

What can we do with massive generative abilities?
**Rethinking Discourse Processing and
Generation**

Jessy Li
The University of Texas at Austin

Discourse Structure

[1] California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax. [2] The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts. [3] The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure... [4] That leaves the state roughly \$2 billion to \$4 billion short. [5] A sales tax increase appears to be the fastest and easiest to raise funds in a hurry. [6] According to the state department of finance, a one-penny increase in the state's six-cent per dollar sales tax could raise \$3 billion.

A text is coherent if each newly introduced proposition is rhetorically connected to another piece of information, and that all anaphoric expressions can be resolved.

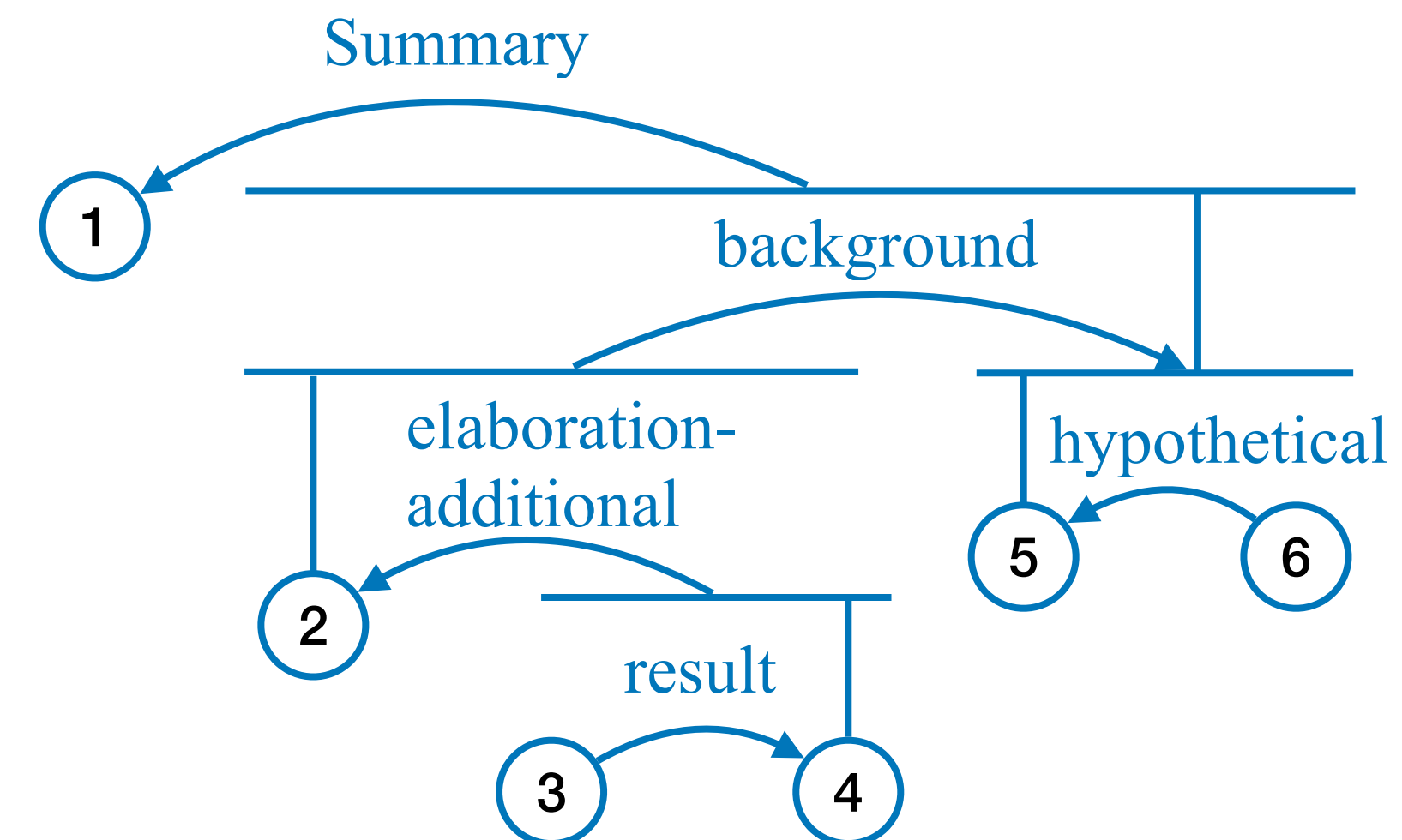
Asher and Lascarides, Logics of Conversation, 2003

- Rhetorical Structure Theory (RST)
- Segmented Discourse Representation Theory

Discourse Structure

[1] California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax. [2] The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts. [3] The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure... [4] That leaves the state roughly \$2 billion to \$4 billion short. [5] A sales tax increase appears to be the fastest and easiest to raise funds in a hurry. [6] According to the state department of finance, a one-penny increase in the state's six-cent per dollar sales tax could raise \$3 billion.

