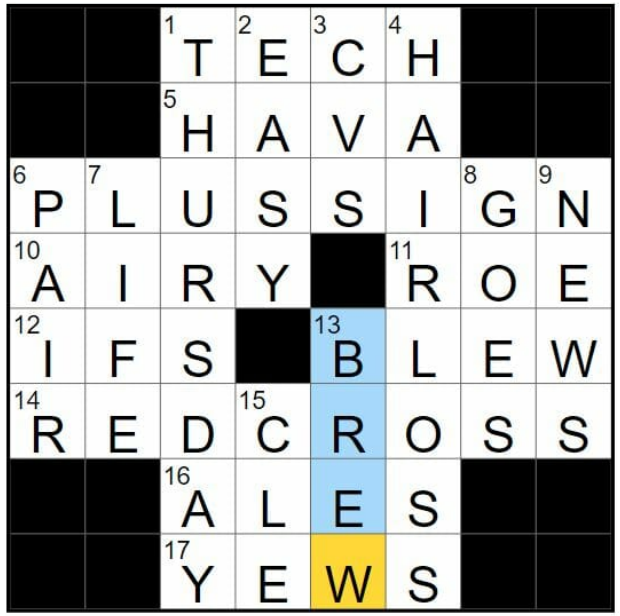
Computational Rationality in LLMs

CS188, Fall 2023

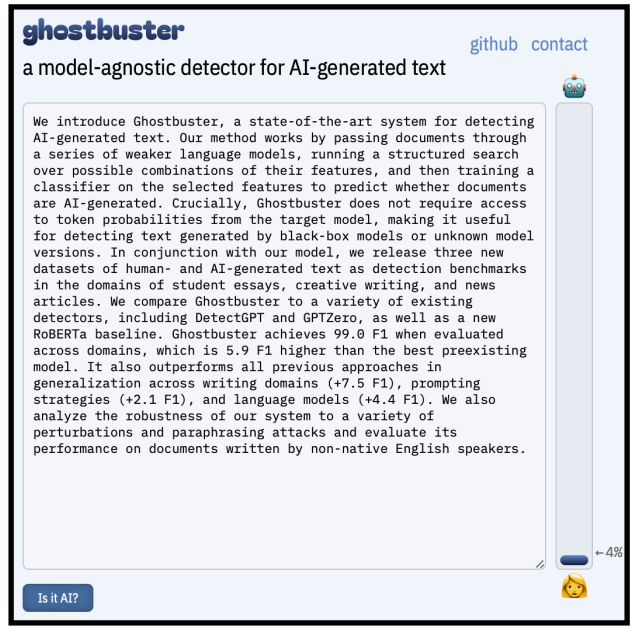
Nicholas Tomlin

Overview

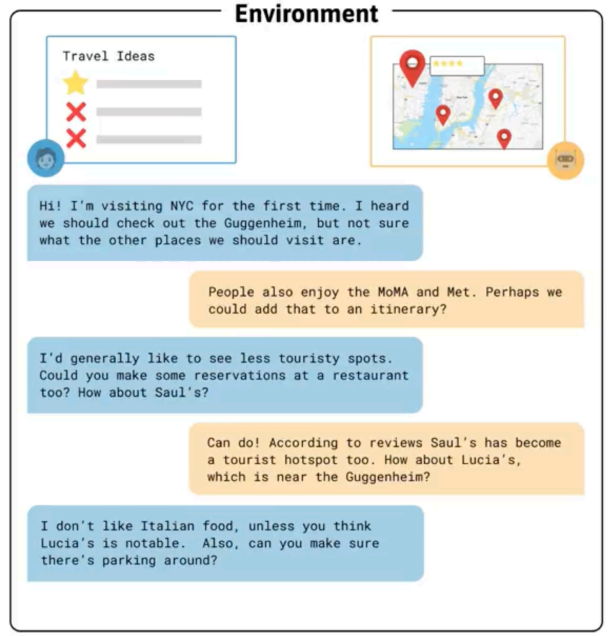
Crossword Solving



Detecting LLMs

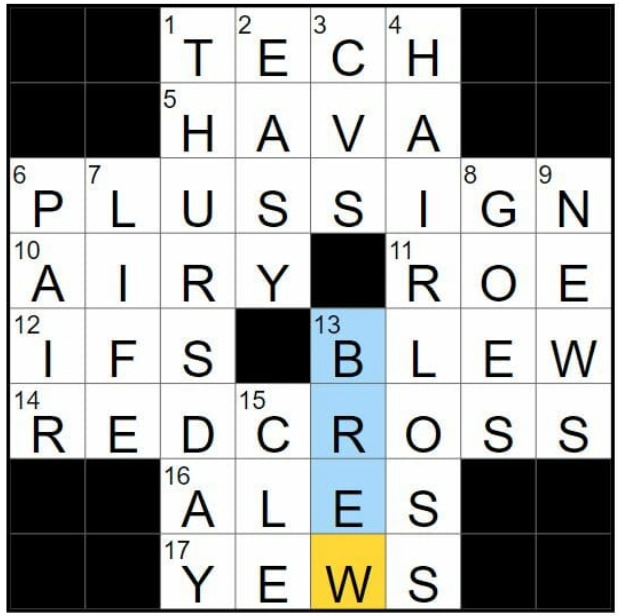


Decision-Making in NL



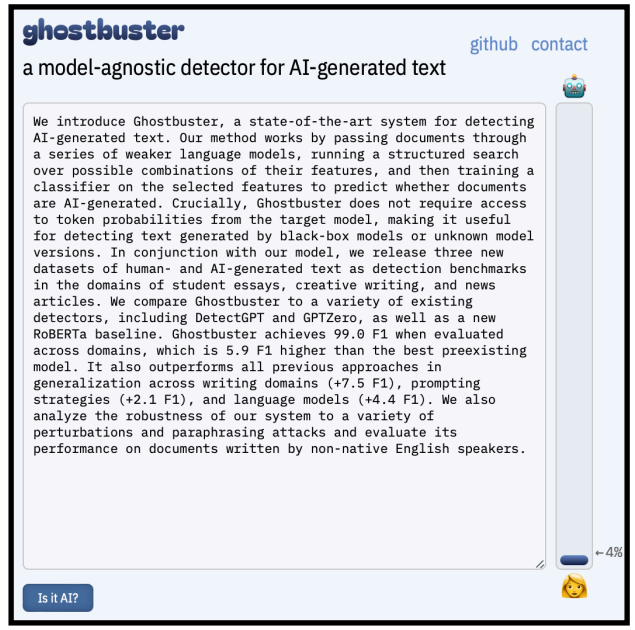
Overview

Crossword Solving



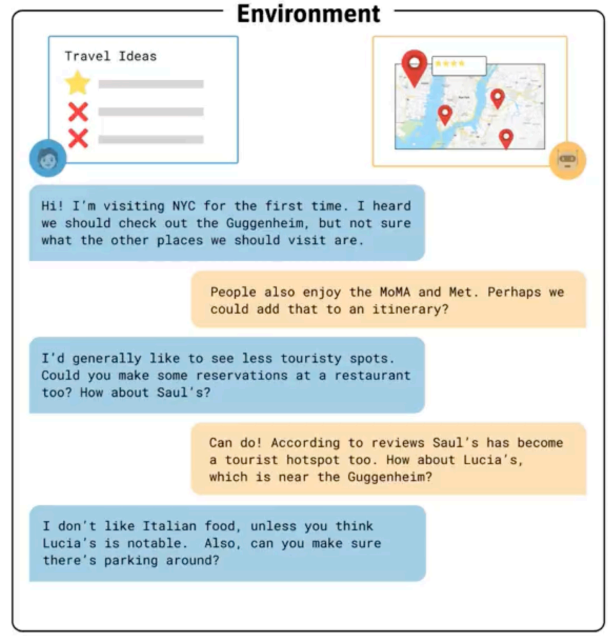
Constraint satisfaction, belief propagation, greedy search, iterative improvement

Detecting LLMs



Classification, feature selection, generalization and overfitting

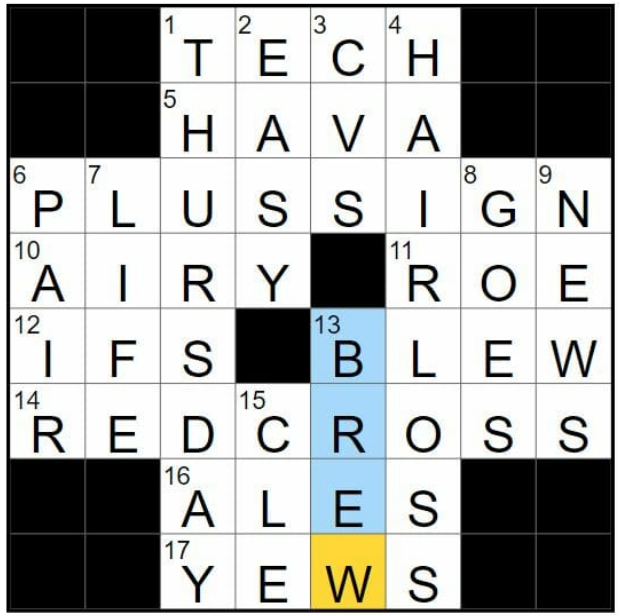
Decision-Making in NL



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

Overview

Crossword Solving



Constraint satisfaction, belief propagation, greedy search, iterative improvement

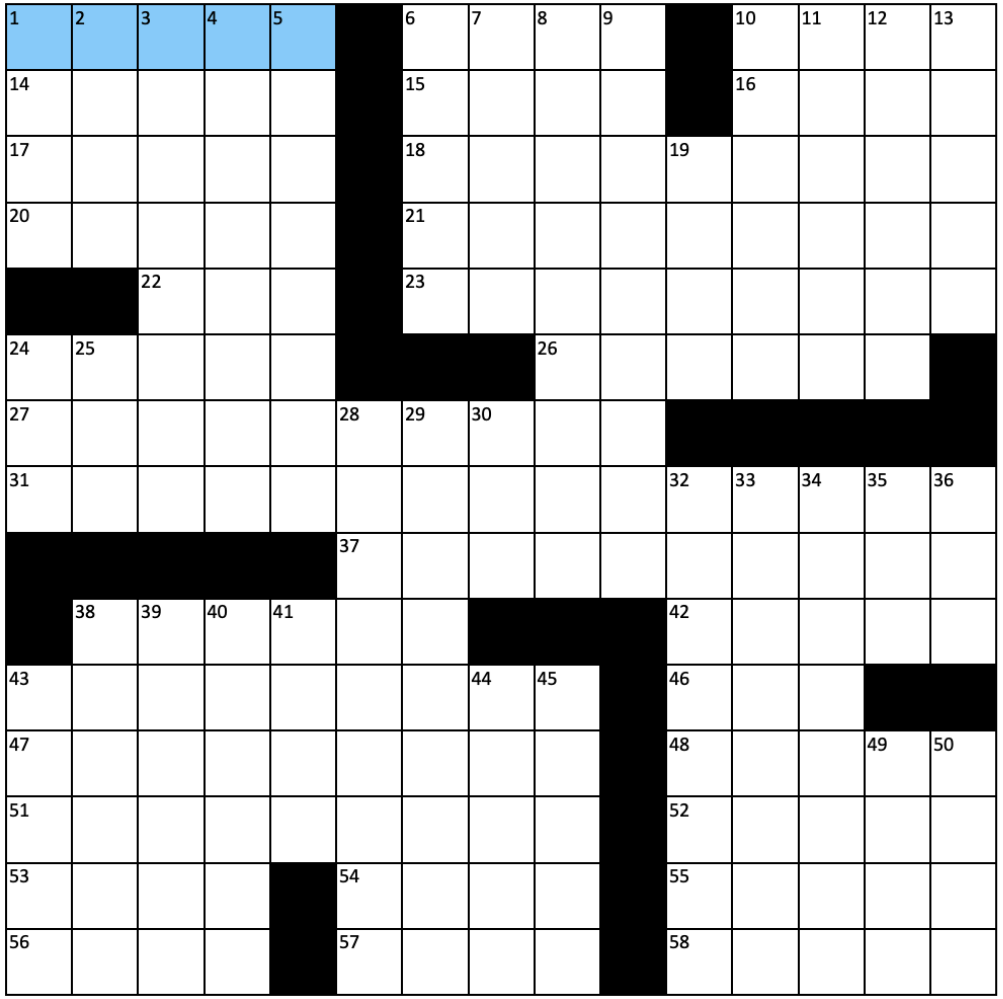
Detecting LLMs

Classification, feature selection, generalization and overfitting

Decision-Making in NL

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

Example Crossword



Across

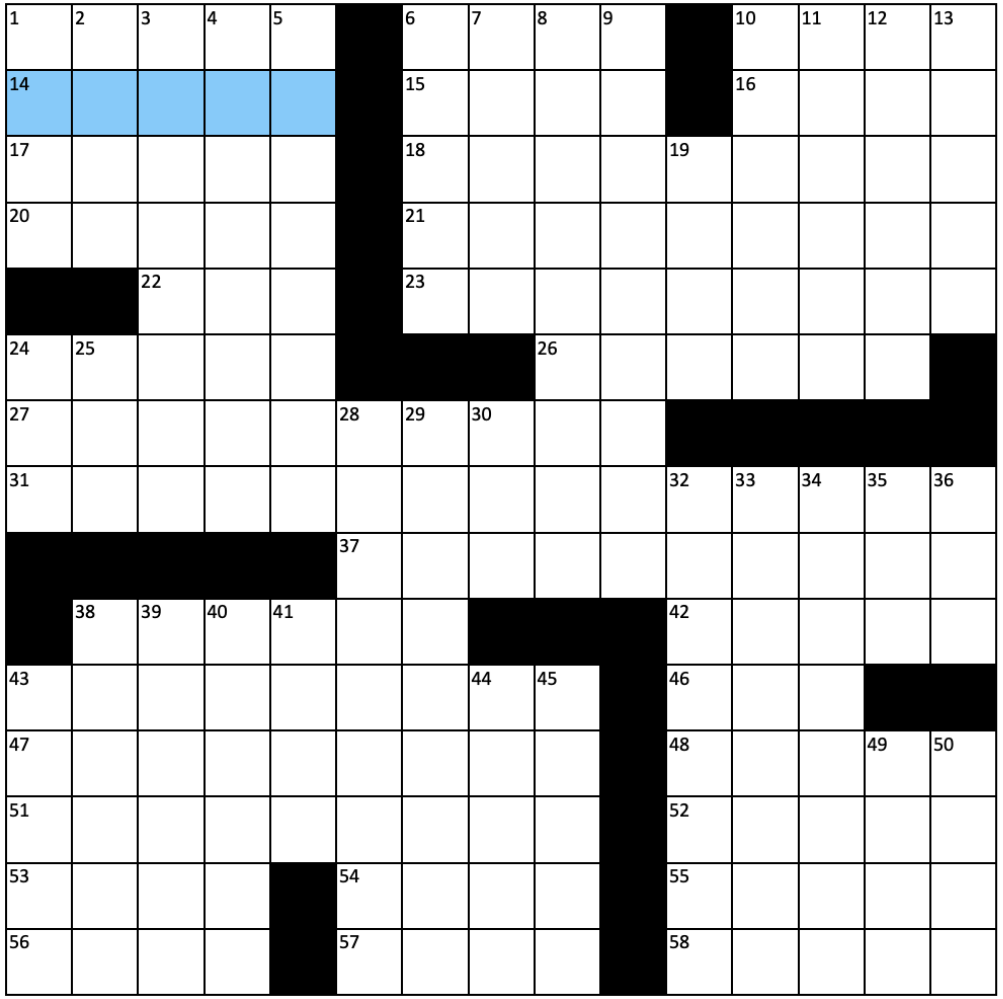
1 Regarding
6 Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi 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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

Example Crossword



Across

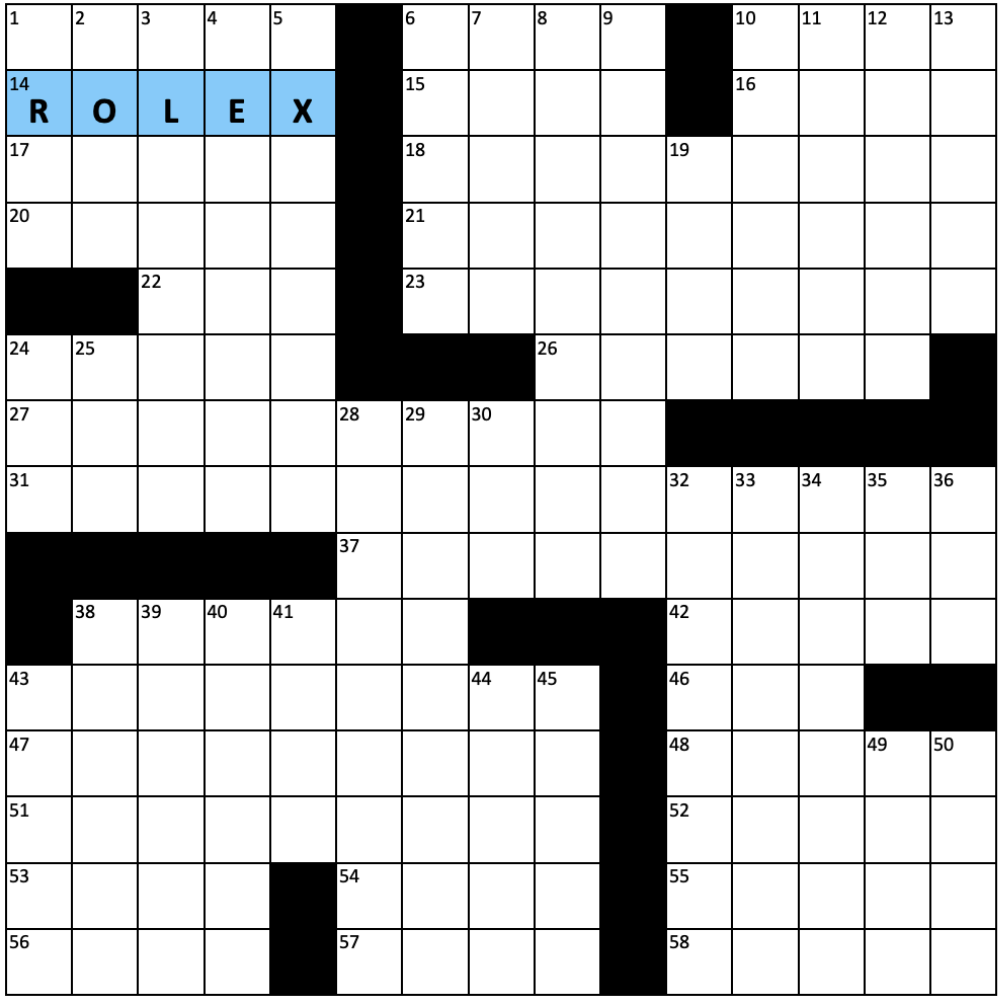
1 Regarding
6 Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi 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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

Example Crossword



Across

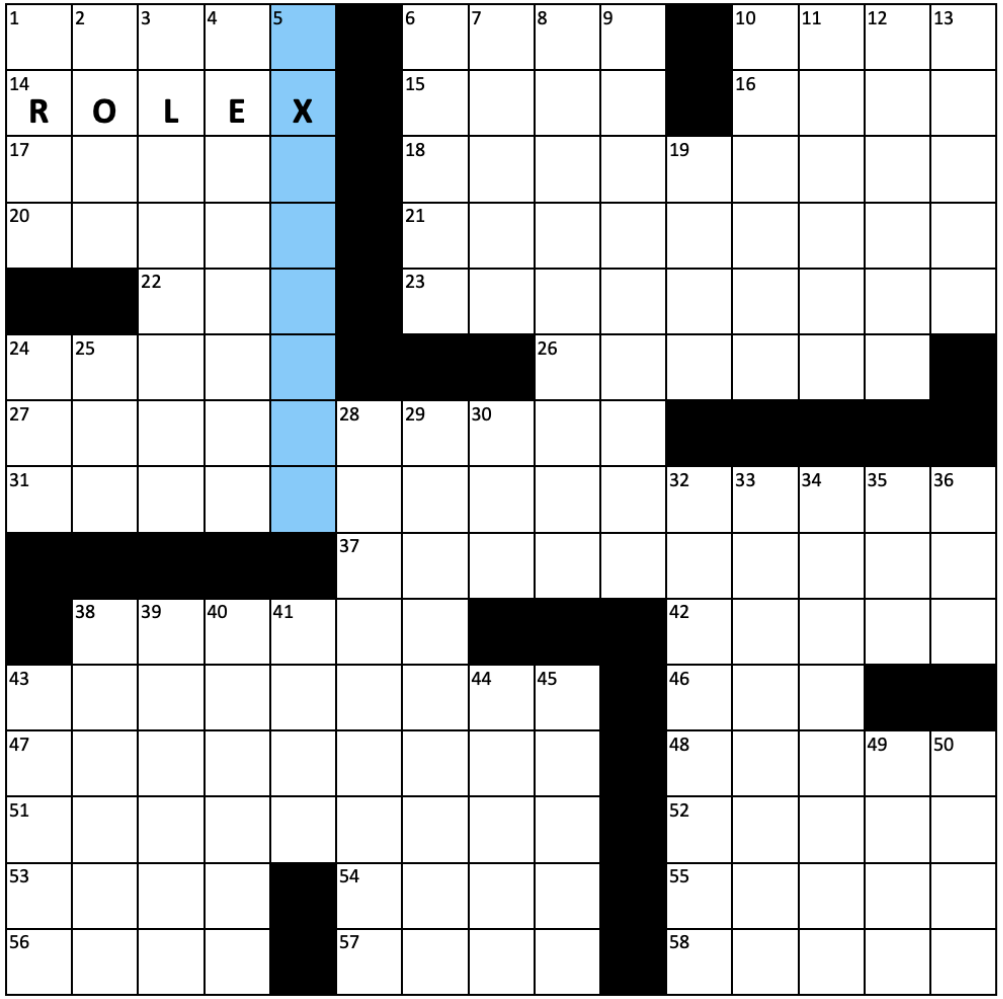
1 Regarding
6 Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi 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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

Example Crossword



Across

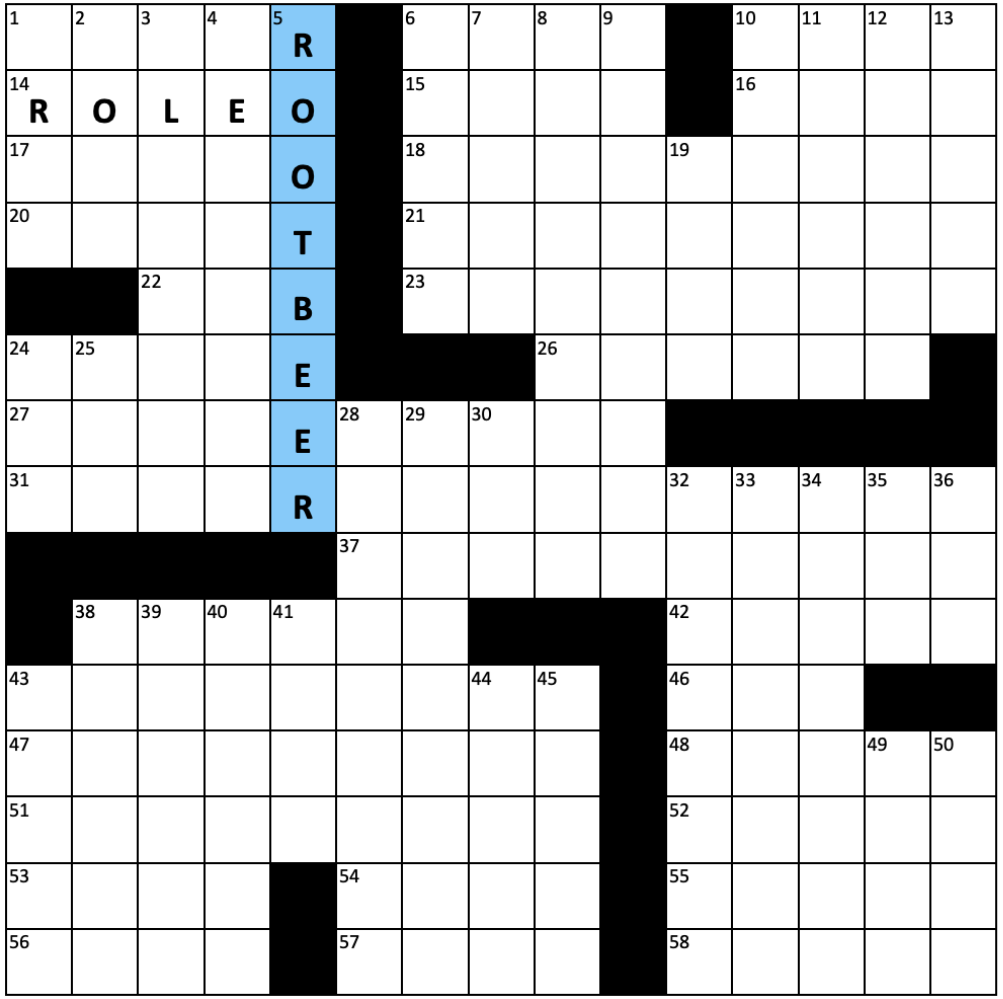
1 Regarding
6 Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi 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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

