried, I'm borderline panicked. Every single year, I've felt my anxiety grow as the temperature drops, all because of one question nagging me between the ears. Am I on the naughty list this year? And there's no answer, there's never an answer because we weren't given a guideline, there's no clear line drawn in the sand that we must not cross lest we end up on the naughty list. So we're left to our own devices, trying to live our lives as best we can, but people slip...</td>
<td>It was a sunny December morning when the annual tradition of receiving Christmas presents took a peculiar turn. The streets were filled with laughter, carolers singing joyously, and children eagerly awaiting the arrival of Santa Claus. But this year, things were different. Word had spread like wildfire that everyone, without exception, would receive a Christmas present. Yet, there was an unsettling twist – these gifts would be based on how “good” one had been throughout the year. Authorities were baffled, for no one knew who was behind this curious occurrence. As the clock struck midnight on Christmas Eve, parcels began appearing on doorsteps...</td>
</tr>
</tbody>
</table>
Data Sources

Reddit /r/writingprompts

Reuters News

IvyPanda Essays

How can we avoid spurious topic correlations between positive and negative examples?
Two Ideas That Didn’t Work
Two Ideas That Didn’t Work

Idea #1: Perplexity Thresholding

$$PPL(X) = \exp \left\{ -\frac{1}{t} \sum_{i}^{t} \log p_{\theta}(x_{i}|x_{<i}) \right\}$$

Core assumption is that language models will assign higher probability to LLM-generated text

**Bad underfitting (80 F1)**
Two Ideas That Didn’t Work

Idea #1: Perplexity Thresholding

$$PPL(X) = \exp \left\{ -\frac{1}{t} \sum_{i}^{t} \log p_{\theta}(x_i|x_{<i}) \right\}$$

Core assumption is that language models will assign higher probability to LLM-generated text

Bad underfitting (80 F1)

Idea #2: Fine-tuned Classifier

Bad overfitting

(98 F1 in-domain, but 85 F1 OOD)
Our Approach: Ghostbuster

It was a typical Friday night, and I had decided to spend it at home, browsing the internet and catching up on some of my favorite shows. My roommate had gone...

<table>
<thead>
<tr>
<th>Stories</th>
<th>News</th>
<th>Essay</th>
</tr>
</thead>
<tbody>
<tr>
<td>max(ada / unigram + ada)</td>
<td>avg-25(ada &gt; davinci / trigram)</td>
<td>avg(ada &gt; davinci / trigram)</td>
</tr>
<tr>
<td>var(davinci / ada - davinci)</td>
<td>var(ada - davinci)</td>
<td>avg(davinci)</td>
</tr>
<tr>
<td>avg(unigram &gt; trigram &gt; davinci)</td>
<td>avg(trigram / ada &gt; davinci)</td>
<td>var(trigram &gt; davinci)</td>
</tr>
<tr>
<td>avg(unigram + trigram &lt; davinci)</td>
<td>var(unigram &gt; ada - davinci)</td>
<td>avg(unigram &gt; trigram + ada)</td>
</tr>
<tr>
<td>var(unigram * davinci)</td>
<td>avg(unigram &gt; trigram / unigram)</td>
<td>avg(unigram / ada &gt; davinci)</td>
</tr>
</tbody>
</table>
Results: Domain Generalization

In-Domain Model Performance

Out-of-Domain Model Performance
Results: Robustness to Perturbations
Results: Varying Document Lengths

![Graph showing F1 Score vs. Document Size for different domains](image)

**Axes:**
- **Y-axis:** F1 Score
- **X-axis:** Document Size (in # of Tokens)

**Legend:**
- In-Domain
- Out-Domain (Claude)
- Out-Domain (WP)
- Out-Domain (Reuter)
- Out-Domain (Essay)
AI-Detectors Biased Against Non-Native English Writers

Don’t put faith in detectors that are “unreliable and easily gamed,” says scholar.

May 15, 2023 | Andrew Myers
Ghostbuster Demo

a model-agnostic detector for AI-generated text

How does it work?
Ghostbuster works by parsing documents through a series of rewrite language models and running a structured search over possible combinations of their features, then training a classifier on the selected features to determine if the target document was AI-generated.