Example Crossword



Across

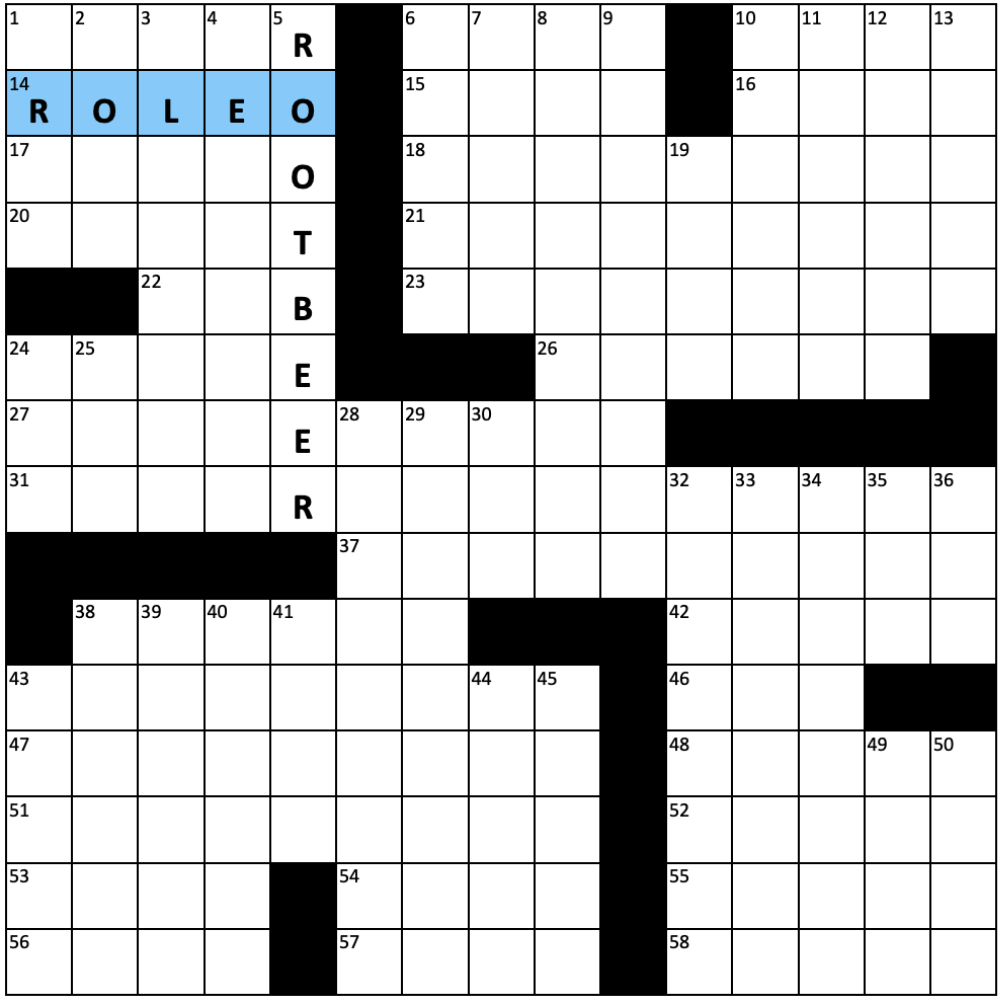
1 Regarding
6 Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi 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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

Example Crossword



Across

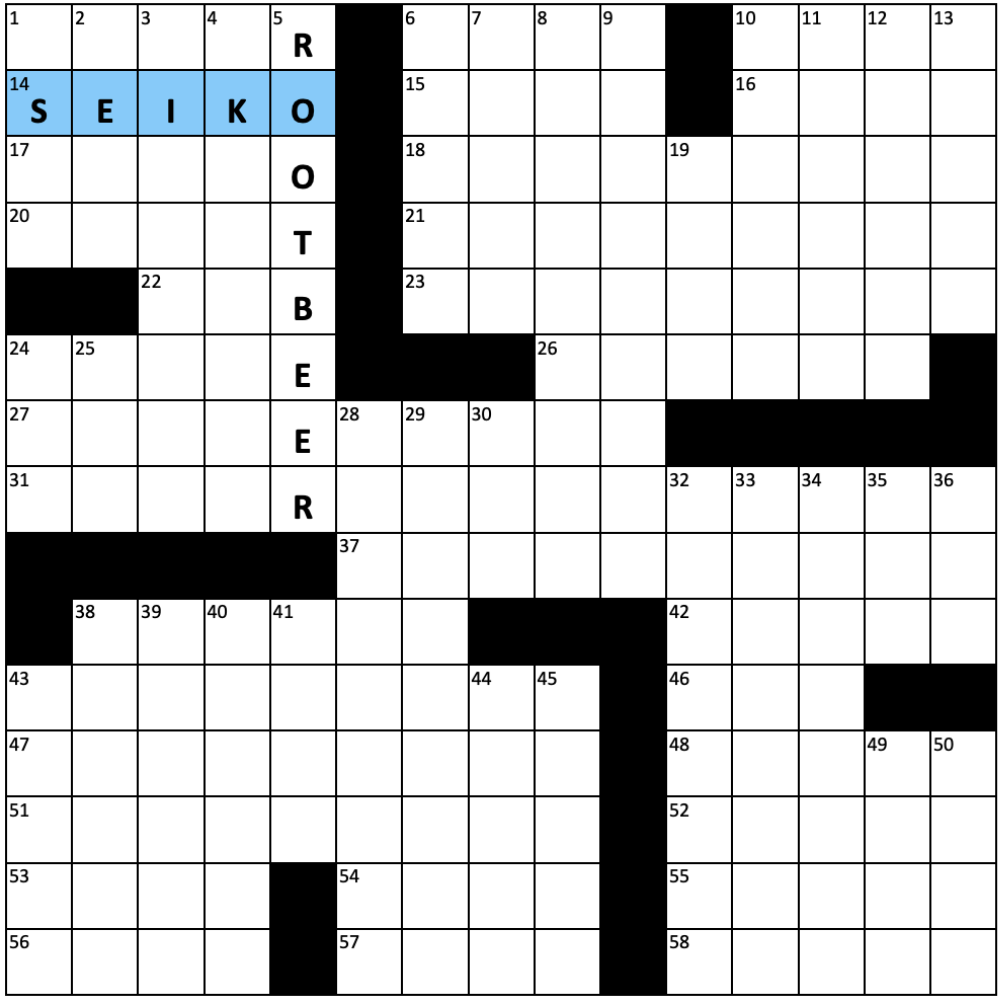
1 Regarding
6 Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi 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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

Example Crossword



Across

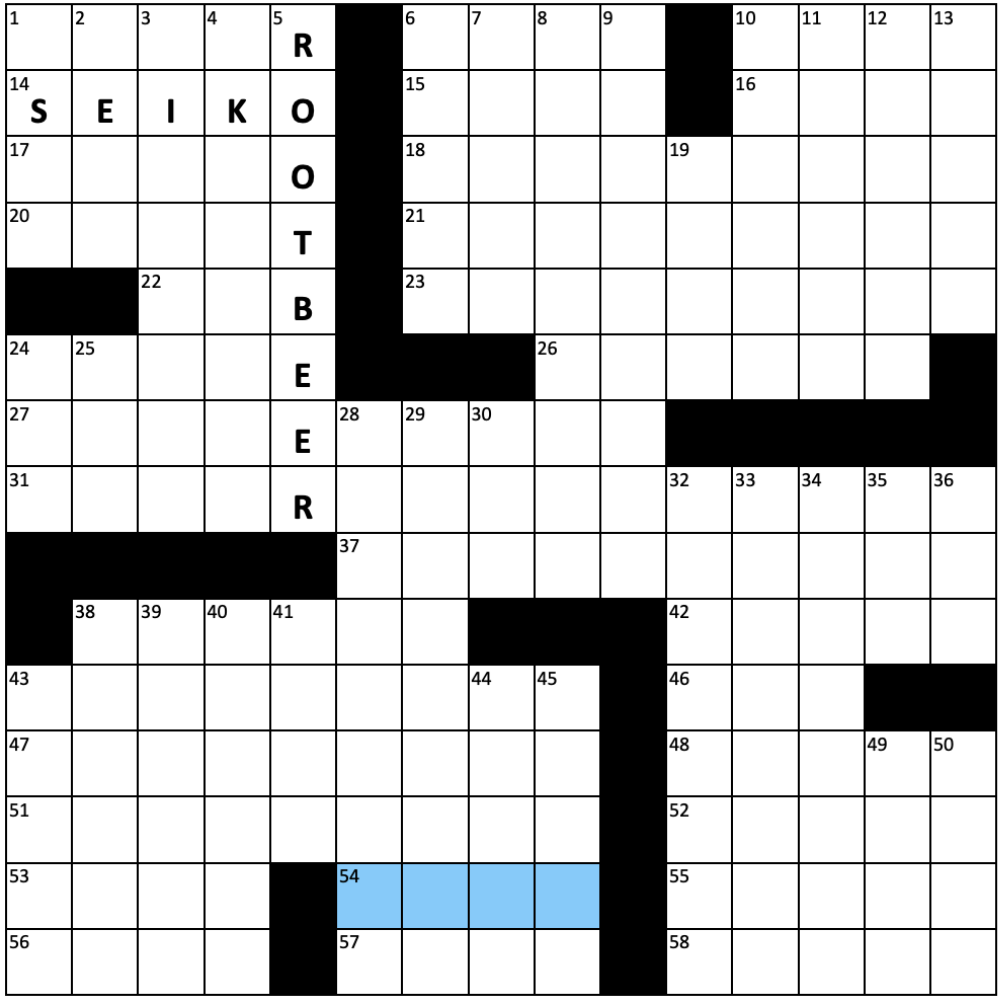
1 Regarding
6 Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi 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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

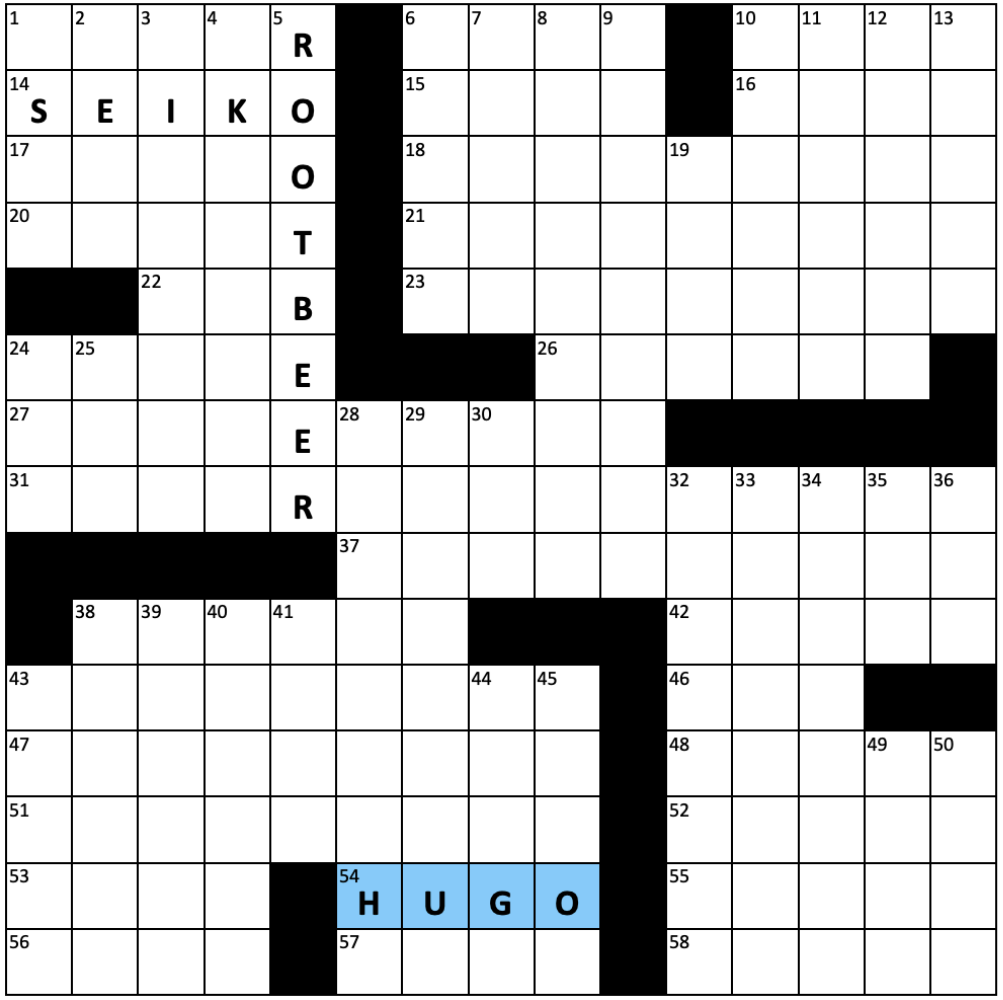
25 "___ noticed ..."

28 Great Hindu sage

29 Late-Triassic flier

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

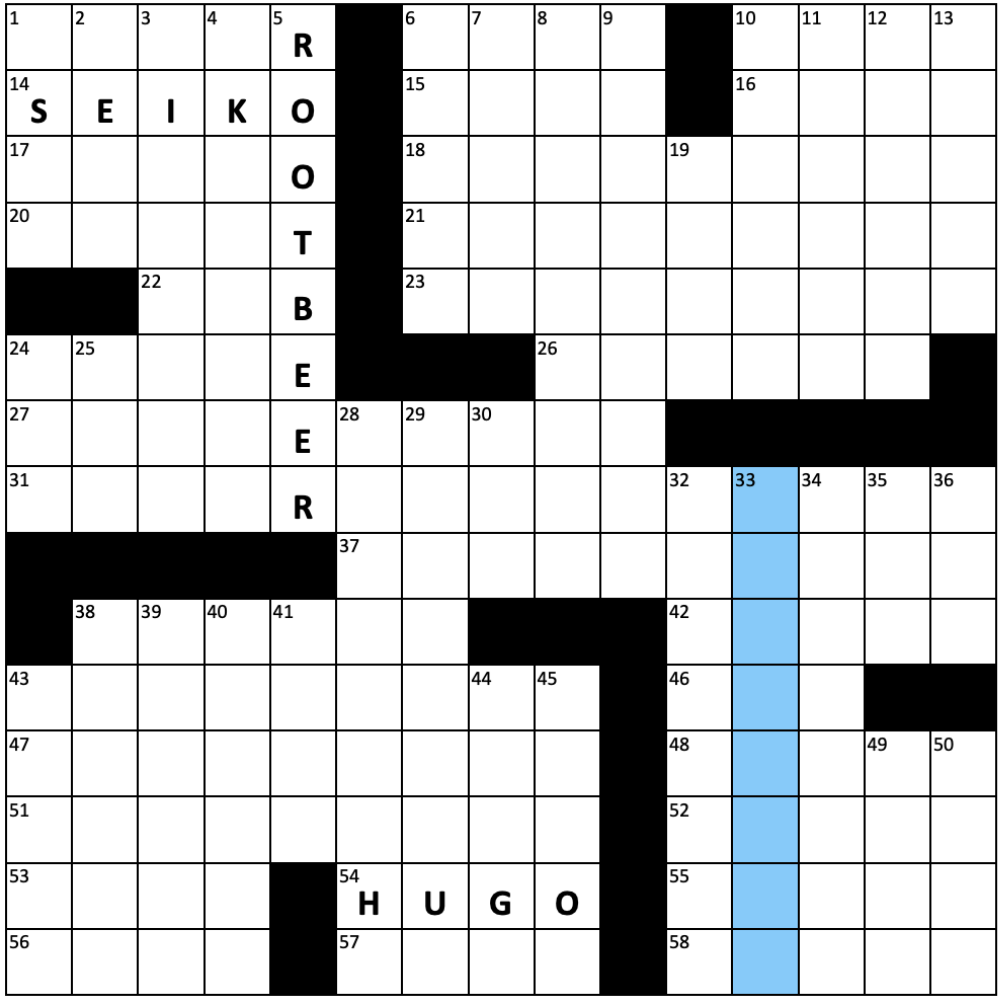
25 "___ noticed ..."

28 Great Hindu sage

29 Late-Triassic flier

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

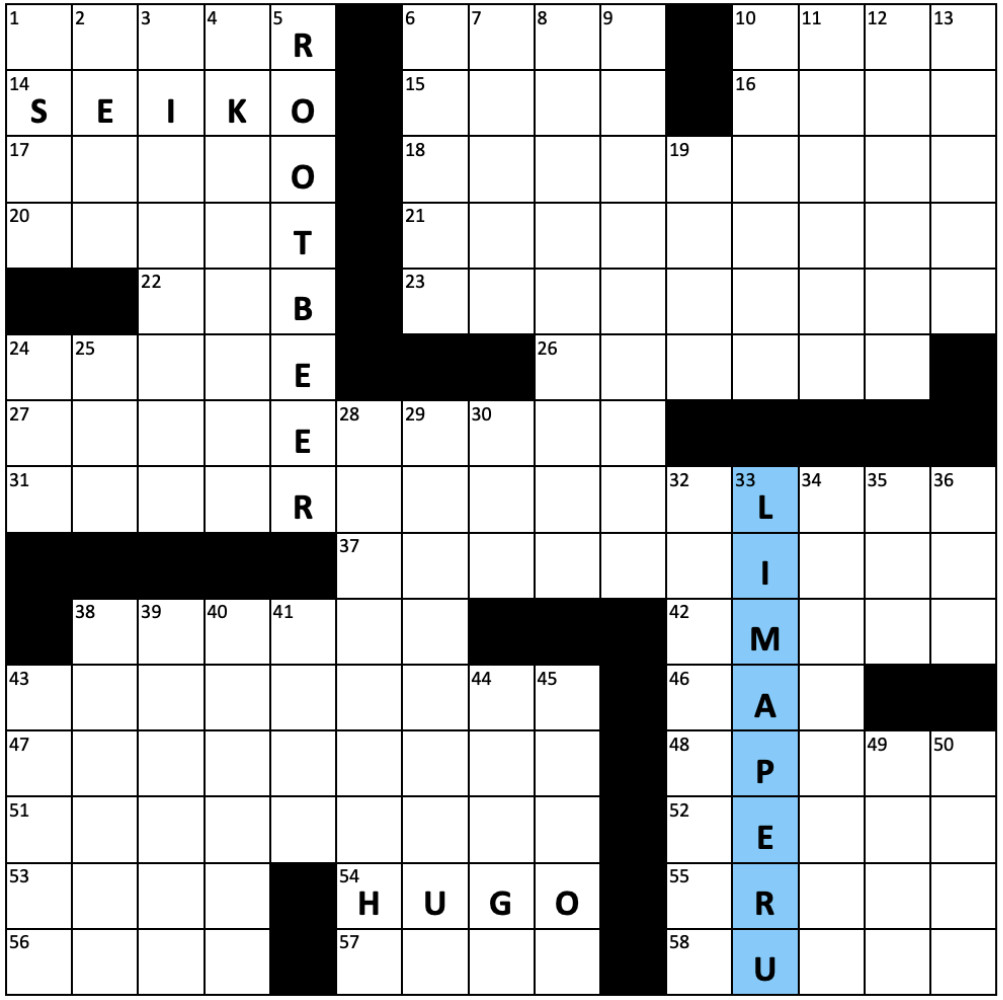
30 It can be found right on a clock

32 "100 Years ... 100 Stars" and others, in brief

33 The world's largest desert capital, after Cairo, Egypt

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

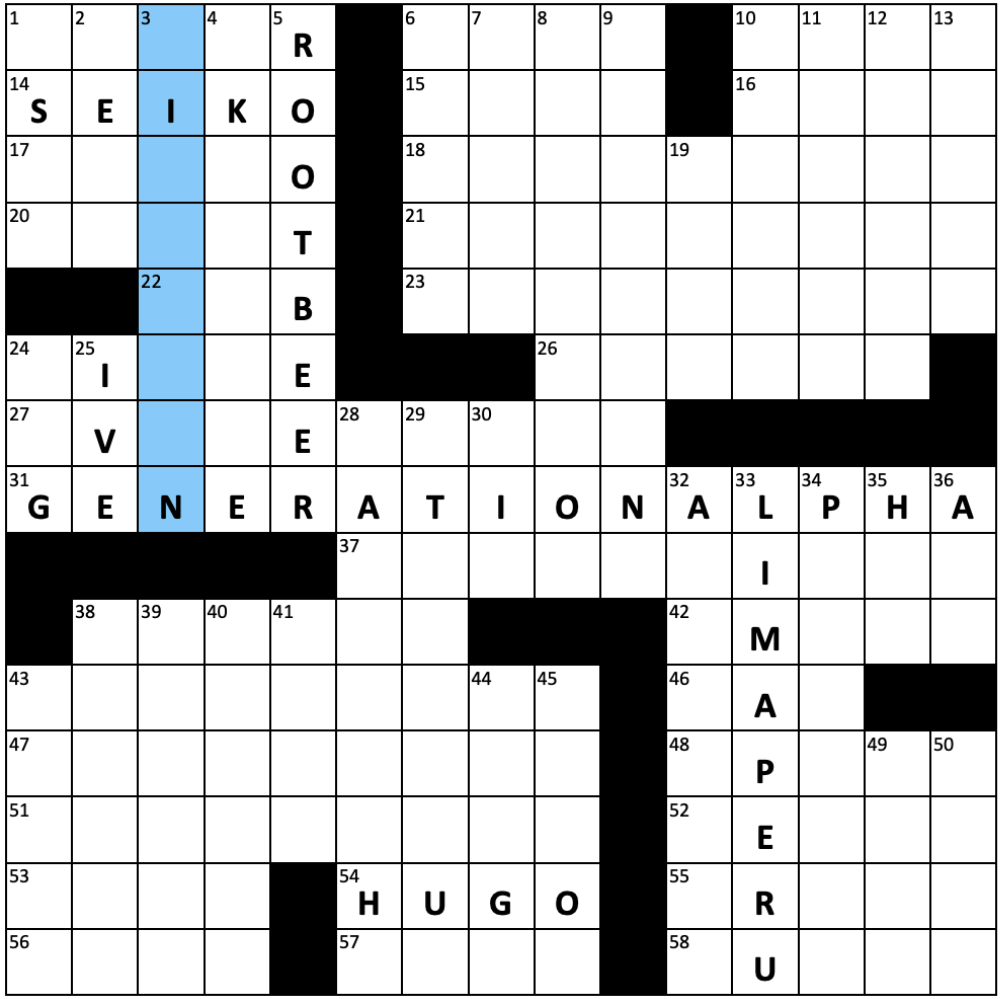
30 It can be found right on a clock

32 "100 Years ... 100 Stars" and others, in brief

33 The world's largest desert capital, after Cairo, Egypt

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

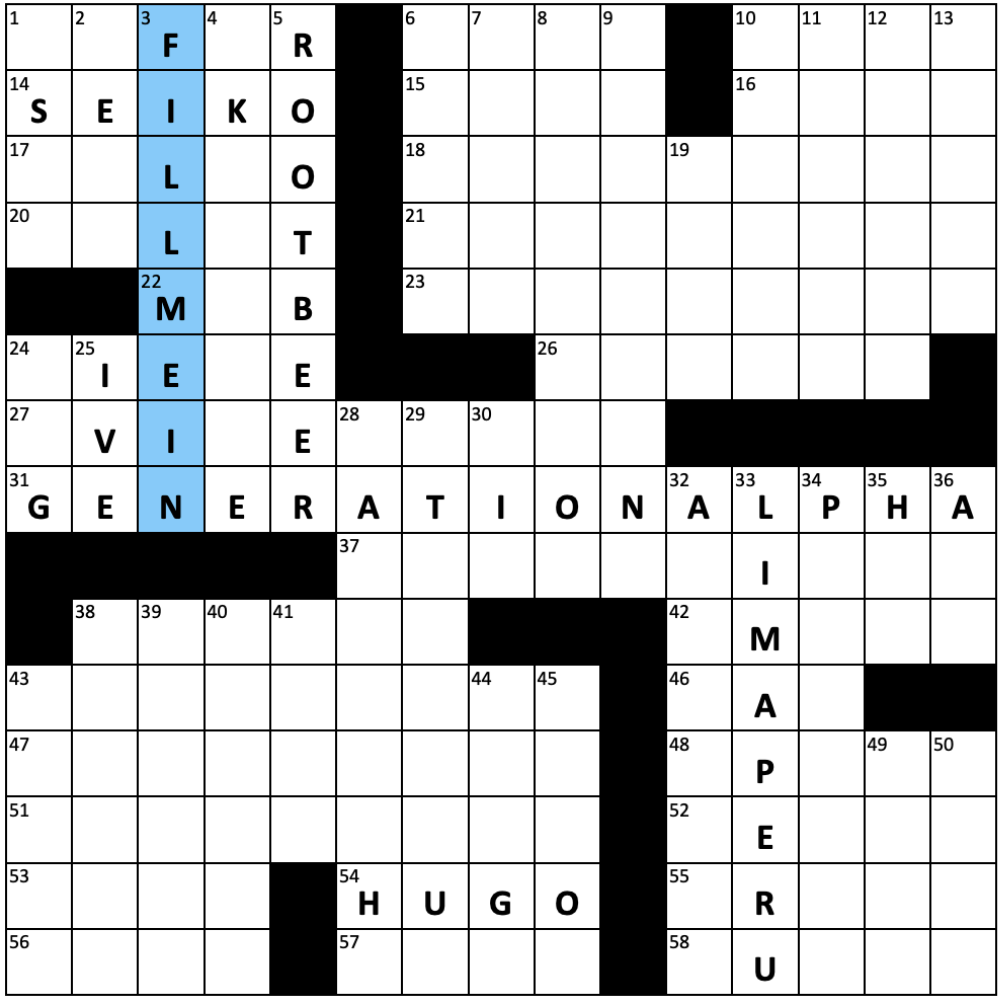
29 Late-Triassic flier

30 It can be found right on a clock

32 "100 Years ... 100 Stars" and others, in brief

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

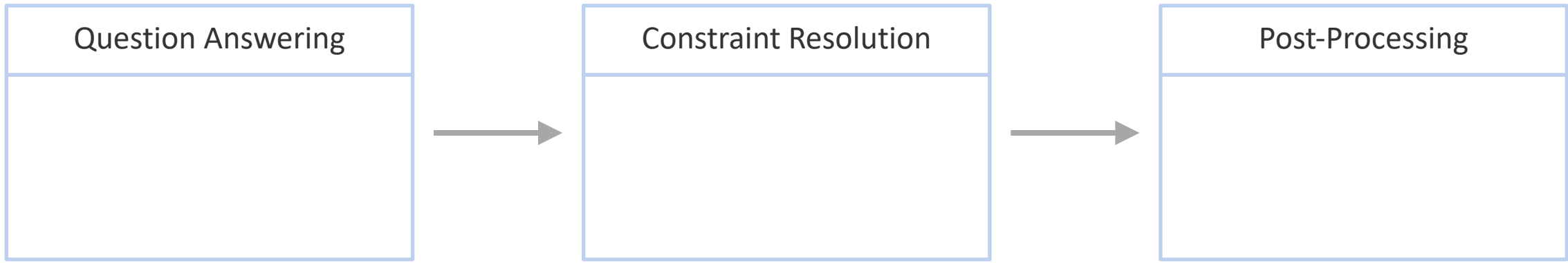
29 Late-Triassic flier

30 It can be found right on a clock

32 "100 Years ... 100 Stars" and others, in brief

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

General Framework



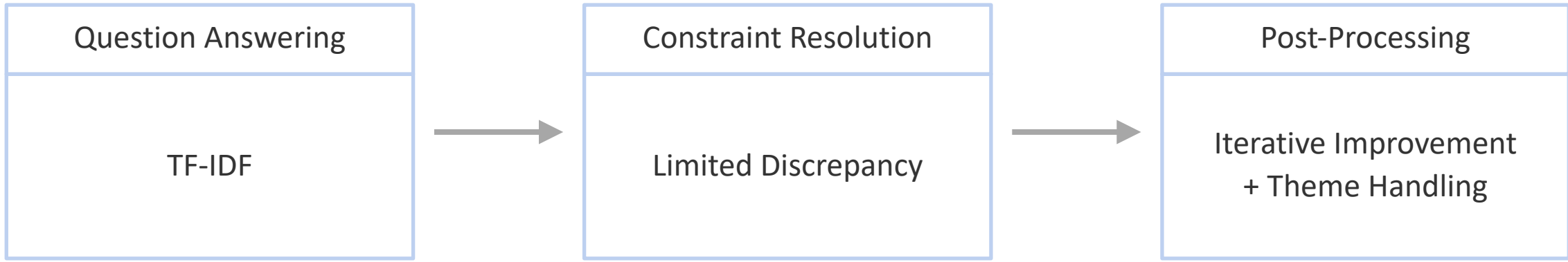
Previous Work

Proverb [Littman et al., 2002]

Question Answering	→	Constraint Resolution	→	Post-Processing
ExactMatch, Movies, Writers, Music, Bigram, Encyclopedia, WordNet, …		Belief Propagation + A* Search		

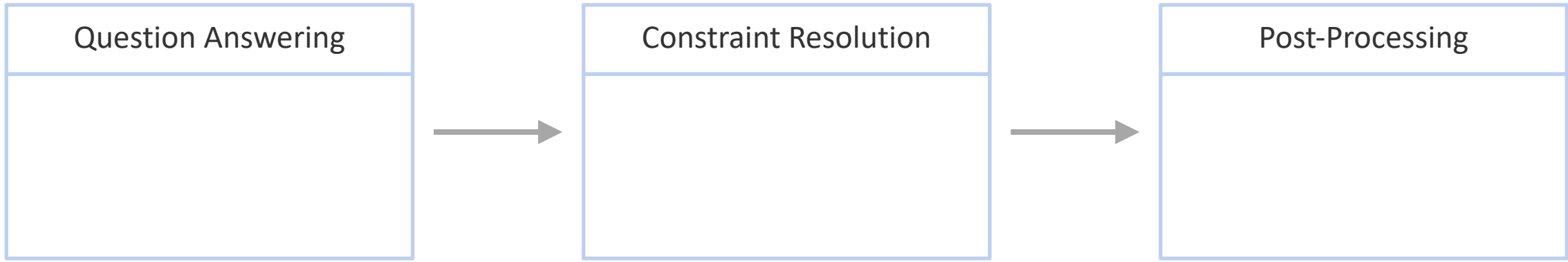
Previous Work

Dr. Fill [Ginsberg, 2011]