Disclaimer
Overview

Crossword Solving

Constraint satisfaction, belief propagation, greedy search, iterative improvement

Detecting LLMs

Classification, feature selection, generalization and overfitting

Decision-Making in NL

Constraint satisfaction, value of information, decision networks, particle filtering
Overview

Crossword Solving

Constraint satisfaction, belief propagation, greedy search, iterative improvement

Detecting LLMs

Classification, feature selection, generalization and overfitting

Decision-Making in NL

Constraint satisfaction, value of information, decision networks, particle filtering
My projector isn’t working. What’s a list of things that could be going wrong?
My projector isn’t working. What’s a list of things that could be going wrong?

My projector stopped working. Can we narrow down what might be going wrong together, step-by-step?
What’s the current price of NVIDIA stock?
What’s the current price of NVIDIA stock?

I’d like to plan our company’s Q4 strategy. Can you take a look at the financial reports and market analysis with me and figure out what we need to prioritize before the end of the year?
Can you book me a flight to LA tomorrow?
Can you book me a flight to LA tomorrow?

Hi! Could you help me plan a trip to LA? I’m giving a talk at USC on Friday and would like to book a nearby hotel. I could fly out of OAK or SFO and am flexible on dates, but I’d like to make sure I have time to explore some nearby restaurants. Could you propose a few options for me and let me choose which one sounds best?
The DialOp Benchmark

**Optimization**
Conference Reviewer Matching

Assign suitable reviewers for conference papers

- matching preferences
- matching preferences

Tim and Spencer could be good fits for the RLHF paper.

I don’t think Tim is working on RL anymore, would Cordelia be a viable match instead?

She’s the only person who can review MT.

**Planning**
Travel Planning

Plan an efficient itinerary the user would enjoy

- preferences
- site locations and features

Hi! I want to check out the Guggenheim but other than that, hang out at cafes...

How about this itinerary?

I don’t like Italian food, unless that place is notable?

**Mediation**
Group Scheduling

Book a set of flights that respect individual and group constraints

- individual constraints and flights
- global visibility

Could we all arrive before Friday 5pm?

I prefer Saturday but I could make Friday work if Claire wants that.
The DialOp Benchmark

**Optimization**
Conference Reviewer Matching
Assign suitable reviewers for conference papers
- matching preferences

Tim and Spencer could be good fits for the RLHF paper.
I don’t think Tim is working on RL anymore, would Cordelia be a viable match instead?
She’s the only person who can review MT.

**Planning**
Travel Planning
Plan an efficient itinerary the user would enjoy
- preferences
- site locations and features

Travel Ideas

Hi! I want to check out the Guggenheim but other than that, hang out at cafes...
How about this itinerary?
I don’t like Italian food, unless that place is notable?

**Mediation**
Group Scheduling
Book a set of flights that respect individual and group constraints
- individual constraints and flights
- global visibility

Could we all arrive before Friday 5pm?
I prefer Saturday but I could make Friday work if Claire wants that.
The DialOp Benchmark

**Optimization**
Conference Reviewer Matching

Assign suitable reviewers for conference papers

- matching preferences
- matching preferences

Tim and Spencer could be good fits for the RLHF paper.

I don’t think Tim is working on RL anymore, would Cordelia be a viable match instead?

She’s the only person who can review MT.

**Planning**
Travel Planning

Plan an efficient itinerary the user would enjoy

- preferences
- site locations and features

Hi! I want to check out the Guggenheim but other than that, hang out at cafes...

How about this itinerary?

I don’t like Italian food, unless that place is notable.

**Mediation**
Group Scheduling

Book a set of flights that respect individual and group constraints

Could we all arrive before Friday 5pm?

I prefer Saturday but I could make Friday work if Claire wants that.
The DialOp Benchmark

**Optimization**
Conference Reviewer Matching
Assign suitable reviewers for conference papers
- matching preferences

**Planning**
Travel Planning
Plan an efficient itinerary the user would enjoy
- preferences
- site locations and features

**Mediation**
Group Scheduling
Book a set of flights that respect individual and group constraints
- individual constraints and flights
- global visibility

---

Tim and Spencer could be good fits for the RLHF paper.