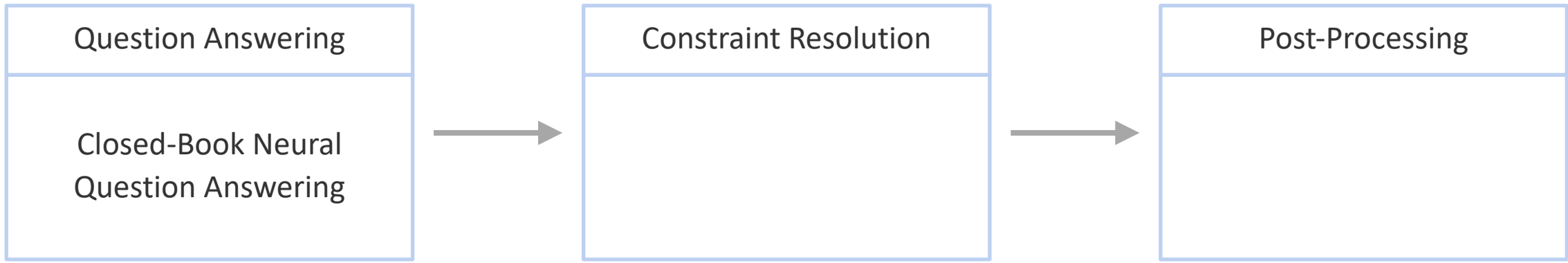
Our Approach

The Berkeley Crossword Solver



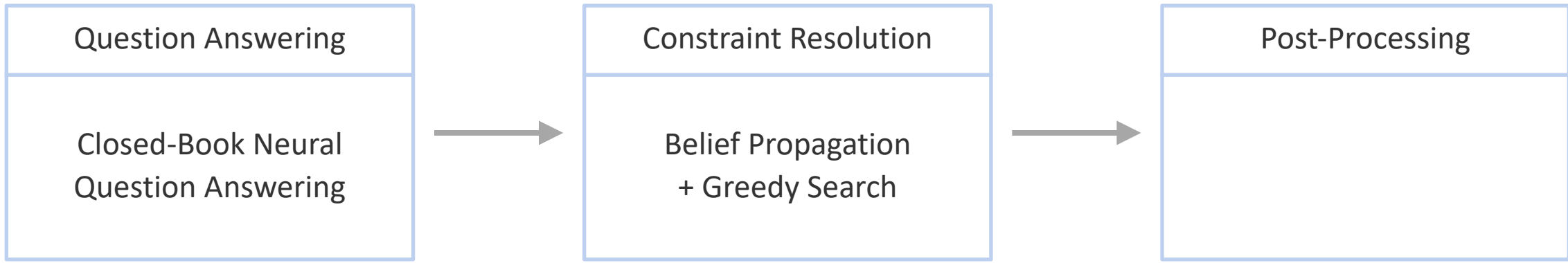
Our Approach

The Berkeley Crossword Solver



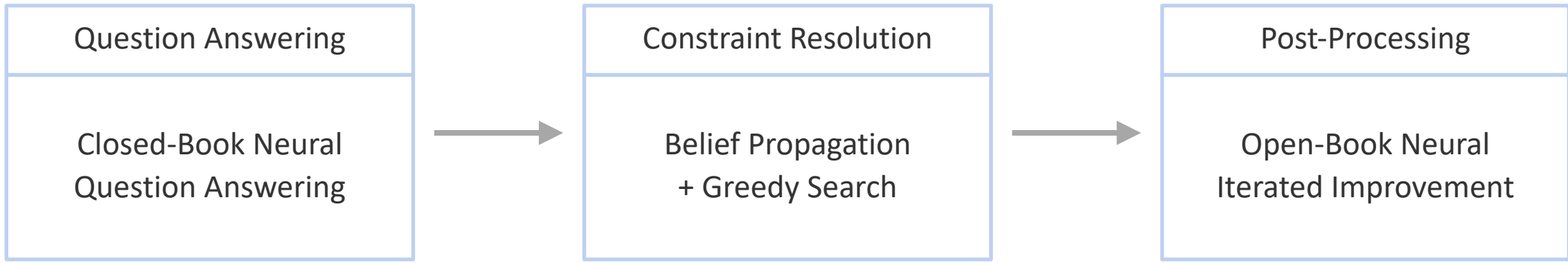
Our Approach

The Berkeley Crossword Solver



Our Approach

The Berkeley Crossword Solver



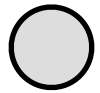
Our Approach

The Berkeley Crossword Solver

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

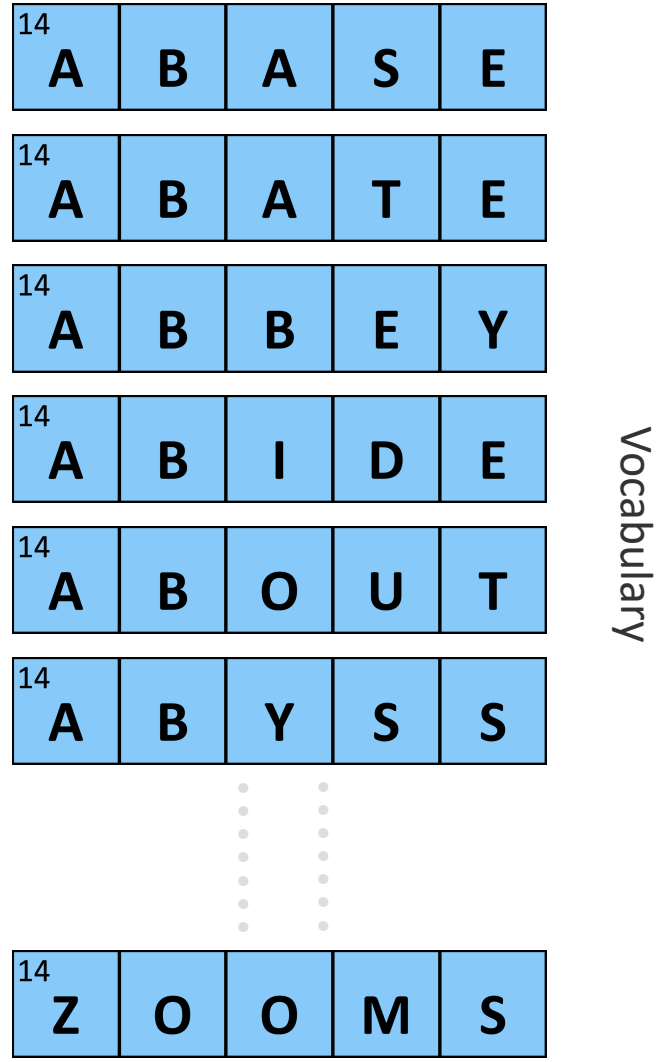
Closed-Book QA

14. Omega competitor



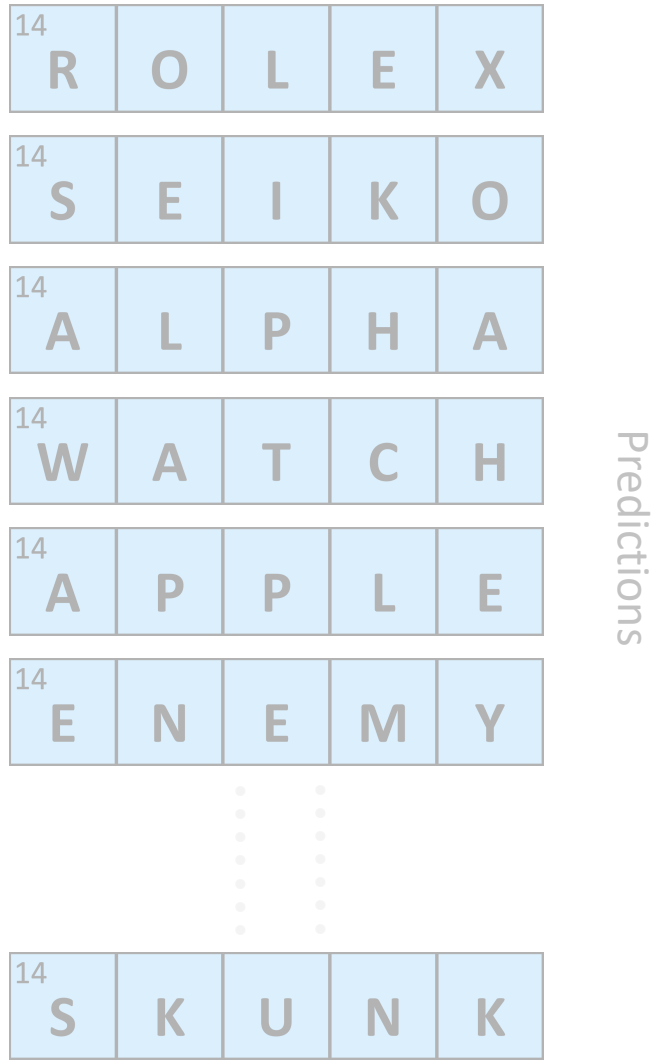
Closed-Book QA

14. Omega competitor



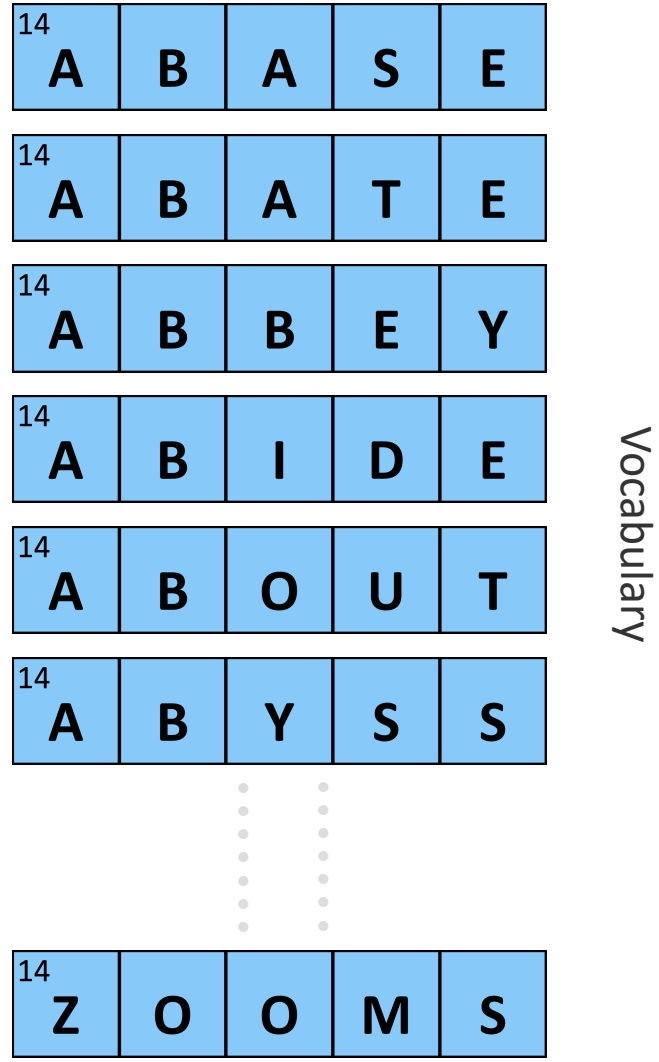
Closed-Book QA

14. Omega competitor



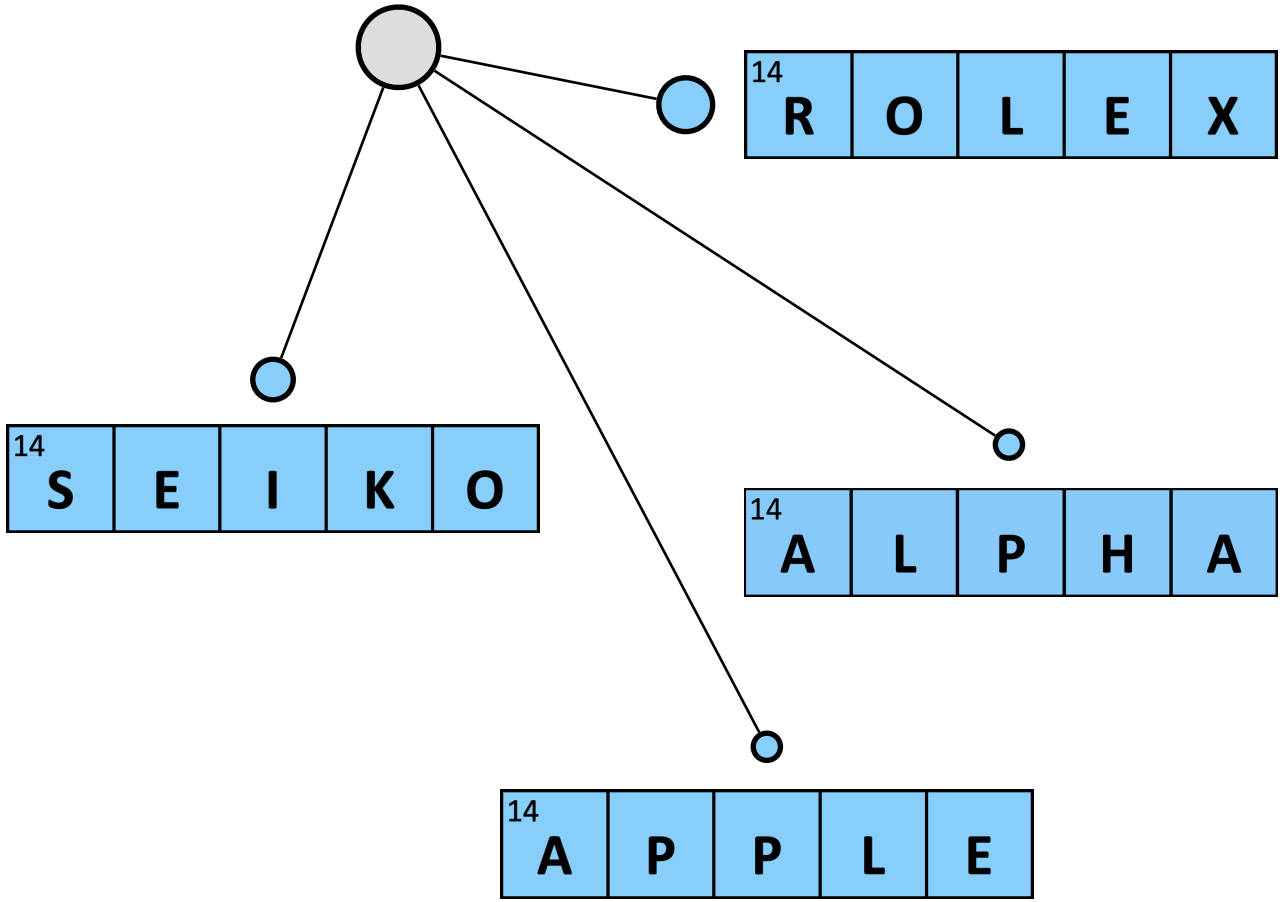
Closed-Book QA

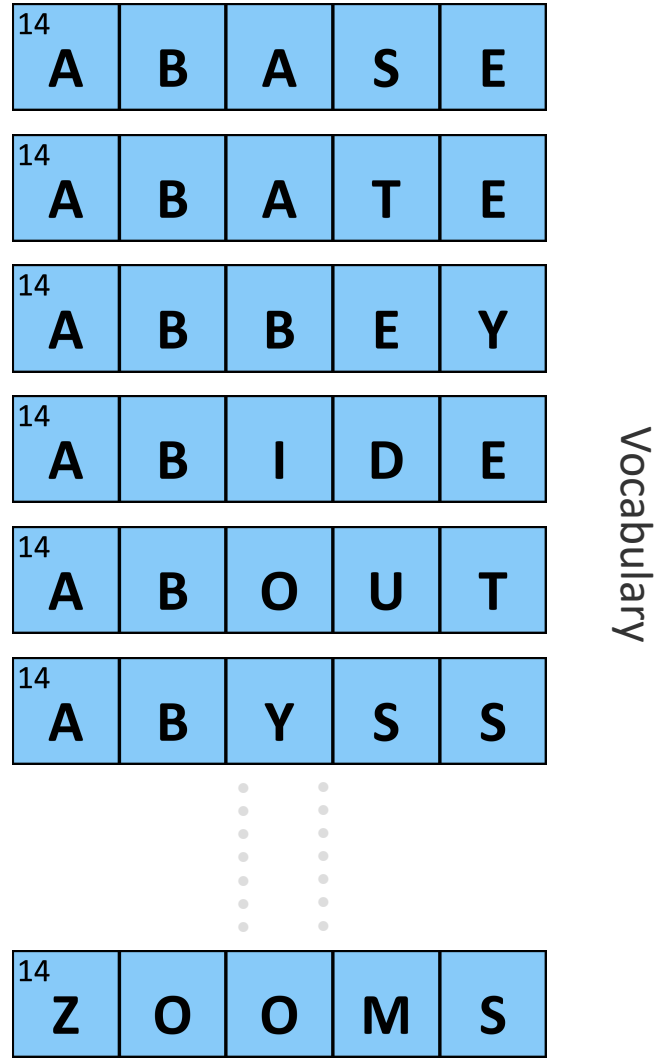
14. Omega competitor



Closed-Book QA

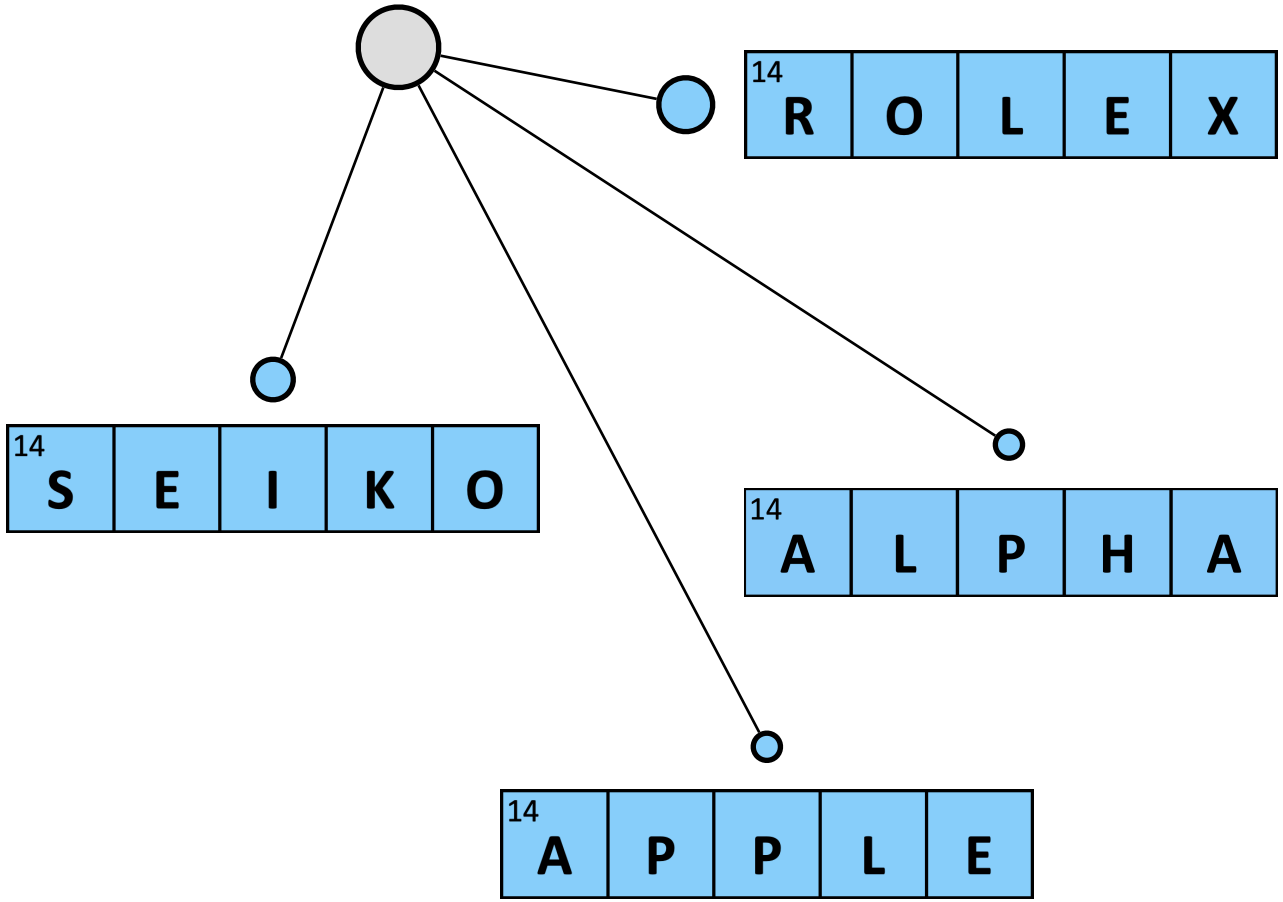
14. Omega competitor

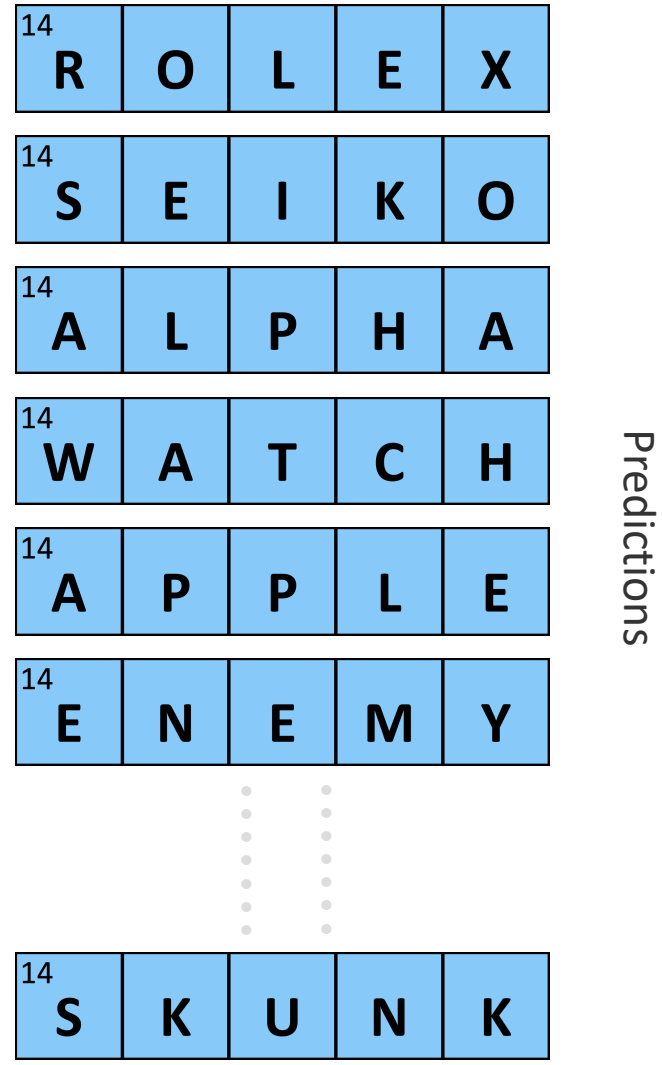




Closed-Book QA

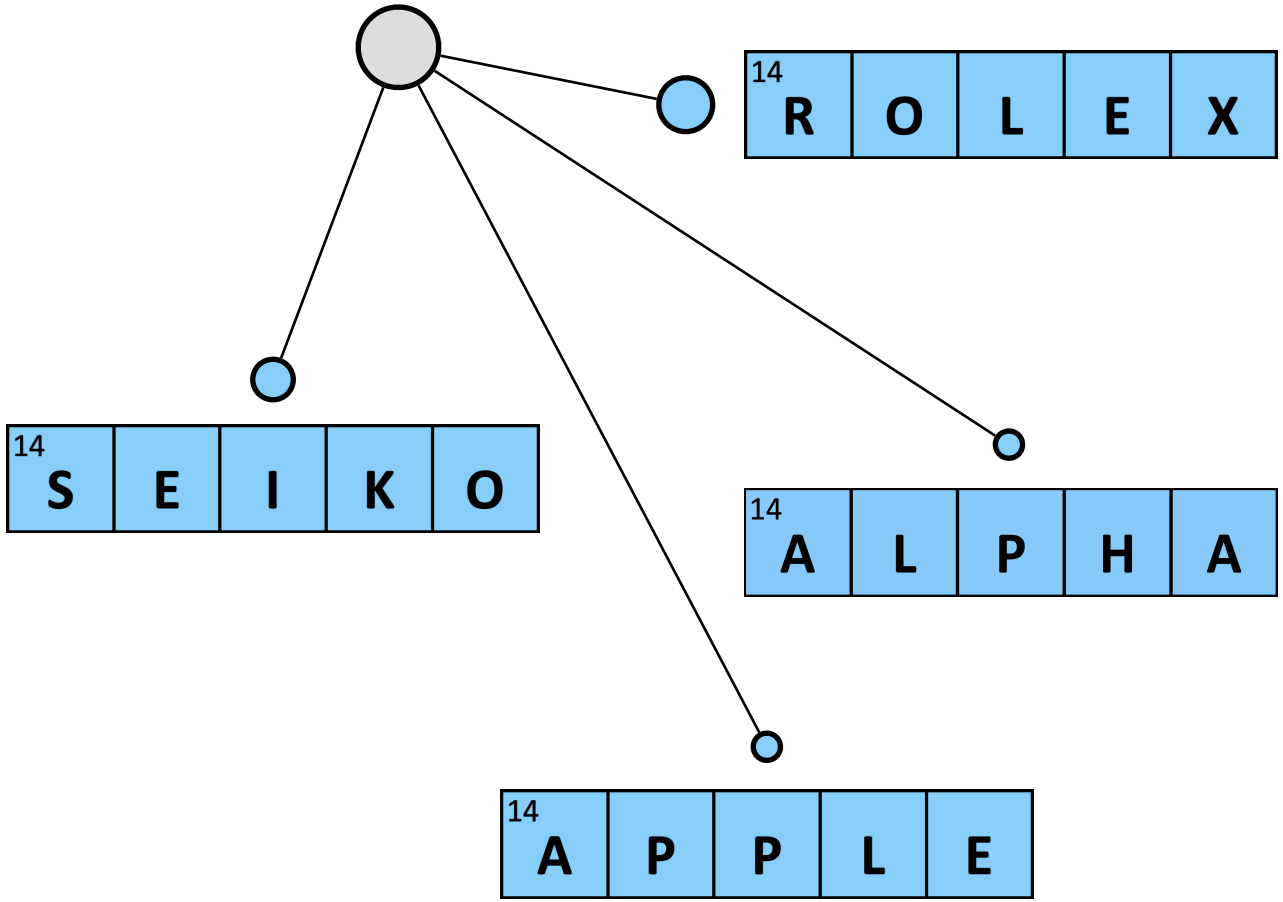
14. Omega competitor

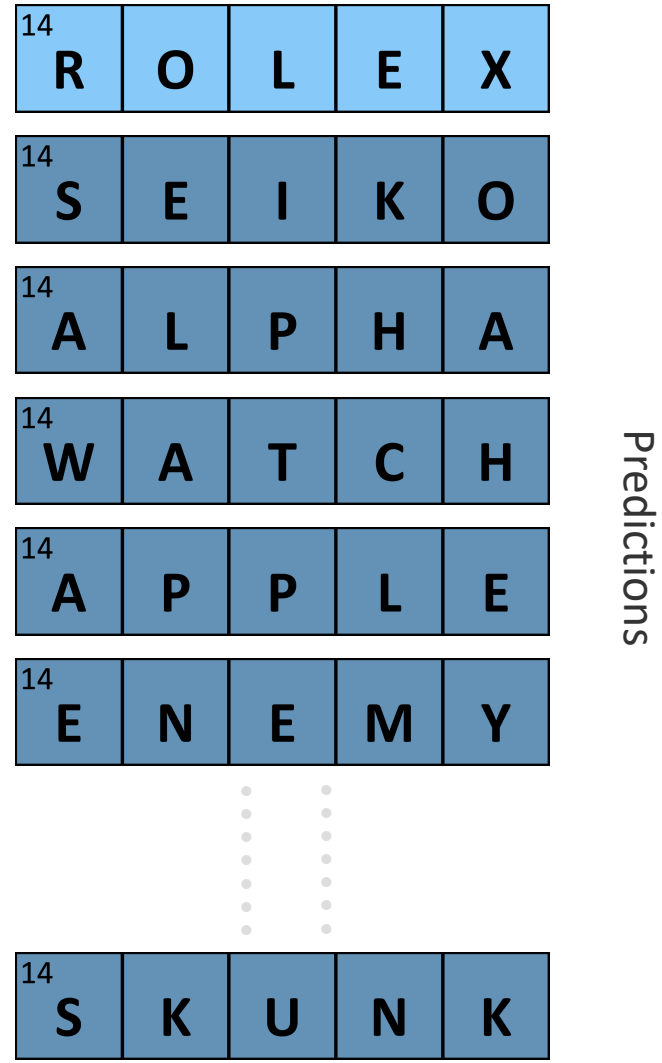




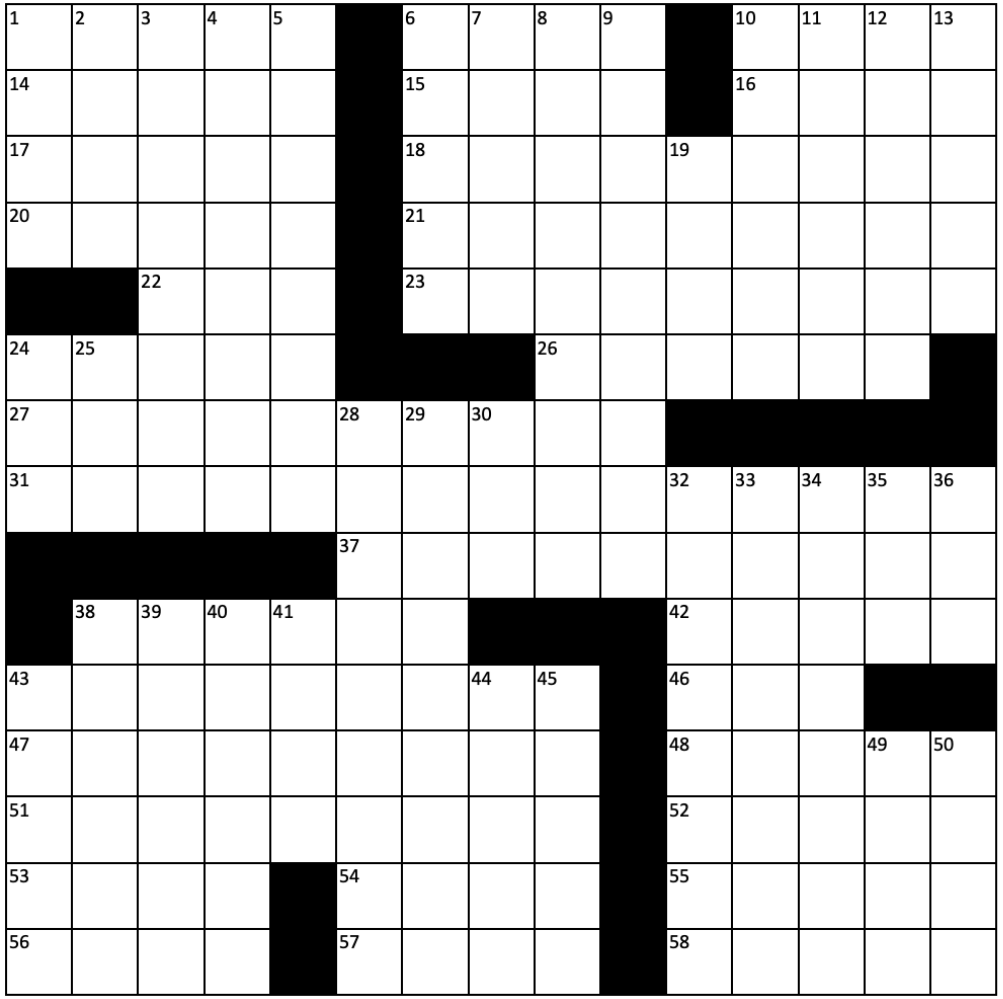
Closed-Book QA

14. Omega competitor





Handling Ambiguous Clues



Across

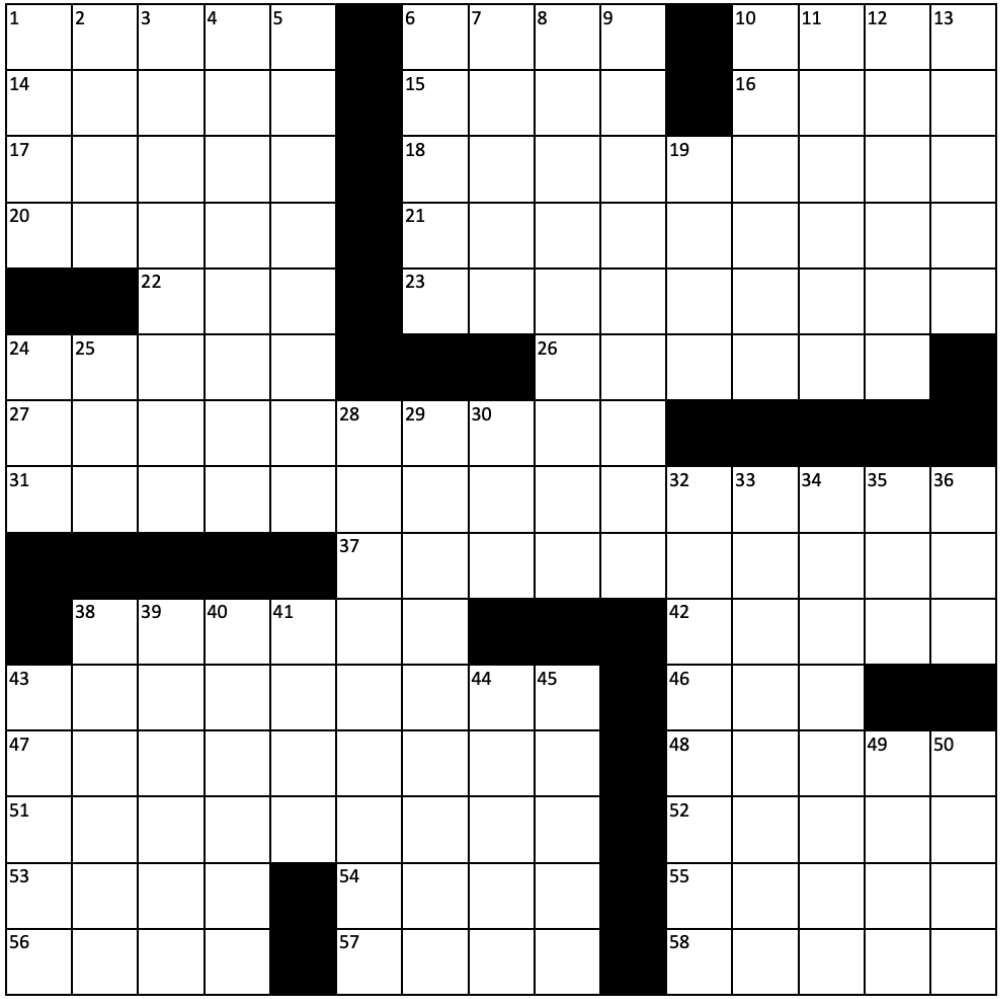
1 Regarding
6 Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi 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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

Handling Ambiguous Clues



Across

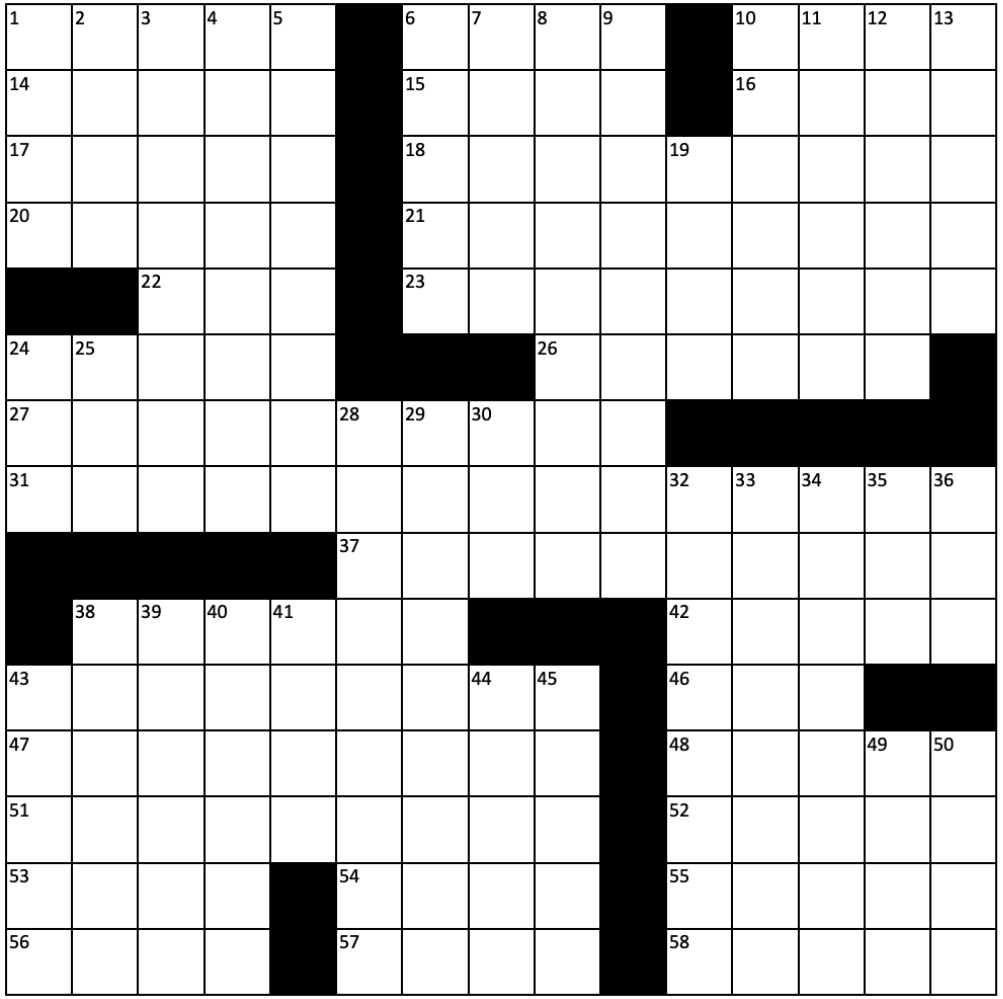
1 Regarding
6 Take back, in a way
10 Start of an aside
14 Omega competitor
15 Something hitting a nerve?
16 Papyrus, e.g.
17 "Take me with you!"
18 Begin flirting with someone, so to speak
20 Assign
21 Hoth, in "Star Wars"
22 ___ rule
23 They don't hold water
24 Feudal figure
26 Panegyric, e.g.
27 Sci-fi 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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

Our Approach

The Berkeley Crossword Solver

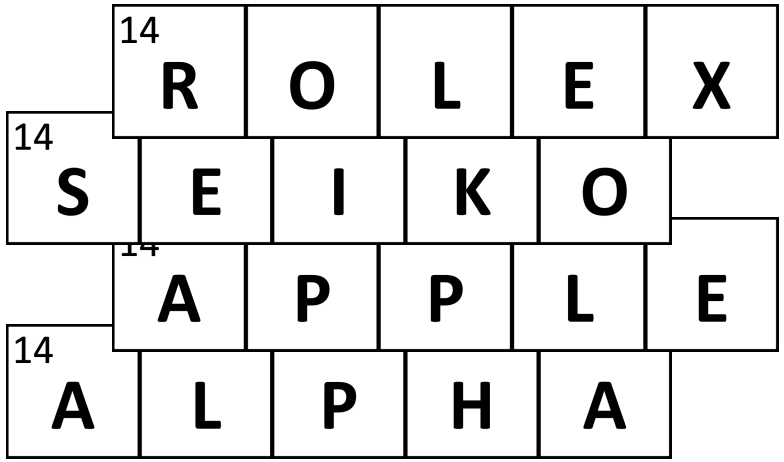
Question Answering	Constraint Resolution	Post-Processing
Closed-Book Neural Question Answering	Belief Propagation + Greedy Search	Open-Book Neural Iterated Improvement

Our Approach

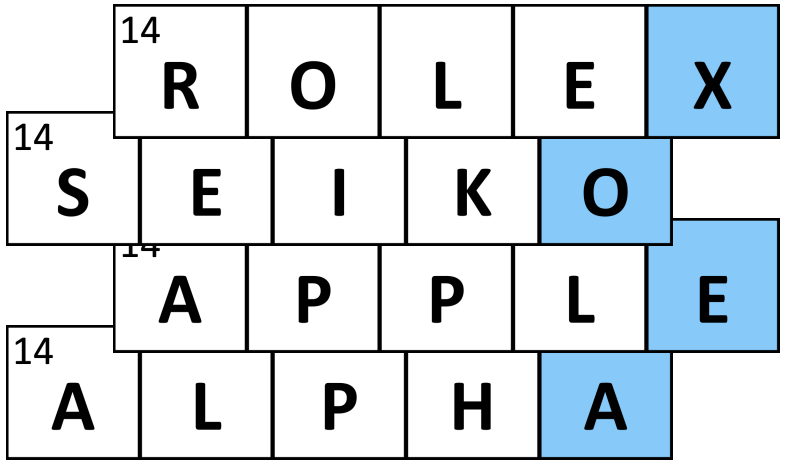
The Berkeley Crossword Solver

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

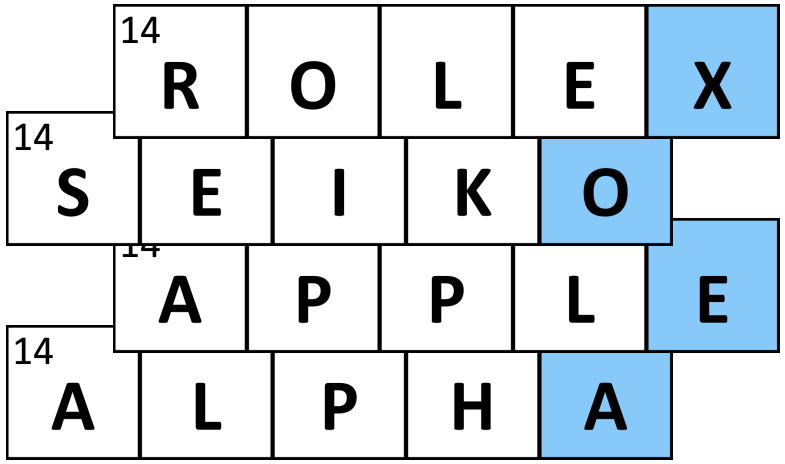
Loopy Belief Propagation

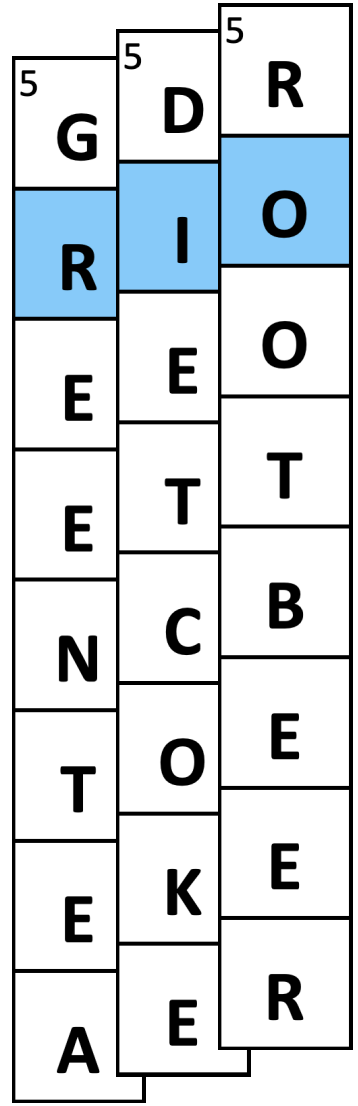


Loopy Belief Propagation

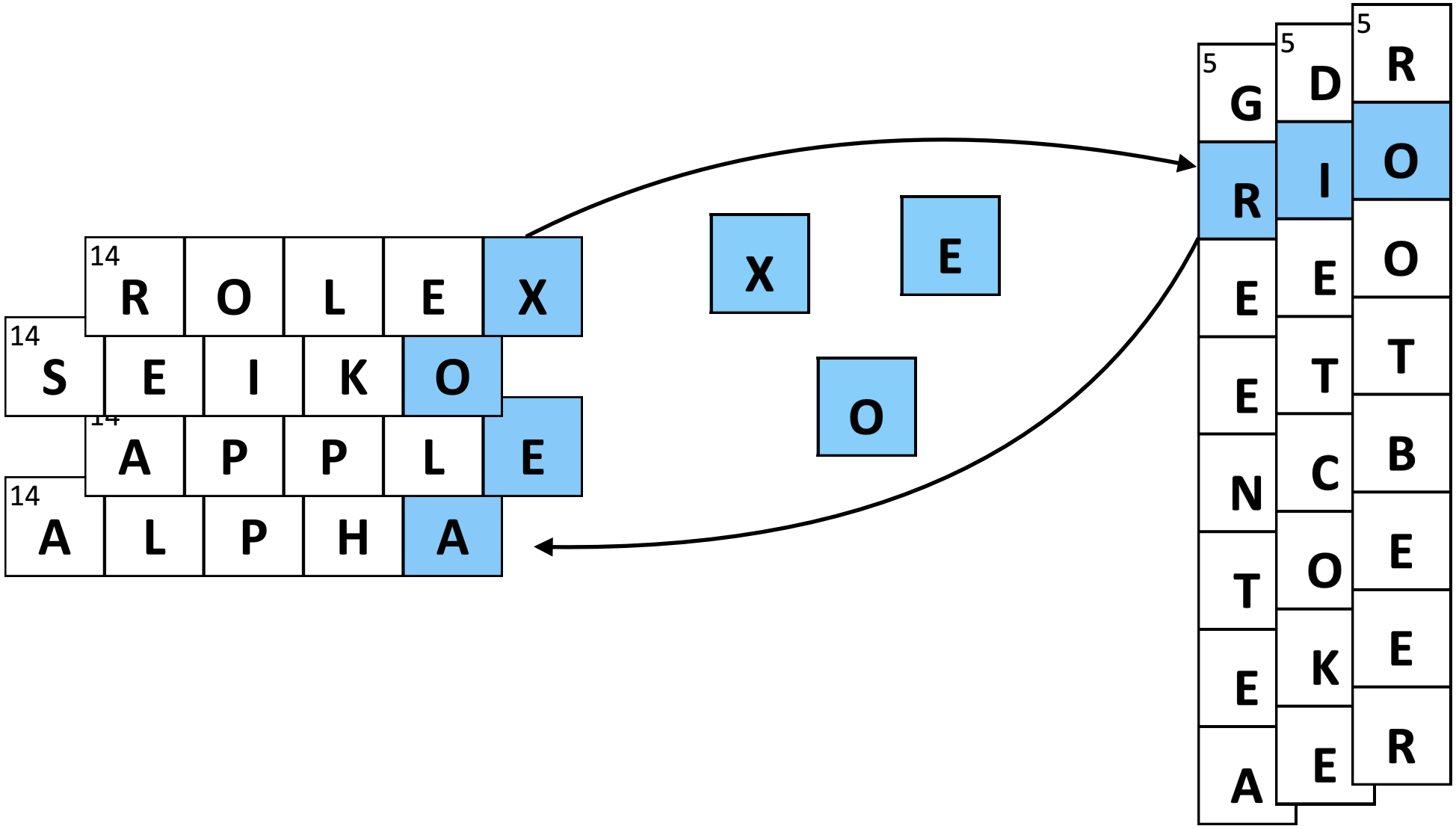


Loopy Belief Propagation

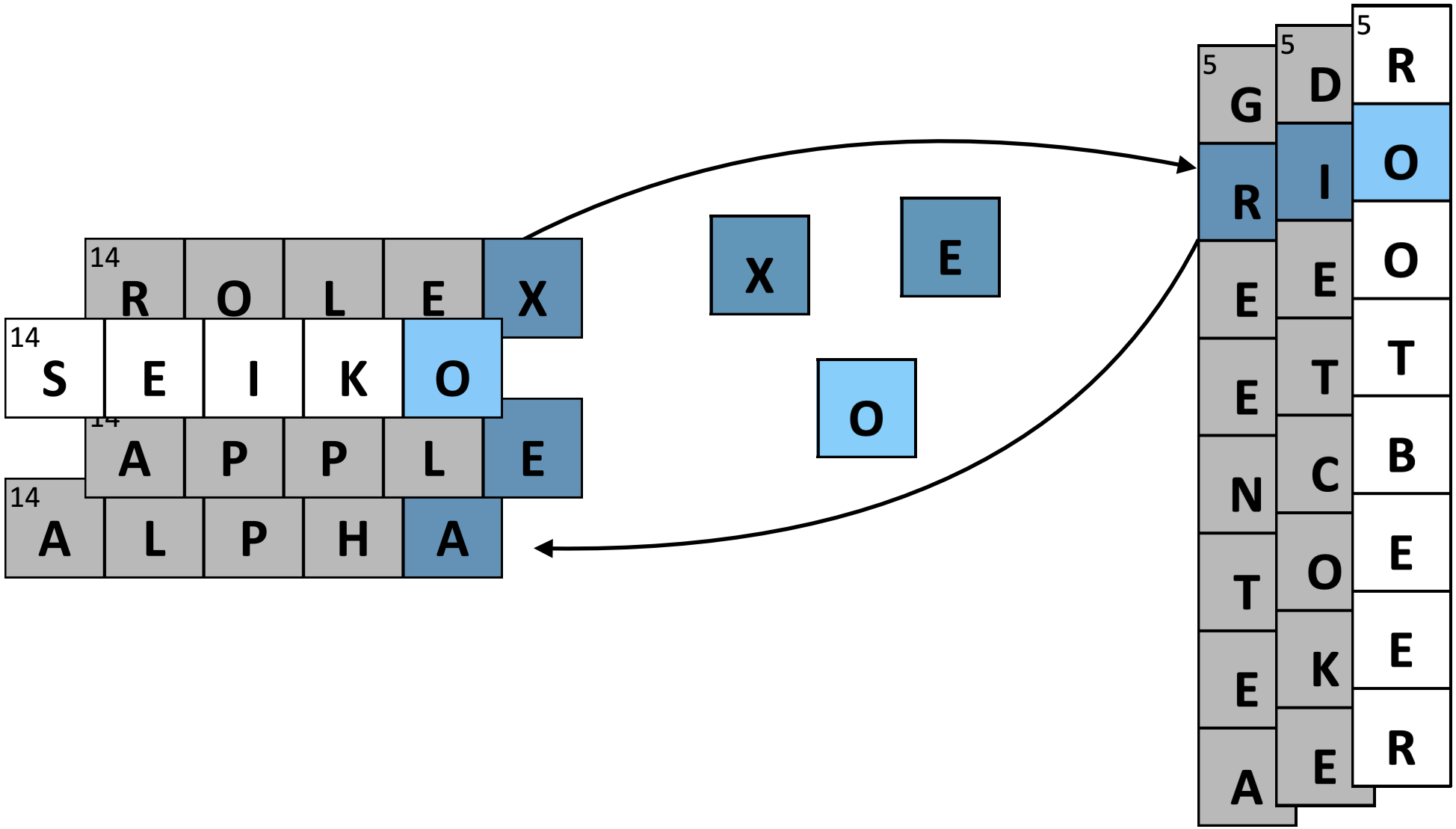




Loopy Belief Propagation



Loopy Belief Propagation



Our Approach

The Berkeley Crossword Solver

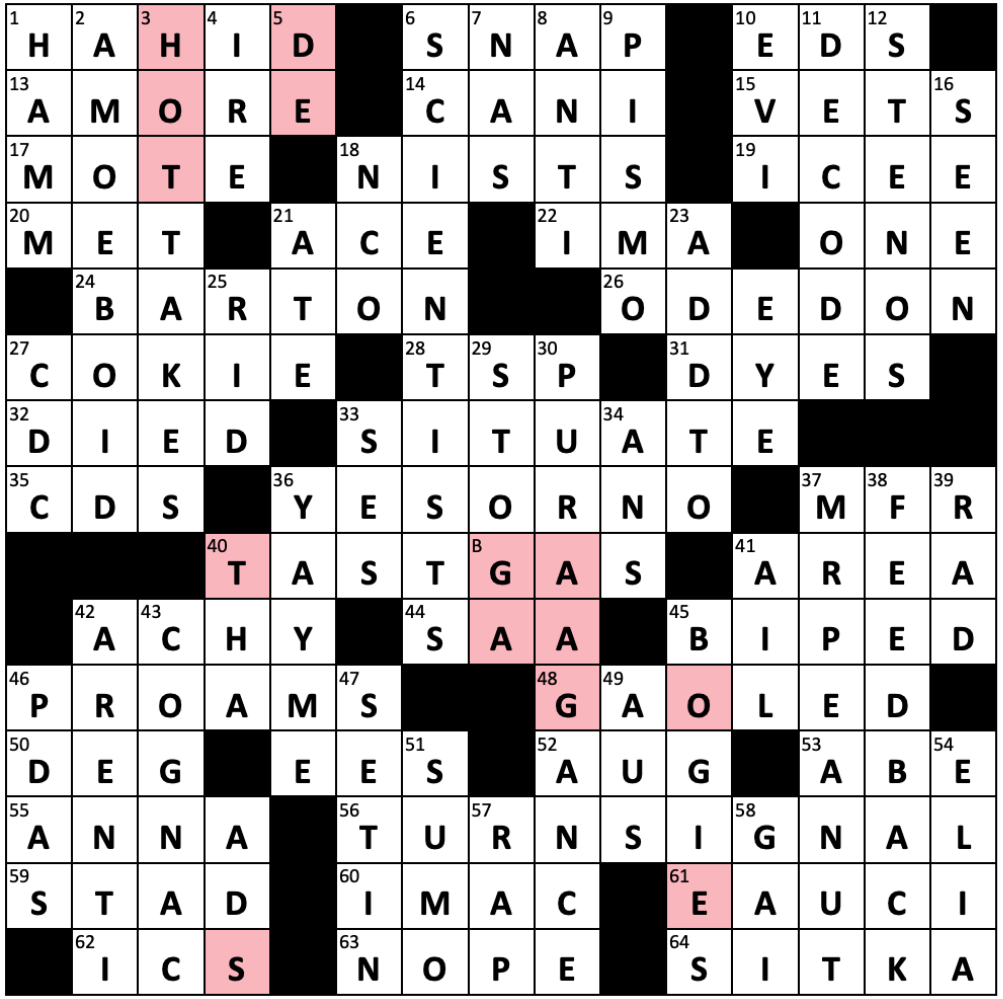
Question Answering	Constraint Resolution	Post-Processing
Closed-Book Neural Question Answering	Belief Propagation + Greedy Search	Open-Book Neural Iterated Improvement

Our Approach

The Berkeley Crossword Solver

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

The New York Times Crossword from Thursday, January 21, 2021. Constructed by Daniel Mauer and edited by Will Shortz.

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-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

The New York Times Crossword from Thursday, January 21, 2021. Constructed by Daniel Mauer and edited by Will Shortz.

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-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

The New York Times Crossword from Thursday, January 21, 2021. Constructed by Daniel Mauer and edited by Will Shortz.

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-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

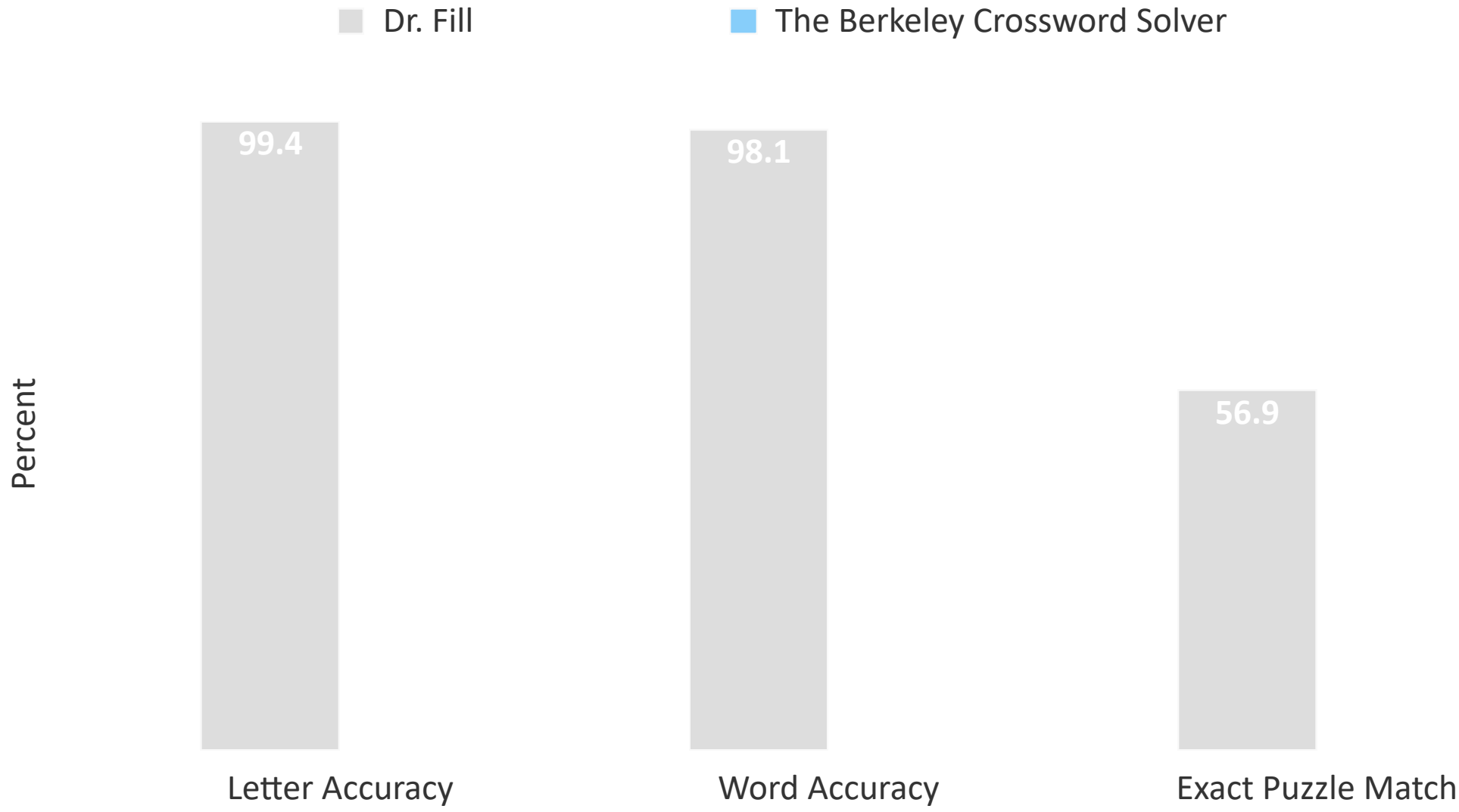
25 Purge (of)

27 Atlanta-based health org.

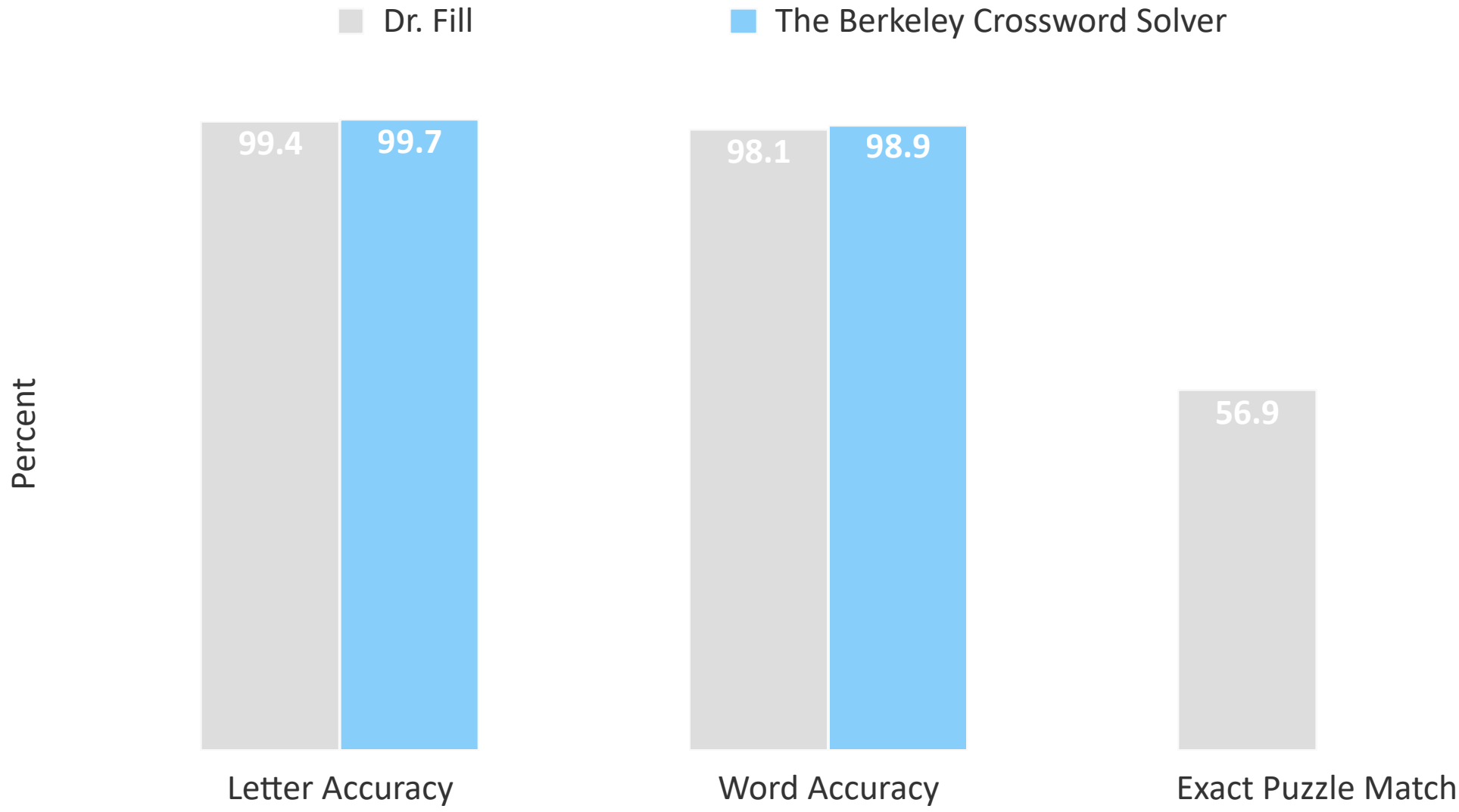
29 Long fur scarfs

The New York Times Crossword from Thursday, January 21, 2021. Constructed by Daniel Mauer and edited by Will Shortz.

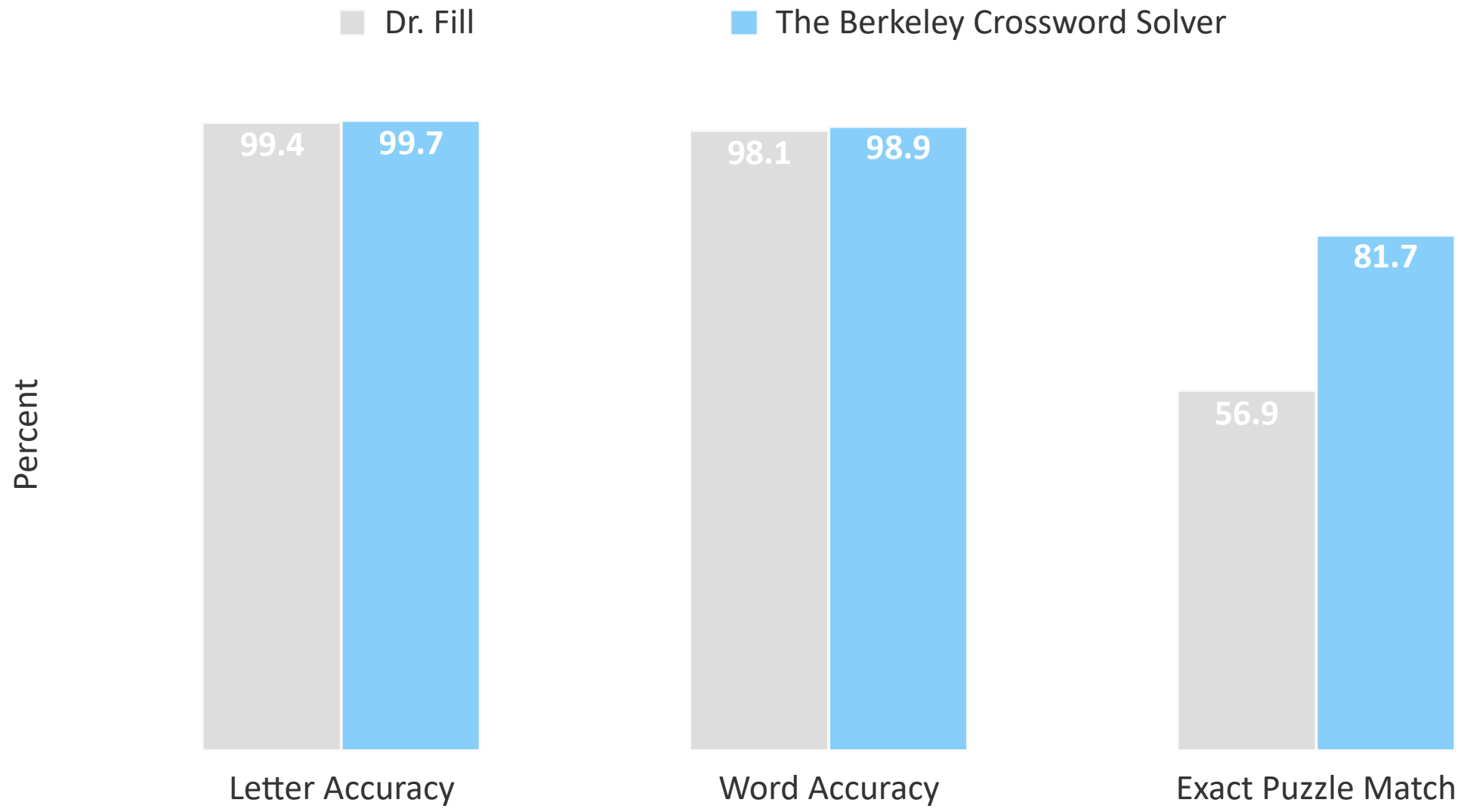
Results: The New York Times



Results: The New York Times



Results: The New York Times



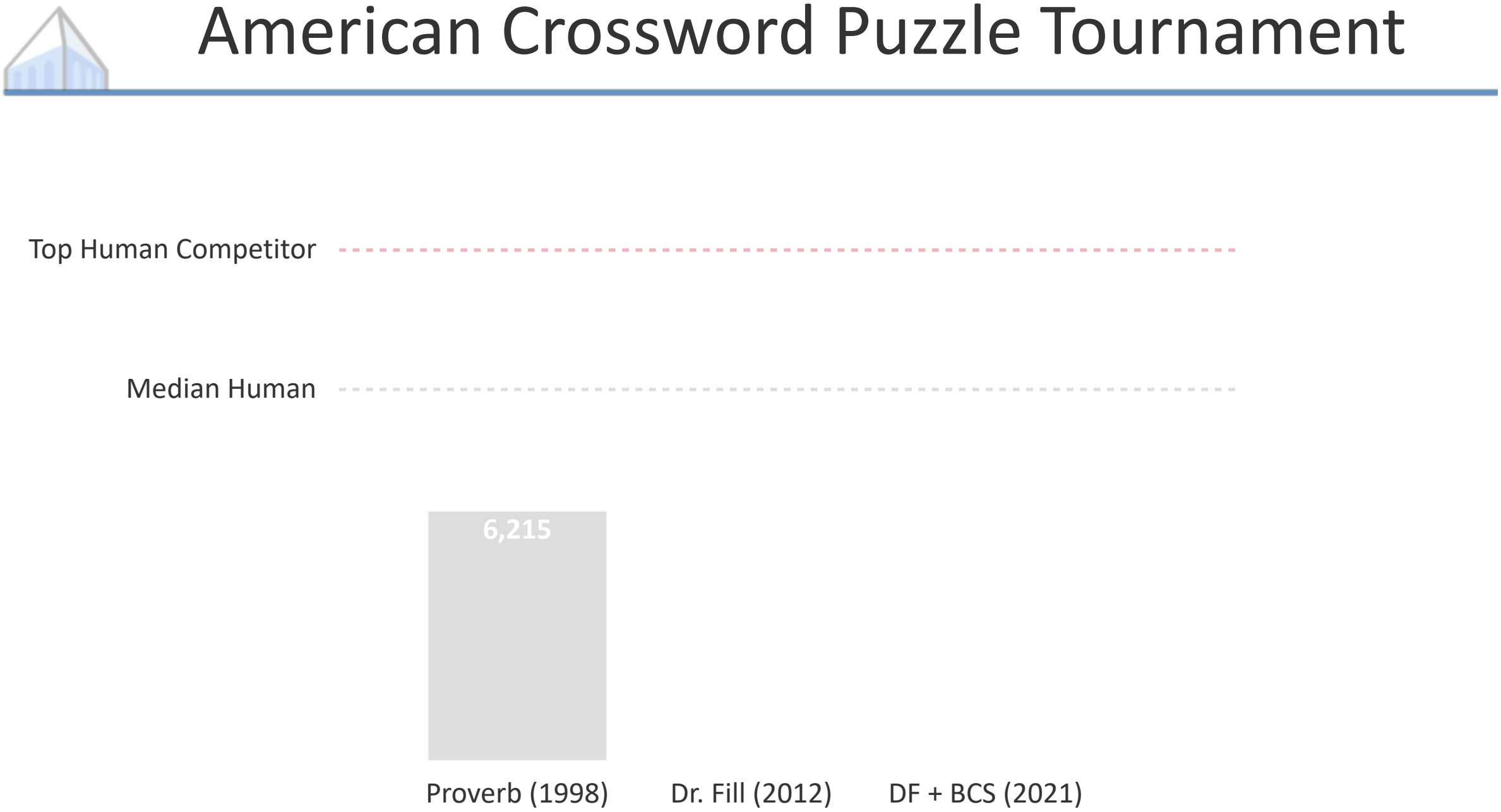
American Crossword Puzzle Tournament

Top Human Competitor

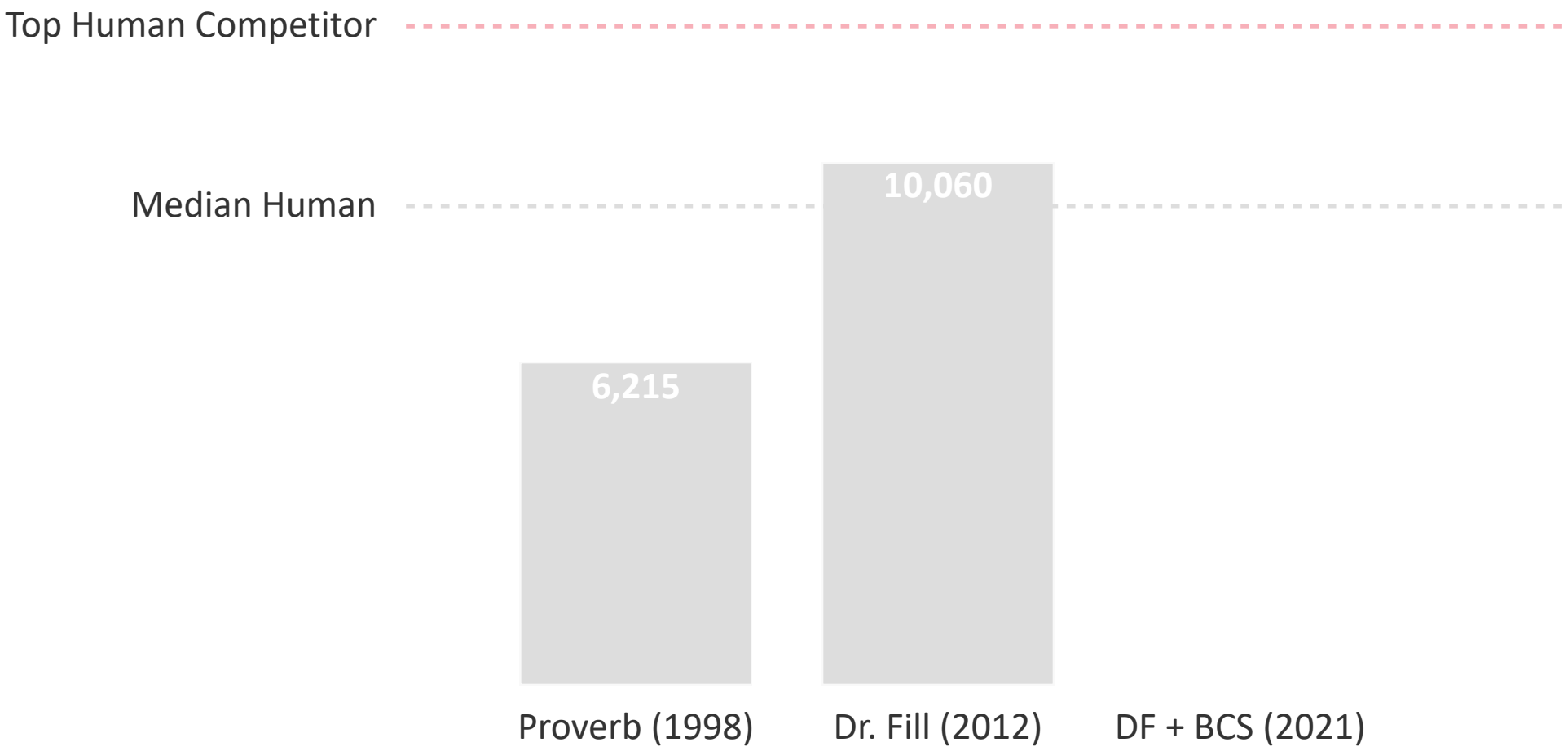
Median Human

Proverb (1998) Dr. Fill (2012) DF + BCS (2021)