I don’t think Tim is working on RL anymore, would Cordelia be a viable match instead?

She’s the only person who can review MT.

Hi! I want to check out the Guggenheim but other than that, hang out at cafes...

How about this itinerary?

I don’t like Italian food, unless that place is notable?

Could we all arrive before Friday 5pm?

I prefer Saturday but I could make Friday work if Claire wants that.
The DialOp Benchmark

**Optimization**
Conference Reviewer Matching
Assign suitable reviewers for conference papers

**Planning**
Travel Planning
Plan an efficient itinerary the user would enjoy

**Mediation**
Group Scheduling
Book a set of flights that respect individual and group constraints

- Tim and Spencer could be good fits for the RLHF paper.
- I don’t think Tim is working on RL anymore, would Cordelia be a viable match instead?
- She’s the only person who can review MT.

- Hi! I want to check out the Guggenheim but other than that, hang out at cafes...
- How about this itinerary?
- I don’t like Italian food, unless that place is notable?
- Could we all arrive before Friday 5pm?
- I prefer Saturday but I could make Friday work if Claire wants that.
User Interface: Planning

Travel Planning
- Generally like easy eating, reasonable options
- Go to at least one bar
- Don’t like Japanese, like Ethiopian
- Don’t like beans, rice, like entrees
- Family trip, bent to go to places that are great for kids
- Group of 7, been 12 we would go to places good for groups
- Keep budget below $90
- Wants option much better than tourist traps
- Would be great if everything is reservable in advance
- Prefer just eating takeout
- Minimize travel distance

Sawcreek
Price: $13
Checking which popular items are satisfied by this itinerary:
- 2 go to at least one bar
- keep budget below $90

1. Lincoln Park
- Estimated price: $40
- Rating: 3
- Accepts reservations
- Open late
- Good for groups
- Vegetarian options
- Estimated Travel: 6.4 miles

2. Luckie’s
- Estimated price: $220
- Rating: 3
- Accepts reservations
- Open late
- Good for groups
- Vegetarian options
- Estimated Travel: 6.2 miles

3. Liberty Memorial
- Estimated price: $80
- Rating: 2
- Open late
- Good for groups
- Vegetarian options
- Estimated Travel: 6.2 miles

Chat
- Request a proposal
- Send Proposal
  - You already sent a proposal, use a message to edit your turn

The Tower
- Type: brunch
- Estimated price: $80
- Has parking
good for groups
- Open late
- Tourist trap
- Accepts reservations
- Estimated Travel: 6.4 miles

Lincoln Park
- Estimated price: $40
- Has parking
good for groups
- Open late
good for kids
tourist trap
- Estimated Travel: 6.4 miles

Luckie’s
- Estimated price: $220
- Has parking
good for groups
- Open late
- Good for kids
tourist trap
- Estimated Travel: 6.2 miles

Liberty Memorial
- Estimated price: $80
- Has parking
- Good for groups
- Open late
- Accepts reservations
- Estimated Travel: 6.2 miles

Chat
- Request a proposal
- Send Proposal

Starting new game...
- I’m looking to visit some places in the city that are good for groups. Would also like to keep a tight budget.
- You made a proposal.
Human Data Collection

134 games
Avg. 18 messages per game
Avg. 8min 9s per game

114 games
Avg. 9 messages per game
Avg. 10min 56s per game

162 games
Avg. 13 messages per game
Avg. 7min 15s per game

Optimization

Planning

Mediation
The DialOp Environment

Flights:
id | carrier | price | times
0 | Southwest | 50 | 5/31 9:17 AM - 5:17 PM
1 | JetBlue | 1456 | 5/31 2:14 PM - 10:14 PM
...