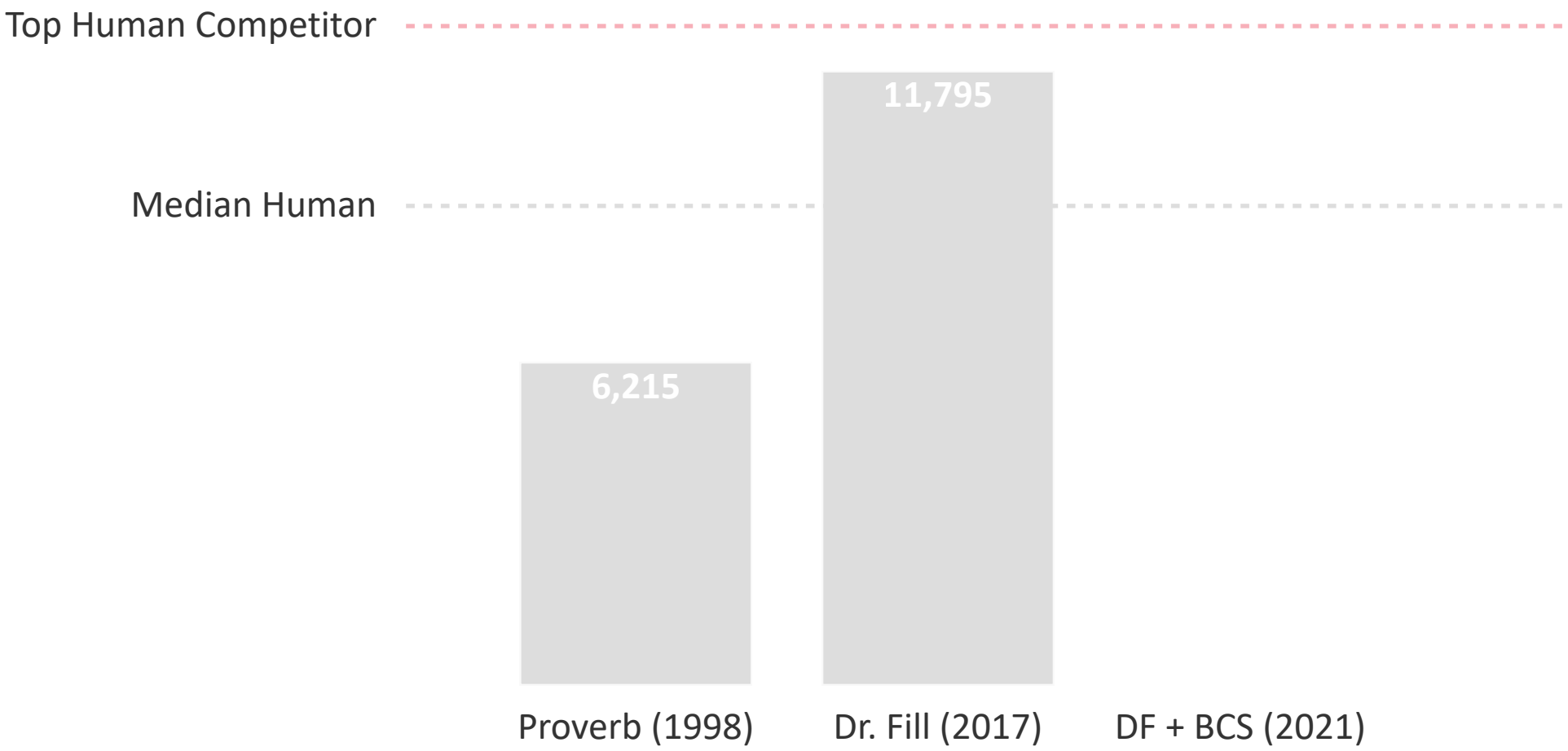
American Crossword Puzzle Tournament



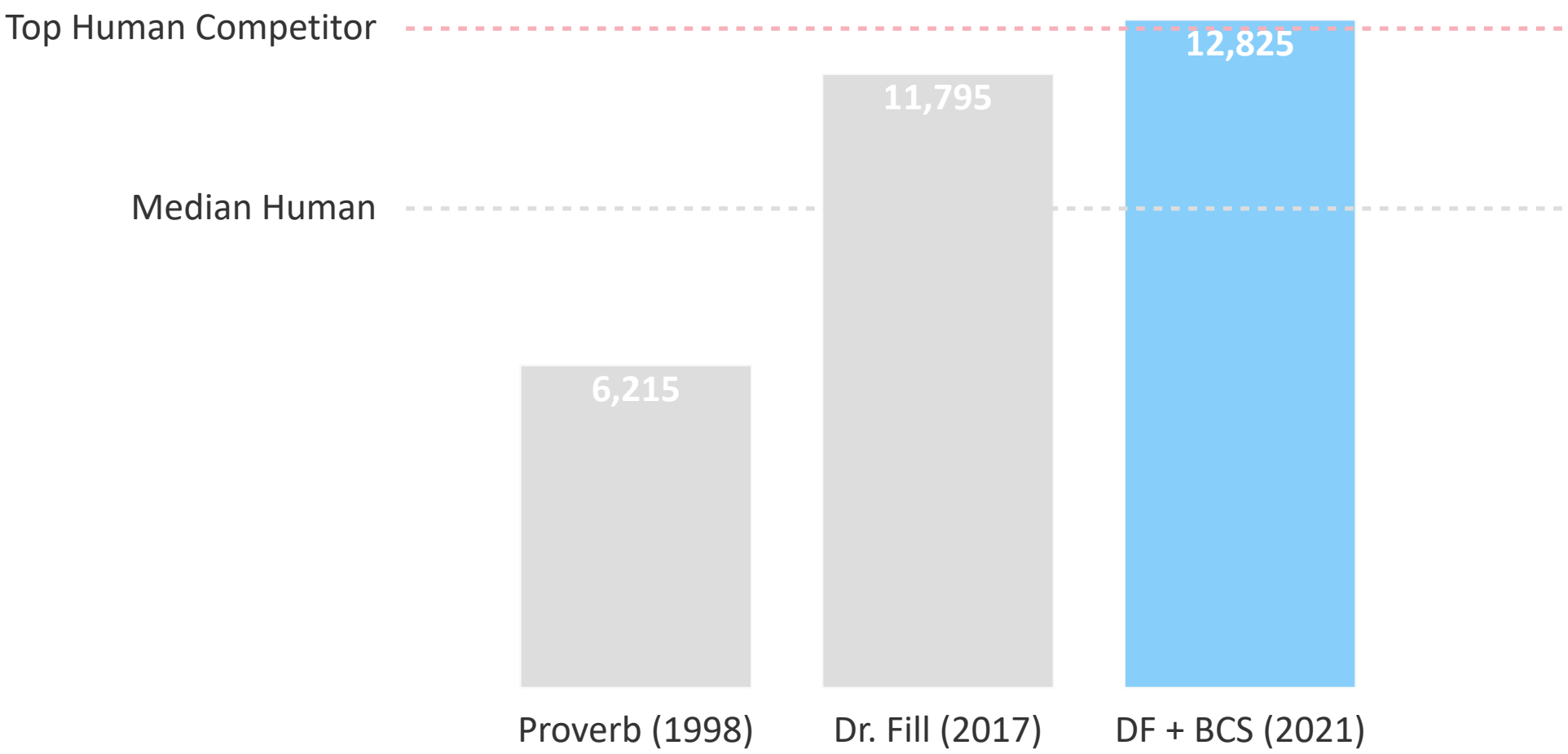
American Crossword Puzzle Tournament



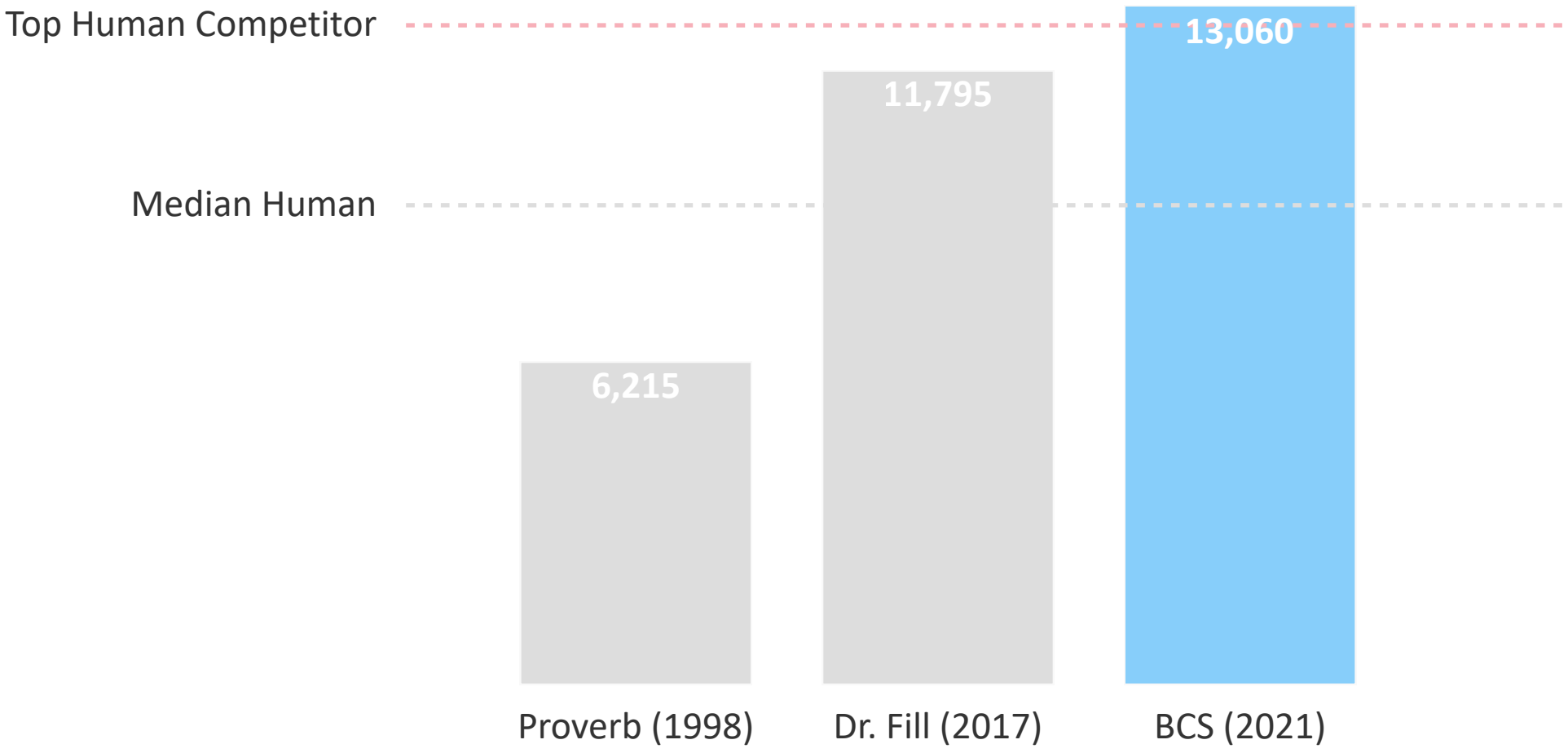
American Crossword Puzzle Tournament



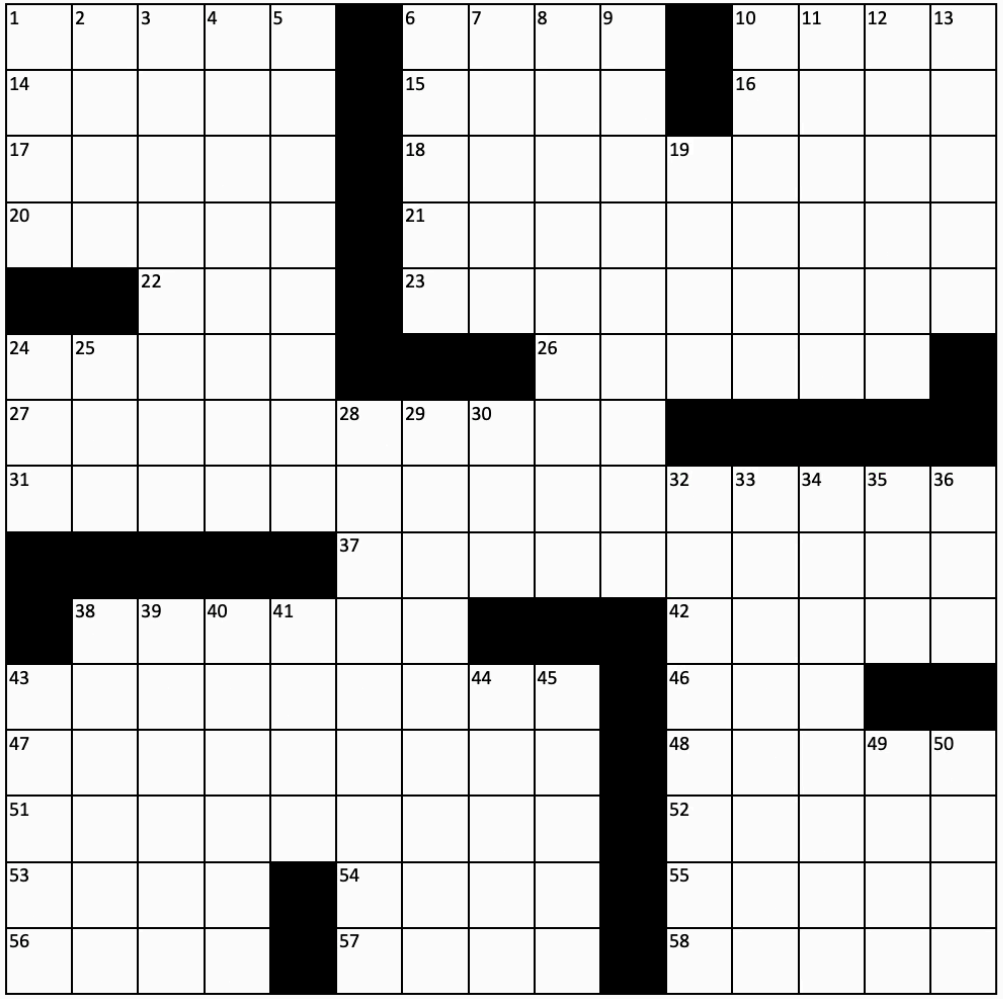
American Crossword Puzzle Tournament



American Crossword Puzzle Tournament



Demo



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

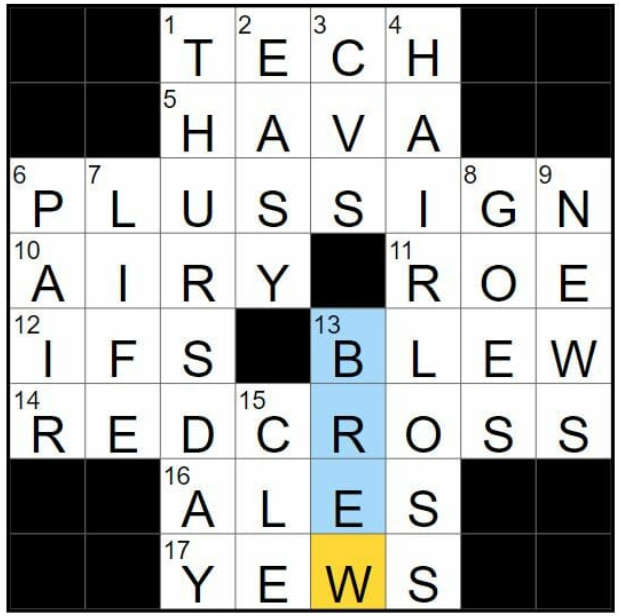
Down

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

The New York Times Crossword from Saturday, May 22, 2021. Constructed by Sid Sivakumar and Matthew Stock and edited by Will Shortz.

Overview

Crossword Solving



Constraint satisfaction, belief propagation, greedy search, iterative improvement

Detecting LLMs

Classification, feature selection, generalization and overfitting

Decision-Making in NL

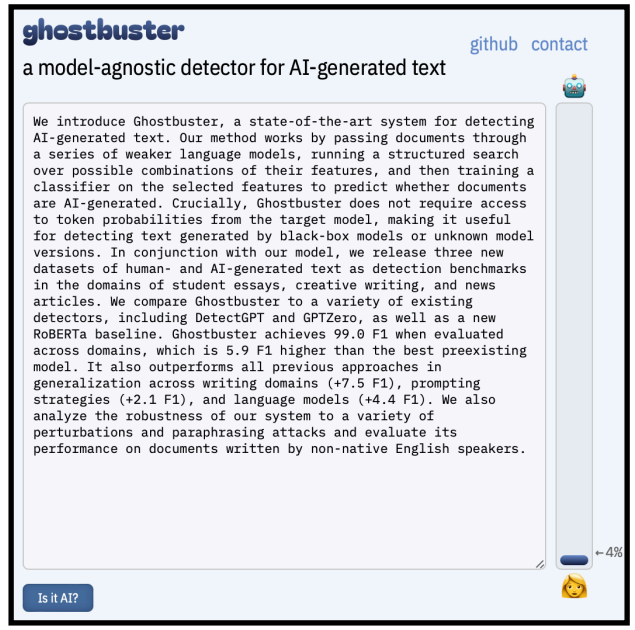
Constraint satisfaction, value of information, decision networks, particle filtering

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

FORBES > LEADERSHIP > CAREERS

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.

Follow

The New York Times

A.I. and Chatbots › | Google's RT-2 Robot | Smart Ways to Use Chatbots | Cha

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

With the rise of the popular new chatbot ChatGPT, col restructuring some courses and taking preventive mea

3.6K

The Washington Post

Sign in

Help Desk | Tech in Your Life | Future of Work | Your Data and Privacy | Inter

TECH IN YOUR LIFE

We tested a new ChatGPT-detector for teachers. It flagged an innocent student.

Five high school students helped our tech columnist test a ChatGPT detector coming from Turnitin to 2.1 million teachers. It missed enough to get someone in trouble.

Analysis by Geoffrey A. Fowler

Columnist | + Follow

Updated April 3, 2023 at 9:47 a.m. EDT | Published April 3, 2023 at 6:00 a.m. EDT

INSIDER

Subscribe

HOME > TECH

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

BACKCHANNEL | BUSINESS | MORE | SUBSCRIBE

RE JAN 30, 2023 1:21 PM

ls Making Universities agiarism

fessors can't decide whether the AI rch tool—or a cheating engine.

WSJ

TECH

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 Alyssa Lukpat Follow

Updated Jan. 6, 2023 at 12:55 pm ET

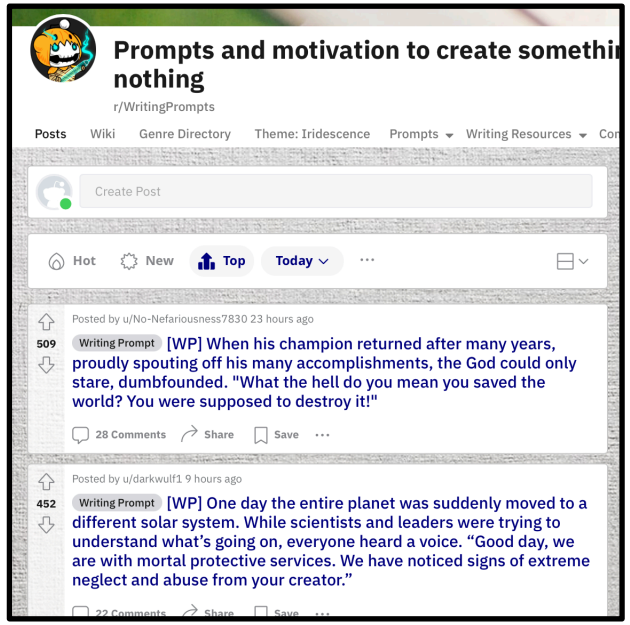
ChatGPT in Schools

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.	
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…	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…

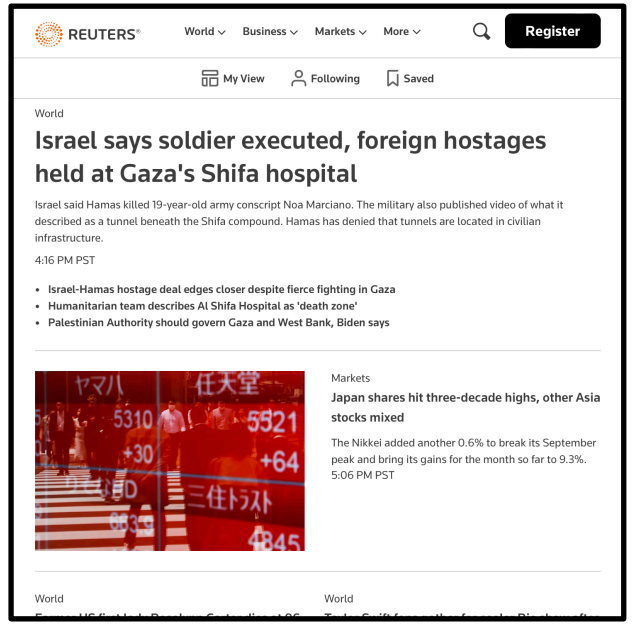
ChatGPT in Schools

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.	
Human Written Text	**ChatGPT Generated Text**
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…	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…

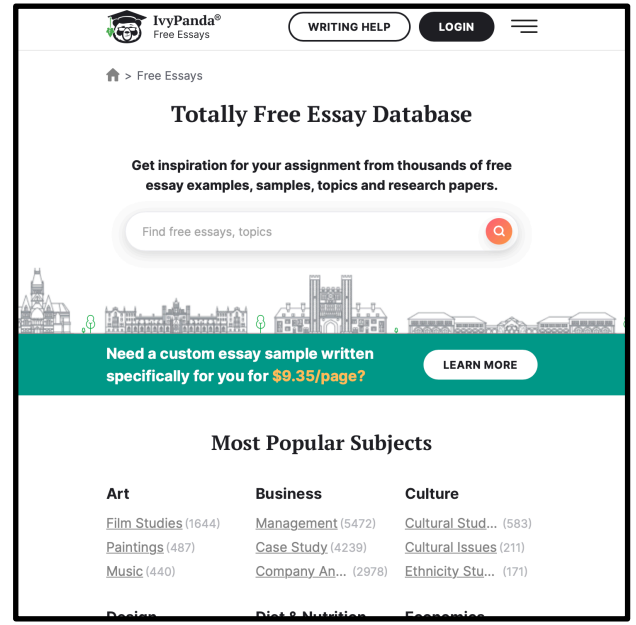
Data Sources



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

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

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

Bad underfitting (80 F1)

Two Ideas That Didn't Work

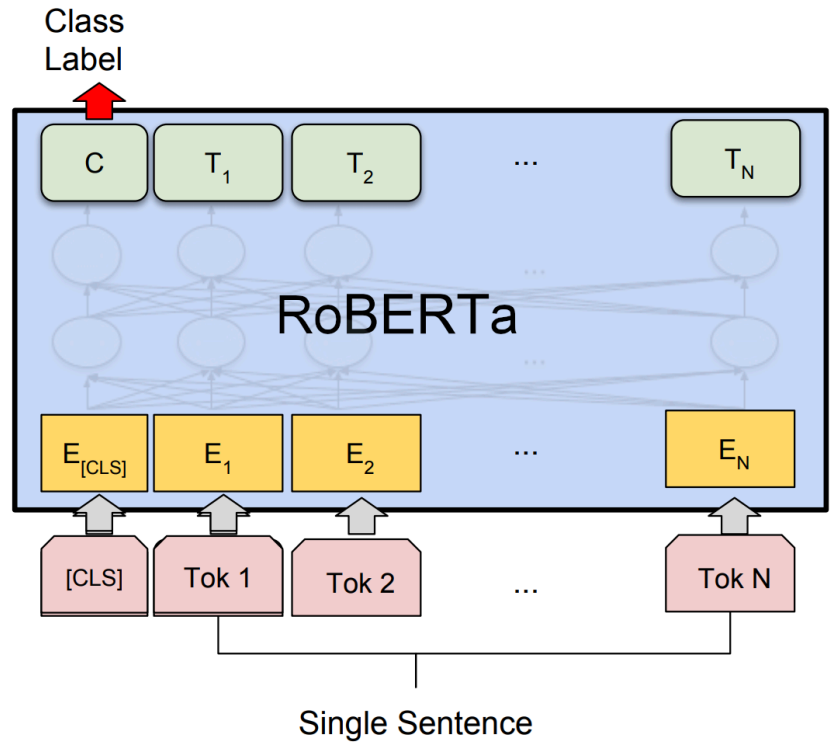
Idea #1: Perplexity Thresholding

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

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

Bad underfitting (80 F1)

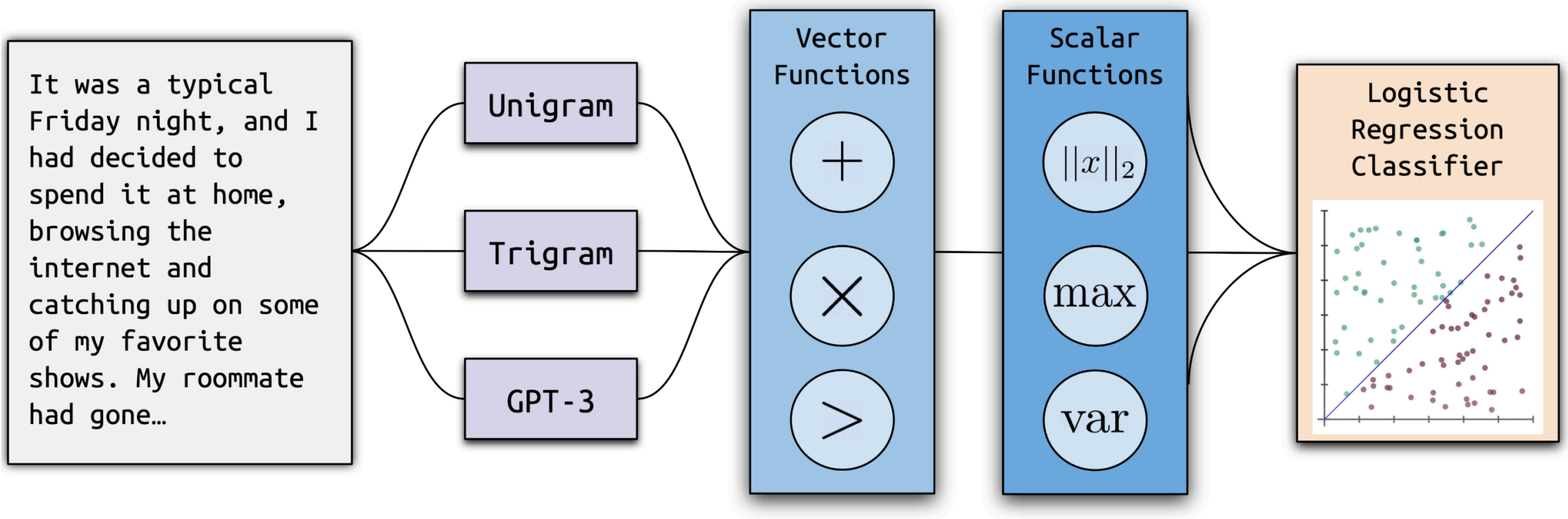
Idea #2: Fine-tuned Classifier



Bad overfitting

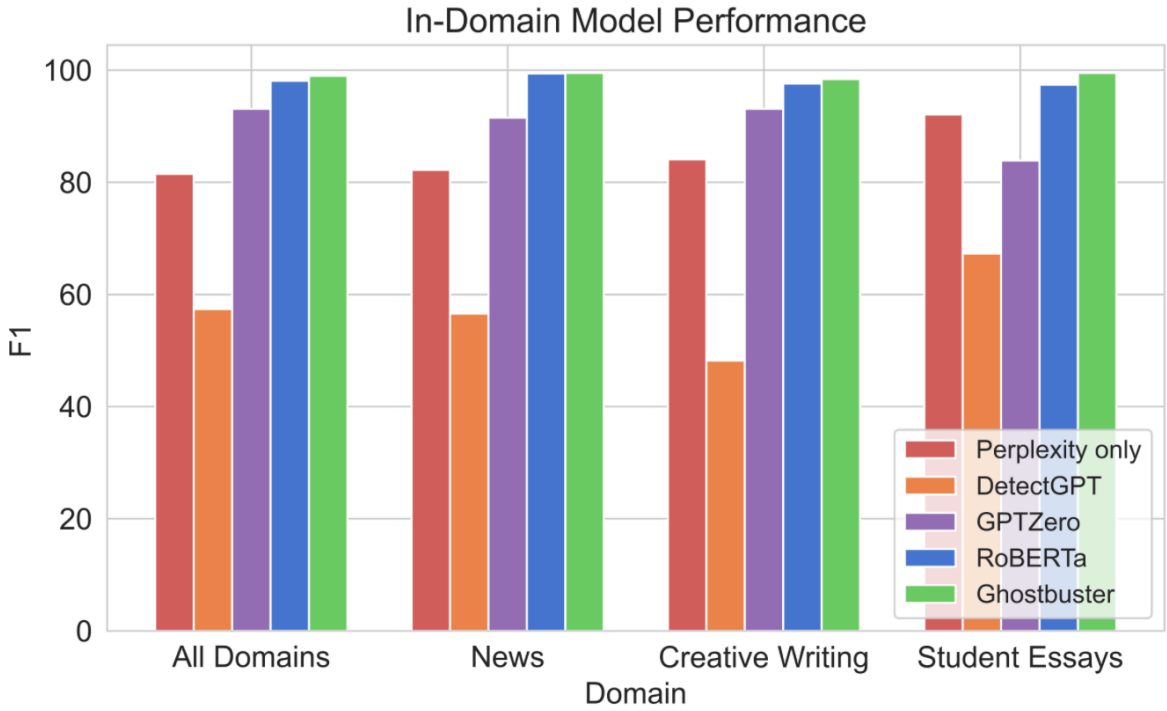
(98 F1 in-domain, but 85 F1 OOD)

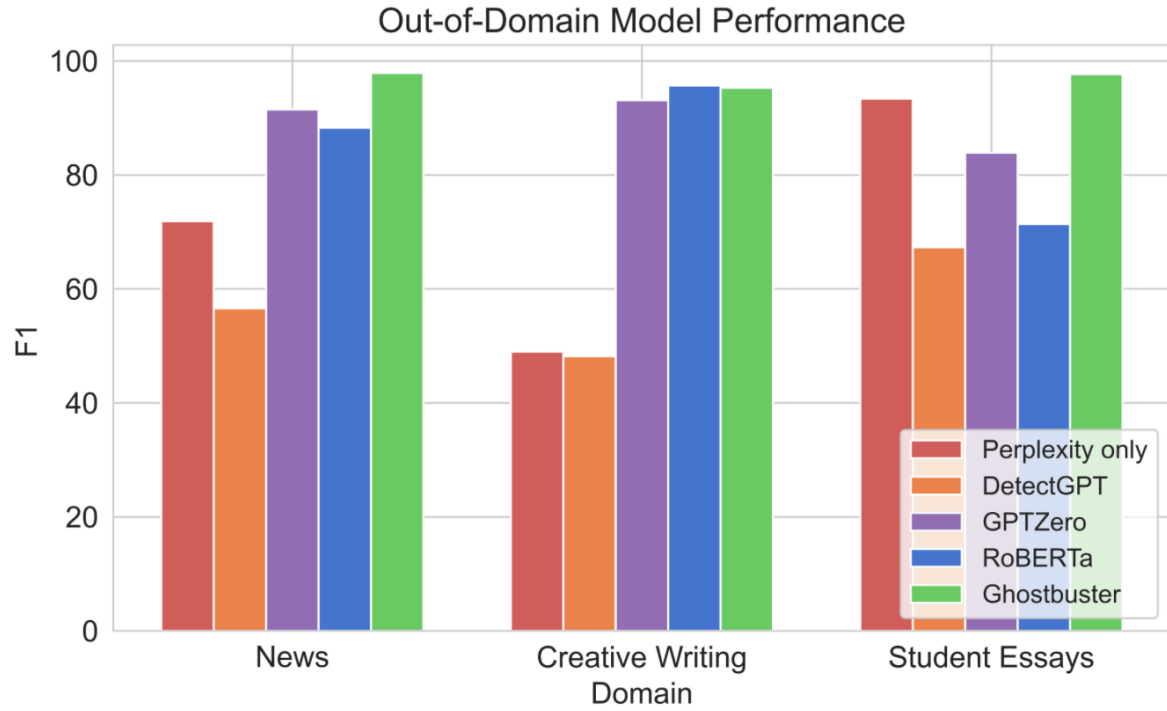
Our Approach: Ghostbuster



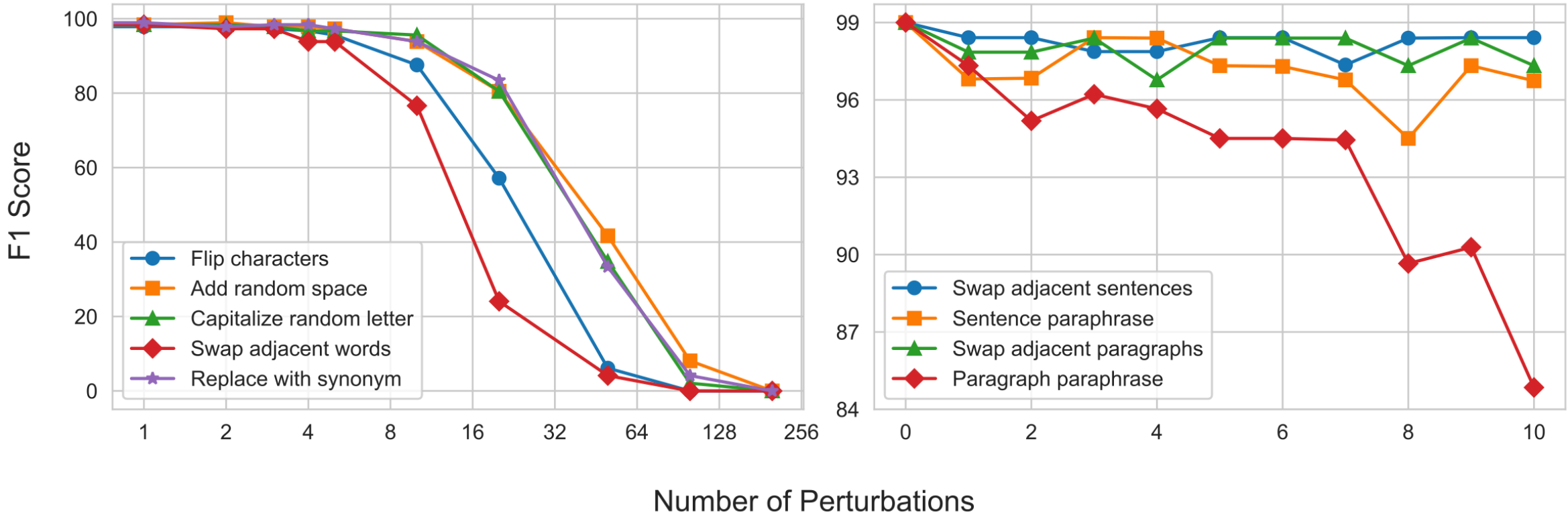
Stories	News	Essay
max(ada / unigram + ada) var(davinci / ada - davinci) avg(unigram > trigram > davinci) avg(unigram + trigram < davinci) var(unigram * davinci) avg-25(unigram - ada > davinci) avg-25(unigram - davinci / unigram)	avg-25(ada > davinci / trigram) var(ada - davinci) avg(trigram / ada > davinci) var(unigram > ada - davinci) avg(unigram > trigram + ada)	avg(ada > davinci / trigram) avg(davinci) var(trigram > davinci) avg(unigram / ada > davinci)

Results: Domain Generalization

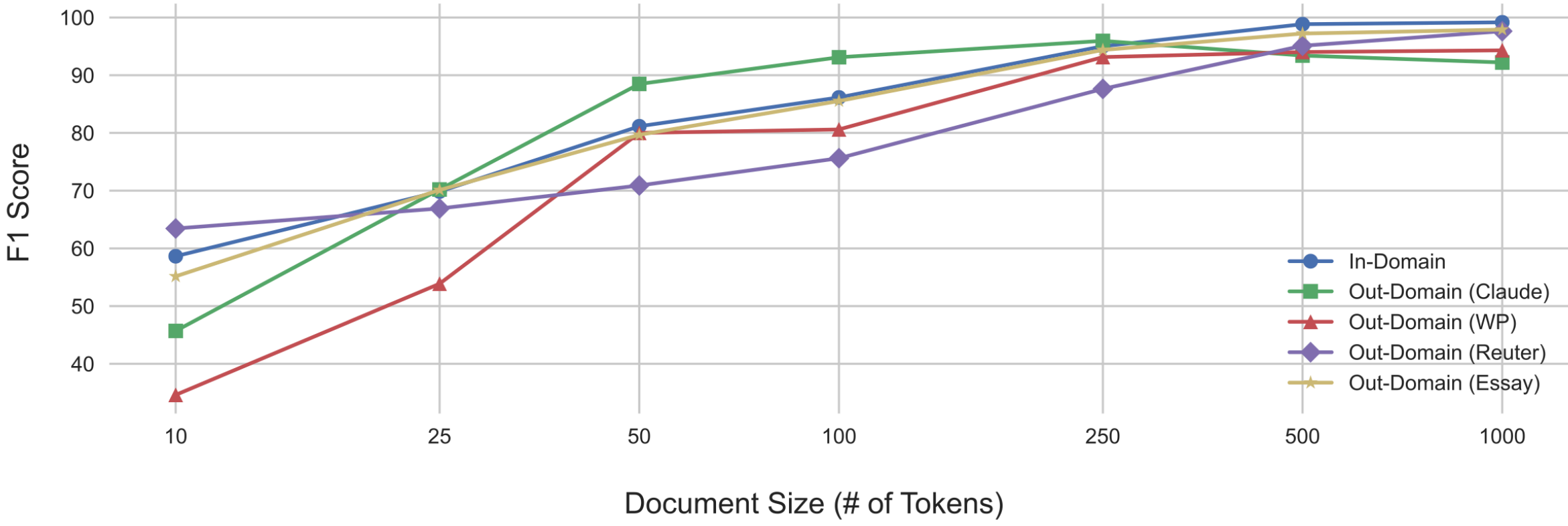




Results: Robustness to Perturbations



Results: Varying Document Lengths



Ethical Considerations

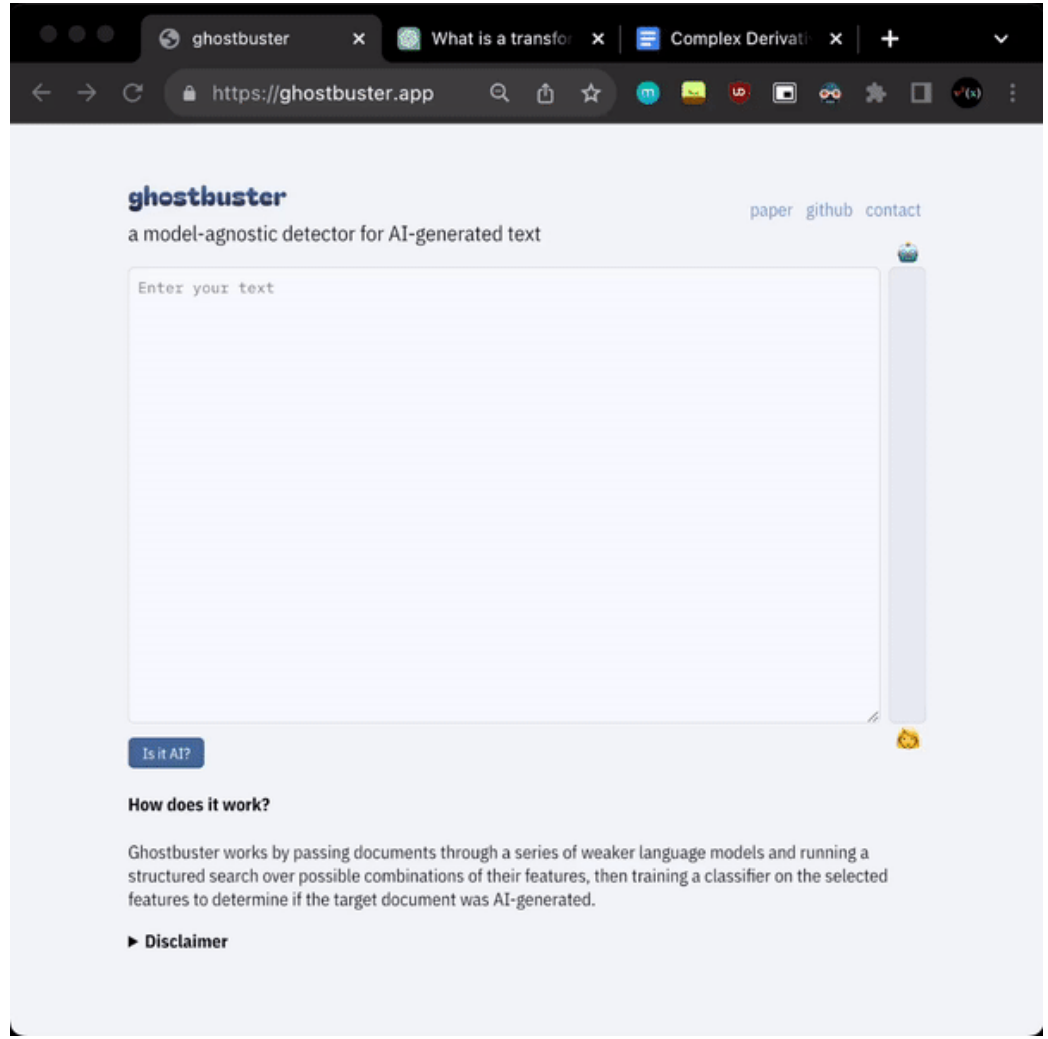
Design and Human-Computer Interaction, Language Processing, Machine Learning

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

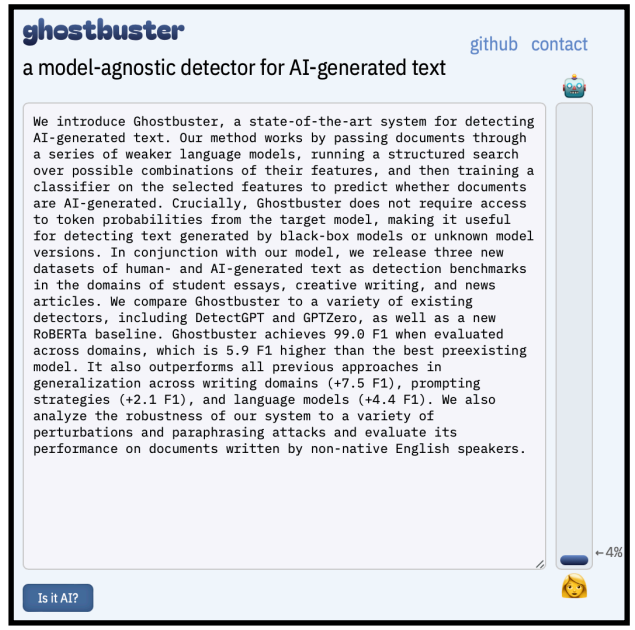


Overview

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

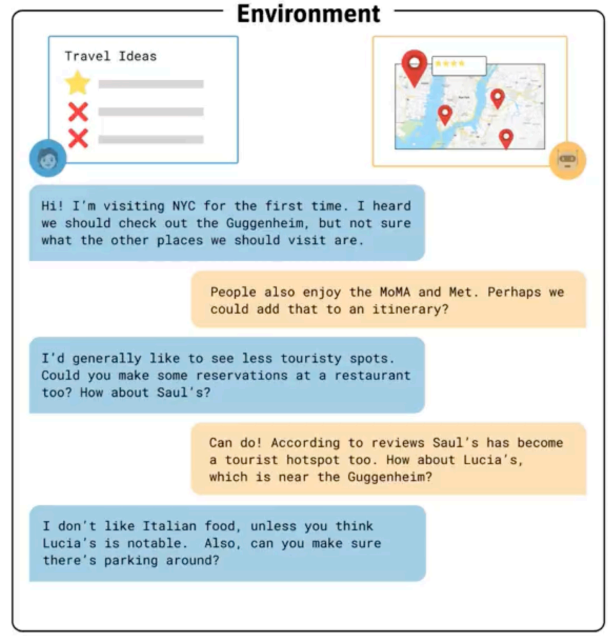
Crossword Solving

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

Assistance in the age of LLMs

Easy

My projector isn’t working. What’s a list of things that could be going wrong?

Assistance in the age of LLMs

Easy

My projector isn’t working. What’s a list of things that could be going wrong?

Hard

My projector stopped working. Can we narrow down what might be going wrong together, step-by-step?

Assistance in the age of LLMs

Easy

What's the current price of NVIDIA stock?

Assistance in the age of LLMs

Easy

What's the current price of NVIDIA stock?

Hard

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?

Assistance in the age of LLMs

Easy

Can you book me a flight to LA tomorrow?

Assistance in the age of LLMs

Easy

Can you book me a flight to LA tomorrow?

Hard

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

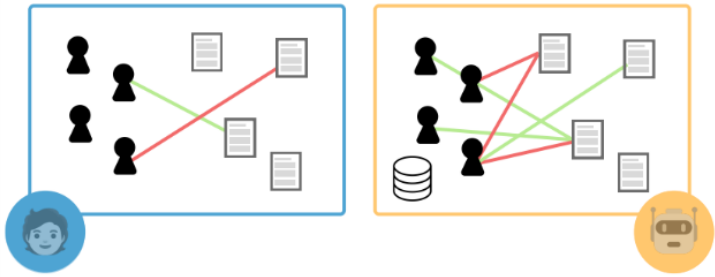
Optimization

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.

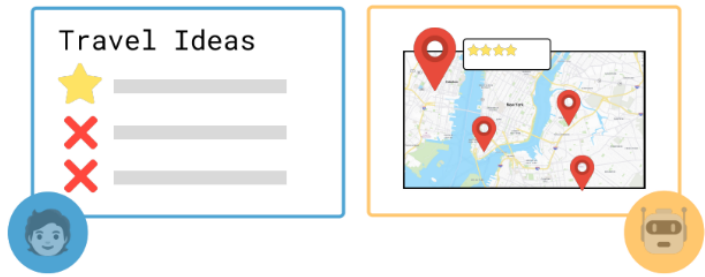
Planning

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?

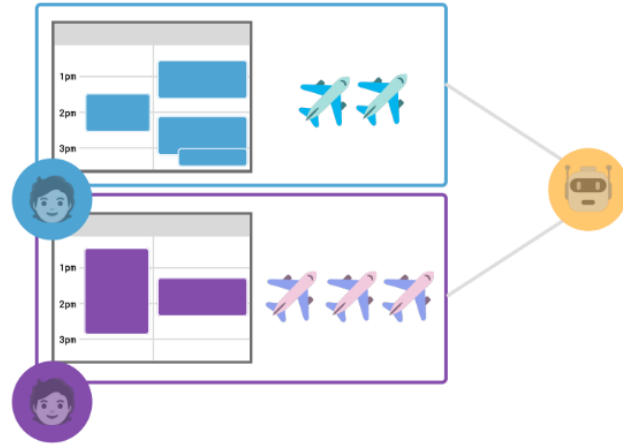
Mediation

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 | 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?

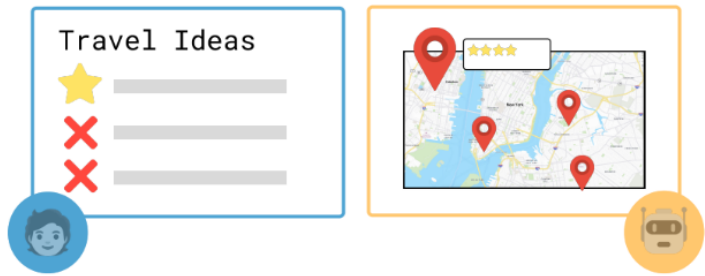
She's the only person who can review MT.

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.