Chat:
User 1: Hi, Monday seems best, any way that would also work for them?
You (Agent) to User 1:
The DialOp Environment

Flights:
- id | carrier | price | times
- 0 | Southwest | 50 | 5/31 9:17 AM - 5:17 PM
- 1 | JetBlue | 1456 | 5/31 2:14 PM - 10:14 PM

Chat:
User 1: Hi, Monday seems best, any way that would also work for them?
You (Agent) to User 1:

...  

User 1: That’s great, thanks!
User 2: Let’s do it!
The DialOp Environment

Flights:
<table>
<thead>
<tr>
<th>id</th>
<th>carrier</th>
<th>price</th>
<th>times</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Southwest</td>
<td>50</td>
<td>5/31 9:17 AM - 5:17 PM</td>
</tr>
<tr>
<td>1</td>
<td>JetBlue</td>
<td>1456</td>
<td>5/31 2:14 PM - 10:14 PM</td>
</tr>
</tbody>
</table>
...

Chat:
User 1: Hi, Monday seems best, any way that would also work for them?
You (Agent) to User 1:
...
You (Agent): How does this set of flights look for you all? ...
User 1: That’s great, thanks!
User 2: Let’s do it!

Reward: 0.77
Evaluating Dialogue Assistants

Gold Standard: Human-LM Evaluation

Assist real human users
Evaluating Dialogue Assistants

Gold Standard: Human-LM Evaluation
- Assist real human users

Proxy: LM-LM Self-Play
- Assist an LM user simulator
Evaluating Dialogue Assistants

Gold Standard: Human-LM Evaluation
Assist real human users

Proxy: LM-LM Self-Play
Assist an LM user simulator

Better Proxy: Prompted Self-Play
Continue assistance from a human-human dialogue prefix
Results: Language Model Self-Play

- Optimization
- Planning
- Mediation

Graphs show the relationship between reward and number of words for Self-Play and Human-Human conditions, with a comparison to random performance.
Results: Prompted Self-Play

- **Optimization**
  - Reward comparison across conditions including Self-Play, 50%, 75%, Proposal, and Human with a random baseline.

- **Planning**
  - Similar comparison as Optimization but with adjusted reward values.

- **Mediation**
  - Highest reward values among all conditions, with a focus on Human performance.
I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)
Example Selfplay Dialogue

I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)
I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)
I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)
Example Selfplay Dialogue

I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)
Example Selfplay Dialogue

I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)
Example Selfplay Dialogue

I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)

Model messages all preferences into the chat

Model queries for already revealed information

(Real Dialogue)
I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)
Example Selfplay Dialogue

I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)
Example Selfplay Dialogue

I'm looking for places good for groups of 7, with vegetarian options, a shop, takeout, Harper and Rye, rated above 3 stars, and no live music. My budget is $70.

I have several options available. Do you have a price point?

As long as it's $70 or under, it should be fine.

Proposal: Cookies Cream, The Cupcake Bar, Asha Tea House

I have found several options. All of these places are within a mile of each other and have vegetarian options. Do you have any other preferences?

Proposal Rejected (64% Match)
Hey I'd like to plan a trip for my family, so I'd like to visit all the touristy spots in town, visit at least one landmark, and find a place with wifi...because I like classy places.

Well Einstein's Summer House is a famous landmark in the area and it's in walking distance to the Coffee Nook which is a cafe that has wifi in the area.

Proposal: Einstein's summer house, Coffee Nook, NULL

That looks pretty decent, could we tack on a visit to the Garden of Wonders? I REALLY WANT TO GO THERE.

Oh that is a great place to visit. Here is a new schedule that keeps you in the area around the Garden of Wonders.
Hey I'd like to plan a trip for my family, so I'd like to visit all the touristy spots in town, visit at least one landmark, and find a place with wifi...because I like classy places.

Well Einstein's Summer House is a famous landmark in the area and it's in walking distance to the Coffee Nook which is a cafe that has wifi in the area.

Proposal: Einstein's summer house, Coffee Nook, NULL

That looks pretty decent, could we tack on a visit to the Garden of Wonders? I REALLY WANT TO GO THERE.