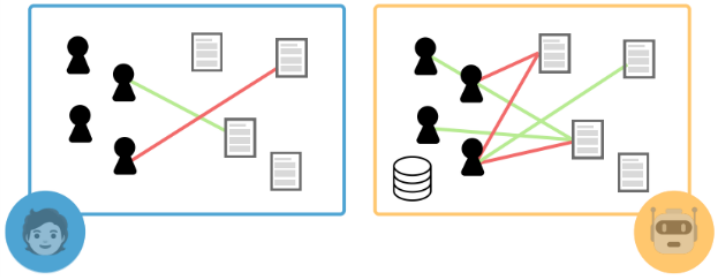
The DialOp Benchmark

Optimization

Conference Reviewer Matching

Assign suitable reviewers for conference papers

matching preferences | 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?

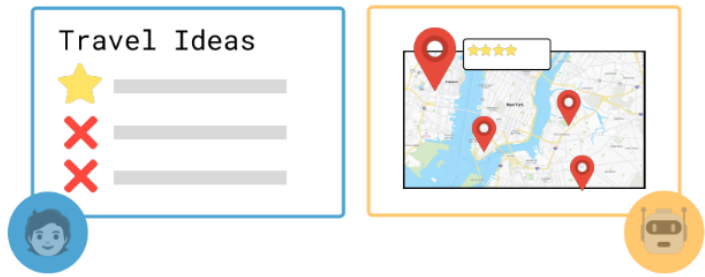
She's the only person who can review MT.

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

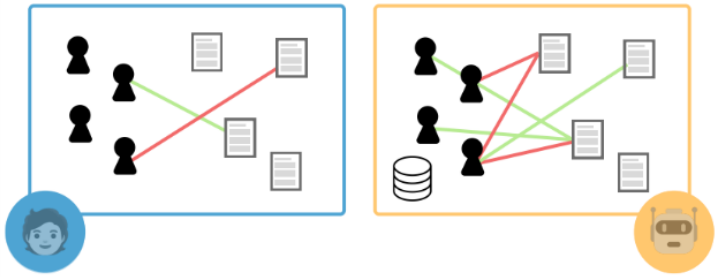
Optimization

Conference Reviewer Matching

Assign suitable reviewers for conference papers

matching preferences

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.

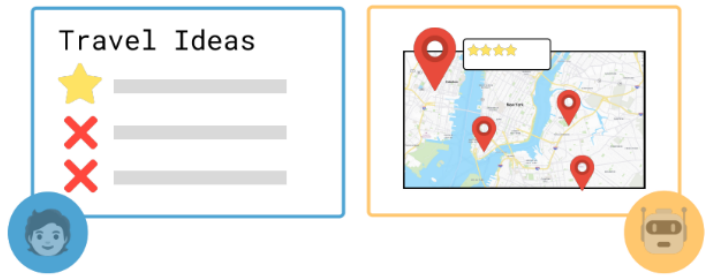
Planning

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?

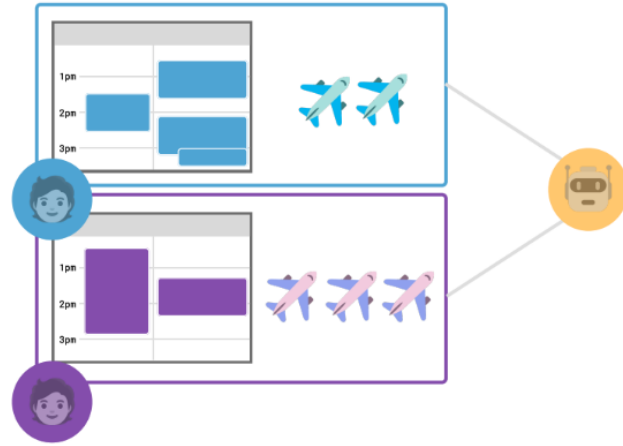
Mediation

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 | 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?

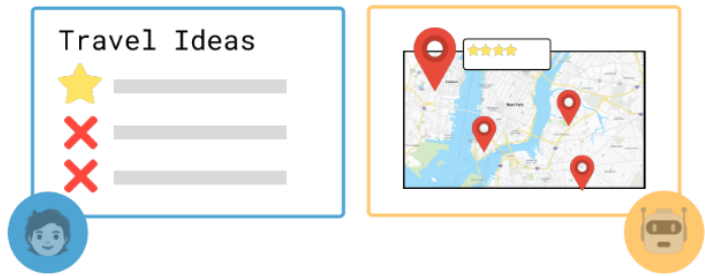
She's the only person who can review MT.

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

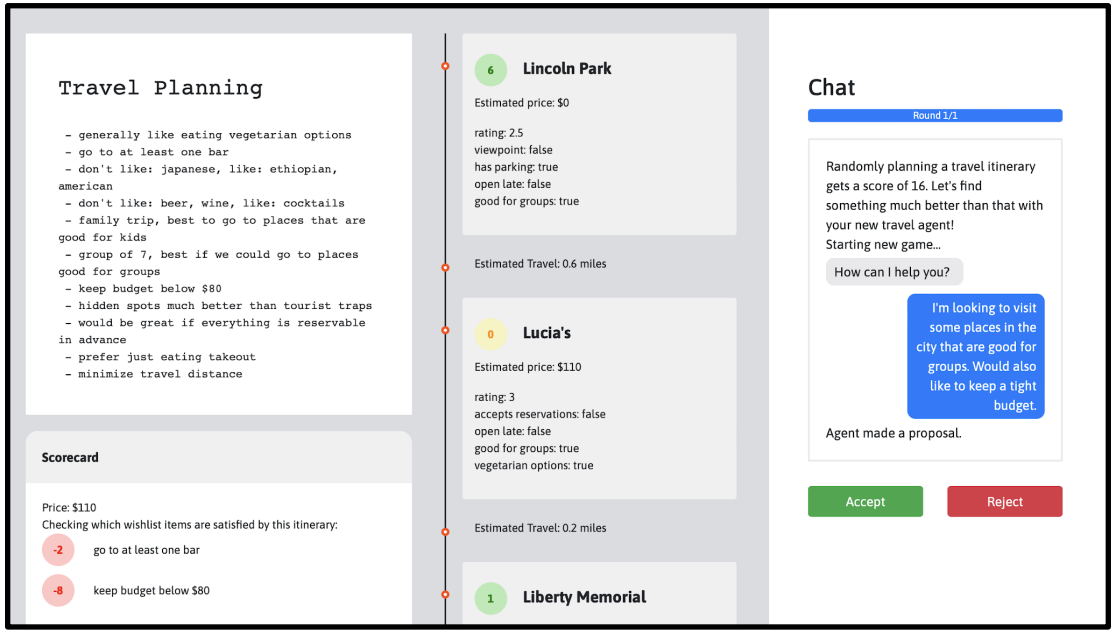
Book a set of flights that respect individual and group constraints

individual constraints and flights | global visibility

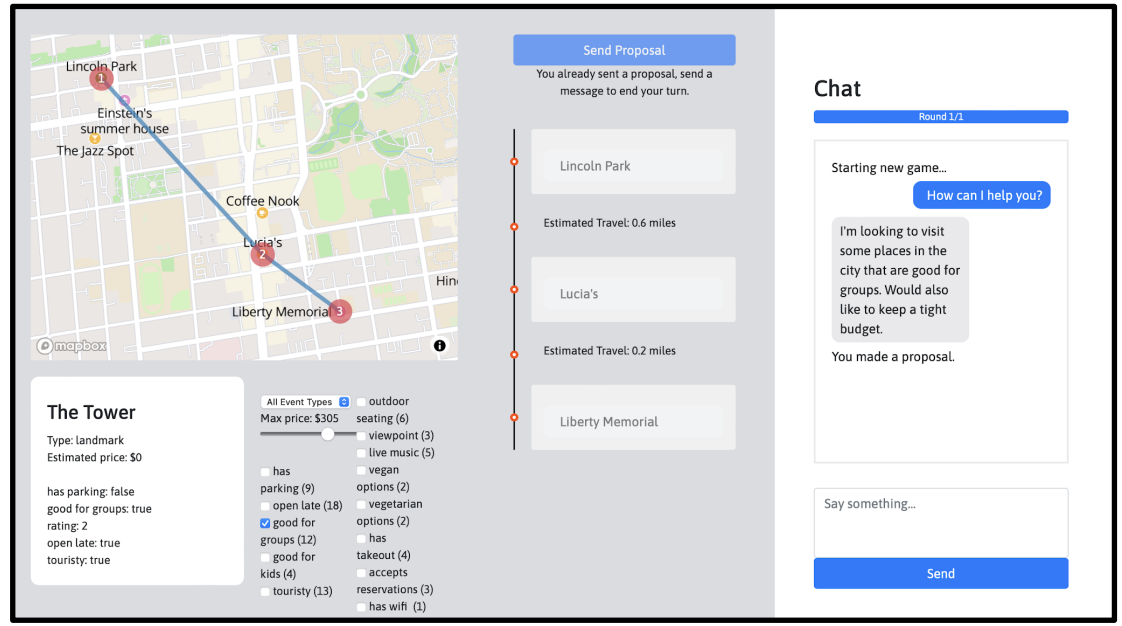
Could we all arrive before Friday 5pm?

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

User Interface: Planning



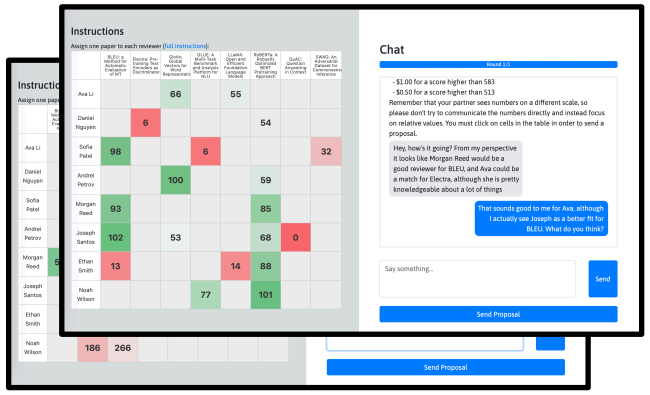
User View



Assistant View

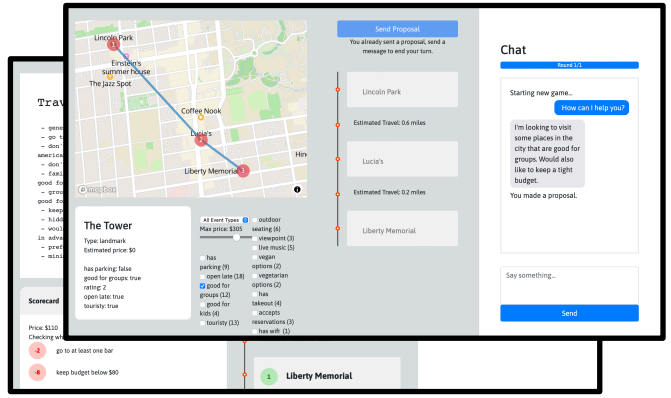
Human Data Collection

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



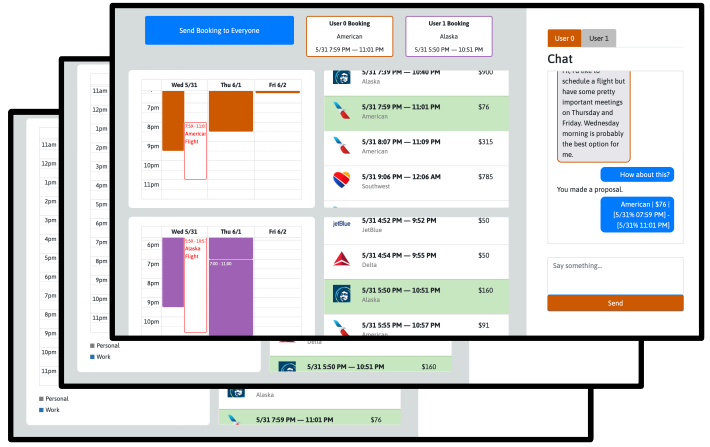
Optimization

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



Planning

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



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:
…
You (Agent): How does this set of flights look for you all? …
User 1: That’s great, thanks!
User 2: Let’s do it!
```

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:
…
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

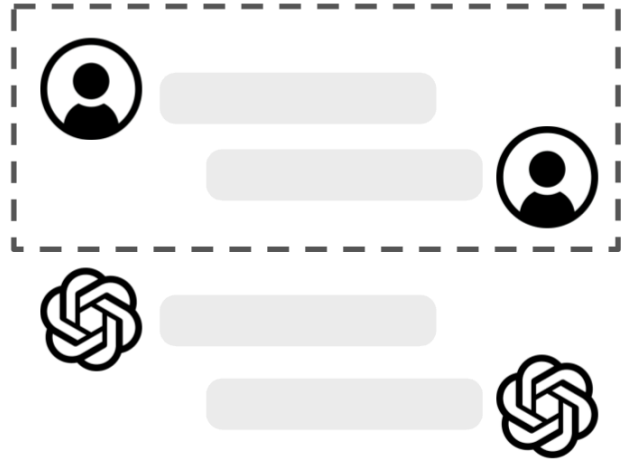
Evaluating Dialogue Assistants

Gold Standard:
Human-LM Evaluation

Assist real human users

Proxy:
LM-LM Self-Play

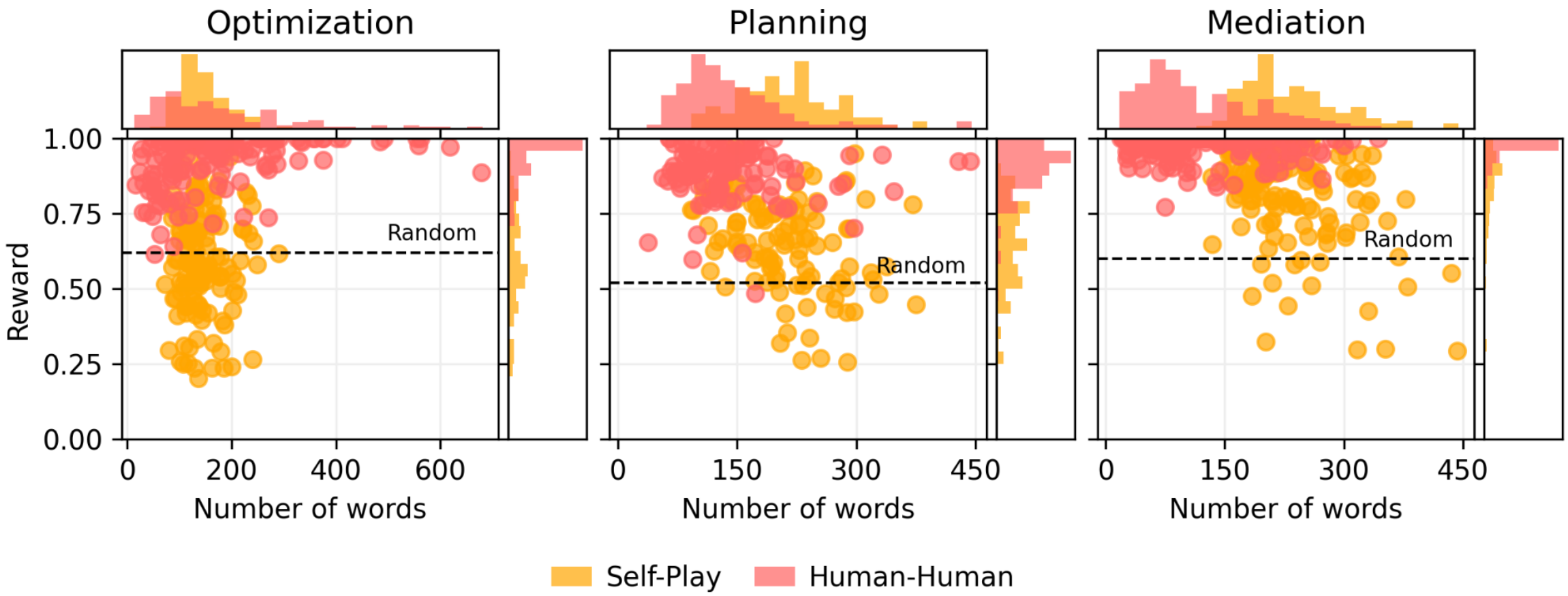
Assist an LM user simulator



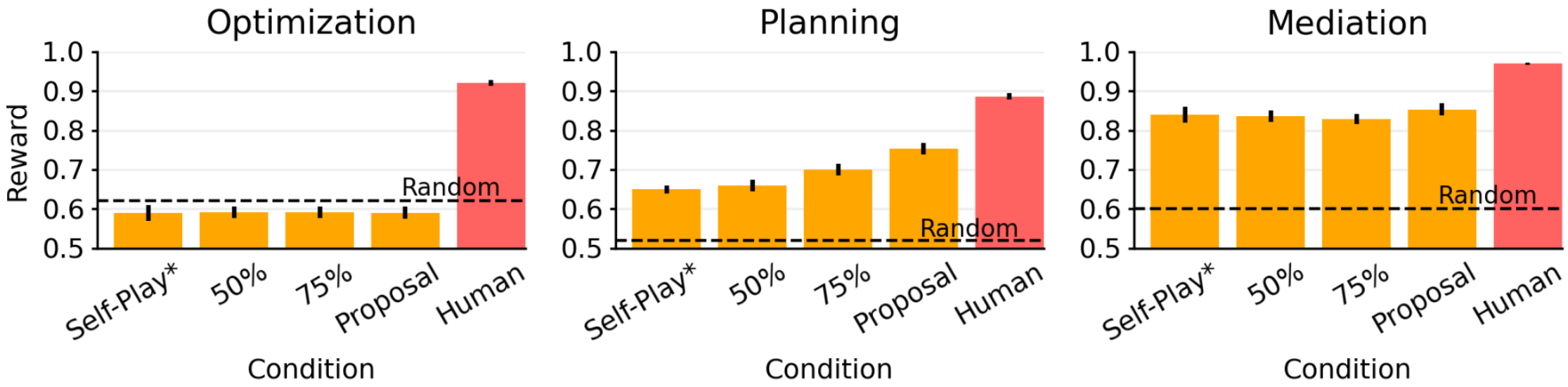
Better Proxy:
Prompted Self-Play

Continue assistance from
a human-human dialogue prefix

Results: Language Model Self-Play



Results: Prompted Self-Play



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)

(Real Dialogue)

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.

Model messages all preferences into the chat

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)

(Real Dialogue)

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.

Model messages all preferences into the chat

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)

(Real Dialogue)

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.

Model messages all preferences into the chat

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

Model queries for already revealed information

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)

(Real Dialogue)

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.

Model messages all preferences into the chat

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

Model queries for already revealed information

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)

(Real Dialogue)

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.

Model messages all preferences into the chat

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

Model queries for already revealed information

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)

(Real Dialogue)

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.

Model messages all preferences into the chat

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

Model queries for already revealed information

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)

(Real Dialogue)

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.

Model messages all preferences into the chat

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

Model queries for already revealed information

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

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

Model makes an initial weak proposal, which will then be rejected

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)

(Real Dialogue)

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.

Model messages all preferences into the chat

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

Model queries for already revealed information

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

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

Model makes an initial weak proposal, which will then be rejected

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)

(Real Dialogue)

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.

Model messages all preferences into the chat

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

Model queries for already revealed information

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

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

Model makes an initial weak proposal, which will then be rejected

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?

Dialogue continues and results in only slightly better outcome

Proposal Rejected (64% Match)

(Real Dialogue)

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

(Real Dialogue)

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

Human sends partial preferences into the chat

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

(Real Dialogue)

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

Human sends partial preferences into the chat

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

(Real Dialogue)

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

Human sends partial preferences into the chat

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

(Real Dialogue)

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

Human sends partial preferences into the chat

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

Human partner makes a partial proposal and provides reasons why it might be a good one

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

(Real Dialogue)

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

Human sends partial preferences into the chat

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

Human partner makes a partial proposal and provides reasons why it might be a good one

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

(Real Dialogue)

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

Human sends partial preferences into the chat

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

Human partner makes a partial proposal and provides reasons why it might be a good one

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

(Real Dialogue)

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

Human sends partial preferences into the chat

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

Human partner makes a partial proposal and provides reasons why it might be a good one

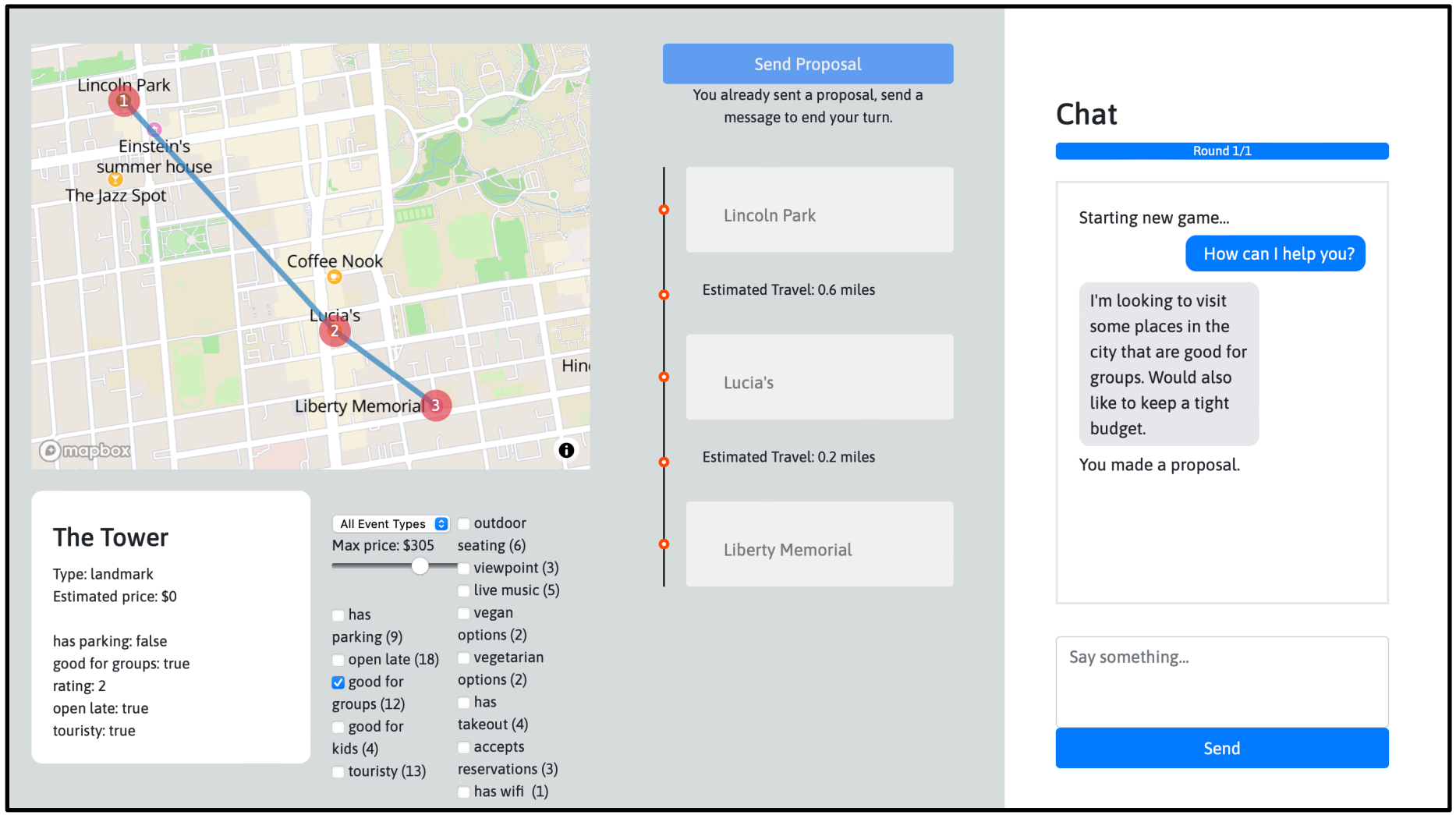
That looks pretty decent, could we tack on a visit to the Garden of Wonders? I REALLY WANT TO GO THERE

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

Dialogue continues as human players incrementally collaborate toward a final solution

(Real Dialogue)

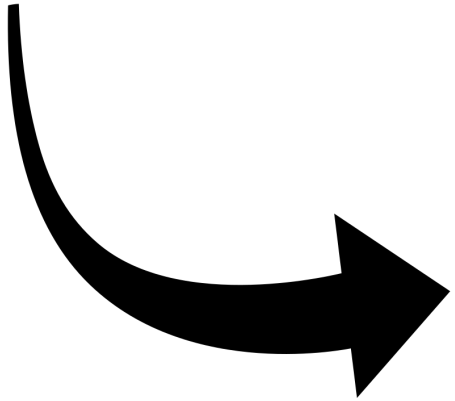
Demo (Time Permitting)



Thanks for listening!

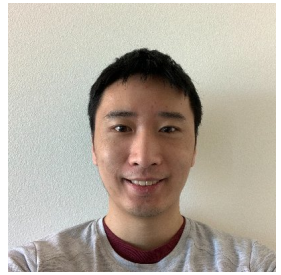
You can find more info about these projects on my website:
`https://people.eecs.berkeley.edu/~nicholas_tomlin/`

My awesome collaborators



Eric Wallace
UC Berkeley

Kevin Yang
UC Berkeley



Albert Xu
USC

Eshaan Pathak
You.com

Matt Ginsberg
Google

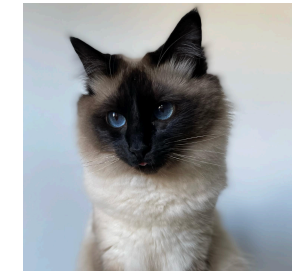
Dan Klein
UC Berkeley

Jessy Lin
UC Berkeley

Jacob Andreas
MIT

Jason Eisner
Johns Hopkins

Vivek Verma
UC Berkeley

Eve Fleisig
UC Berkeley

Coco
My Cat