Oh that is a great place to visit. Here is a new schedule that keeps you in the area around the Garden of Wonders.
Hey I'd like to plan a trip for my family, so I'd like to visit all the touristy spots in town, visit at least one landmark, and find a place with wifi...because I like classy places

Well Einstein's Summer House is a famous landmark in the area and it's in walking distance to the Coffee Nook which is a cafe that has wifi in the area

Proposal: Einstein's summer house, Coffee Nook, NULL

That looks pretty decent, could we tack on a visit to the Garden of Wonders? I REALLY WANT TO GO THERE

Oh that is a great place to visit. Here is a new schedule that keeps you in the area around the Garden of Wonders
Hey I'd like to plan a trip for my family, so I'd like to visit all the touristy spots in town, visit at least one landmark, and find a place with wifi...because I like classy places.

Well Einstein's Summer House is a famous landmark in the area and it's in walking distance to the Coffee Nook which is a cafe that has wifi in the area.

Proposal: Einstein's summer house, Coffee Nook, NULL

That looks pretty decent, could we tack on a visit to the Garden of Wonders? I REALLY WANT TO GO THERE!

Oh that is a great place to visit. Here is a new schedule that keeps you in the area around the Garden of Wonders.
Hey I'd like to plan a trip for my family, so I'd like to visit all the touristy spots in town, visit at least one landmark, and find a place with wifi...because I like classy places

Well Einstein's Summer House is a famous landmark in the area and it's in walking distance to the Coffee Nook which is a cafe that has wifi in the area

Proposal: Einstein's summer house, Coffee Nook, NULL

That looks pretty decent, could we tack on a visit to the Garden of Wonders? I REALLY WANT TO GO THERE

Oh that is a great place to visit. Here is a new schedule that keeps you in the area around the Garden of Wonders
Example Human-Human Dialogue

Hey I'd like to plan a trip for my family, so I'd like to visit all the touristy spots in town, visit at least one landmark, and find a place with wifi...because I like classy places

Well Einstein's Summer House is a famous landmark in the area and it's in walking distance to the Coffee Nook which is a cafe that has wifi in the area

Proposal: Einstein's summer house, Coffee Nook, NULL

That looks pretty decent, could we tack on a visit to the Garden of Wonders? I REALLY WANT TO GO THERE

Oh that is a great place to visit. Here is a new schedule that keeps you in the area around the Garden of Wonders
Hey I'd like to plan a trip for my family, so I'd like to visit all the touristy spots in town, visit at least one landmark, and find a place with wifi...because I like classy places.

Well Einstein's Summer House is a famous landmark in the area and it's in walking distance to the Coffee Nook which is a cafe that has wifi in the area.

Proposal: Einstein's summer house, Coffee Nook, NULL

That looks pretty decent, could we tack on a visit to the Garden of Wonders? I REALLY WANT TO GO THERE.

Oh that is a great place to visit. Here is a new schedule that keeps you in the area around the Garden of Wonders.
Example Human-Human Dialogue

Hey I'd like to plan a trip for my family, so I'd like to visit all the touristy spots in town, visit at least one landmark, and find a place with wifi...because I like classy places

Well Einstein's Summer House is a famous landmark in the area and it's in walking distance to the Coffee Nook which is a cafe that has wifi in the area

Proposal: Einstein's summer house, Coffee Nook, NULL

That looks pretty decent, could we tack on a visit to the Garden of Wonders? I REALLY WANT TO GO THERE

Oh that is a great place to visit. Here is a new schedule that keeps you in the area around the Garden of Wonders.
Demo (Time Permitting)
Thanks for listening!

You can find more info about these projects on my website:
https://people.eecs.berkeley.edu/~nicholas_tomlin/

My awesome collaborators

Eric Wallace
UC Berkeley

Kevin Yang
UC Berkeley

Albert Xu
USC

Eshaan Pathak
You.com

Matt Ginsberg
Google

Dan Klein
UC Berkeley

Jessy Lin
UC Berkeley

Jacob Andreas
MIT

Jason Eisner
Johns Hopkins

Vivek Verma
UC Berkeley

Eve Fleisig
UC Berkeley

Coco
My Cat