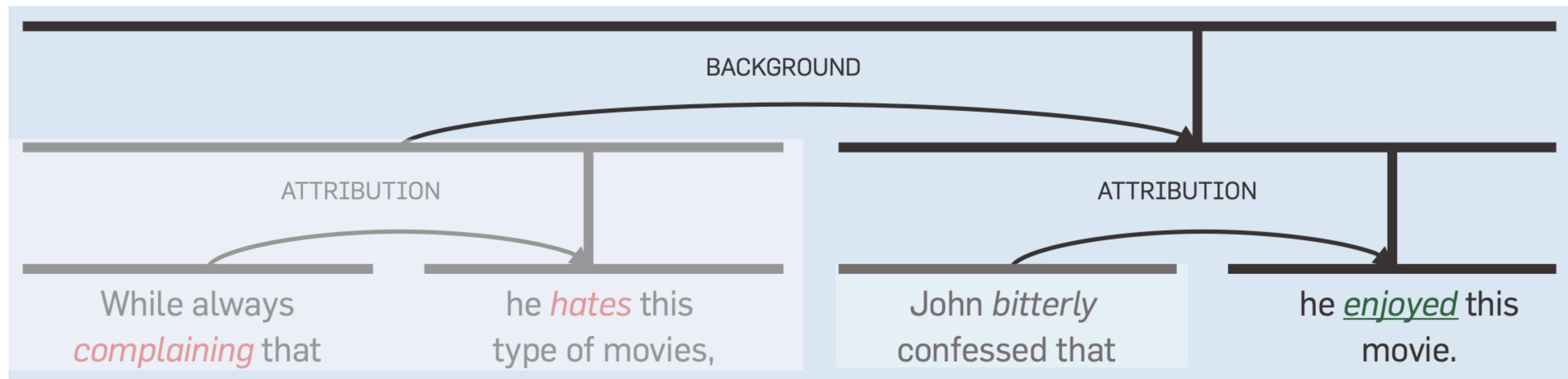
An RST tree:



Discourse is (was?) useful!

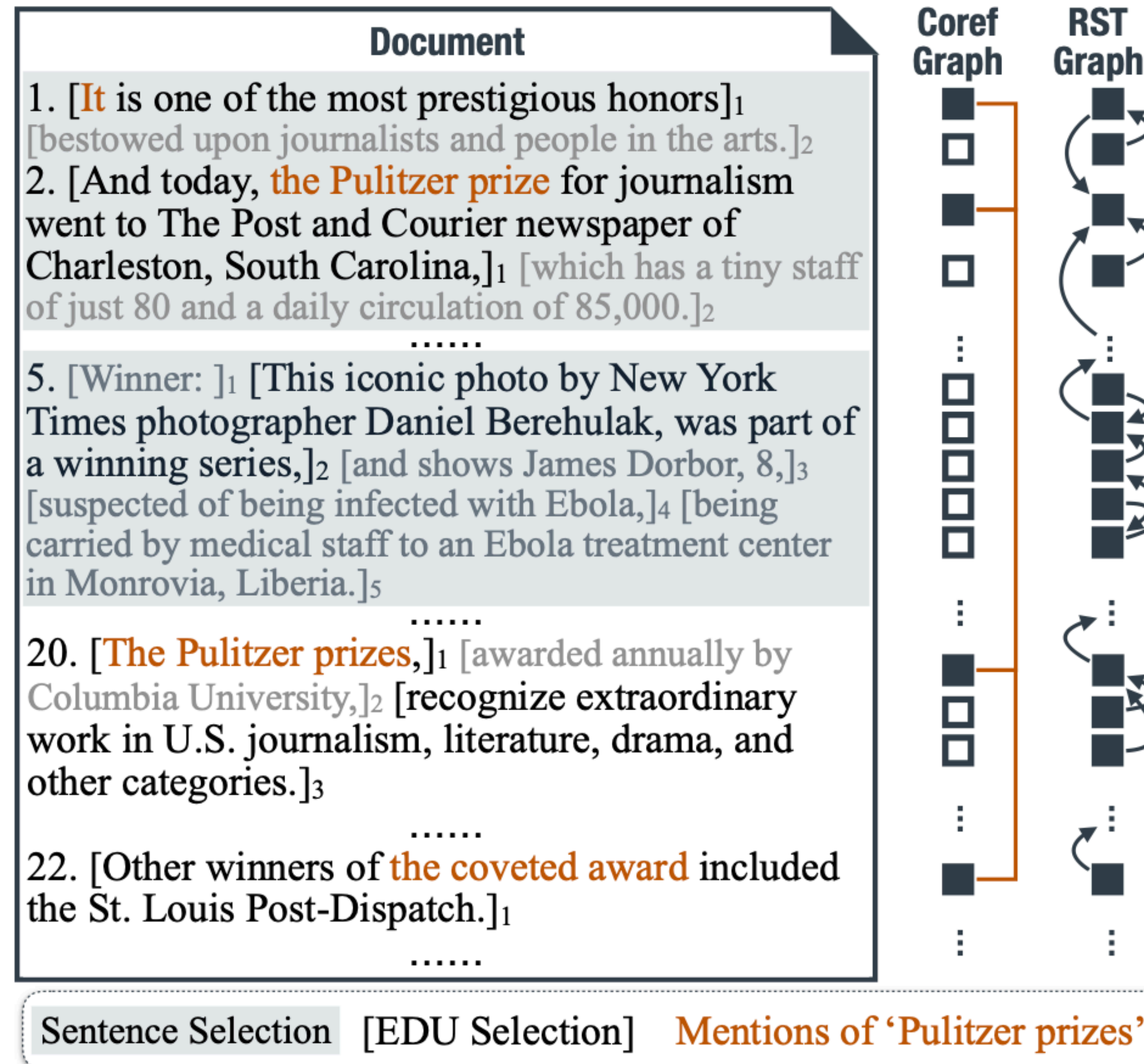
Sentiment analysis

While always complaining that he hates this type of movies, John bitterly confessed that he enjoyed this movie.



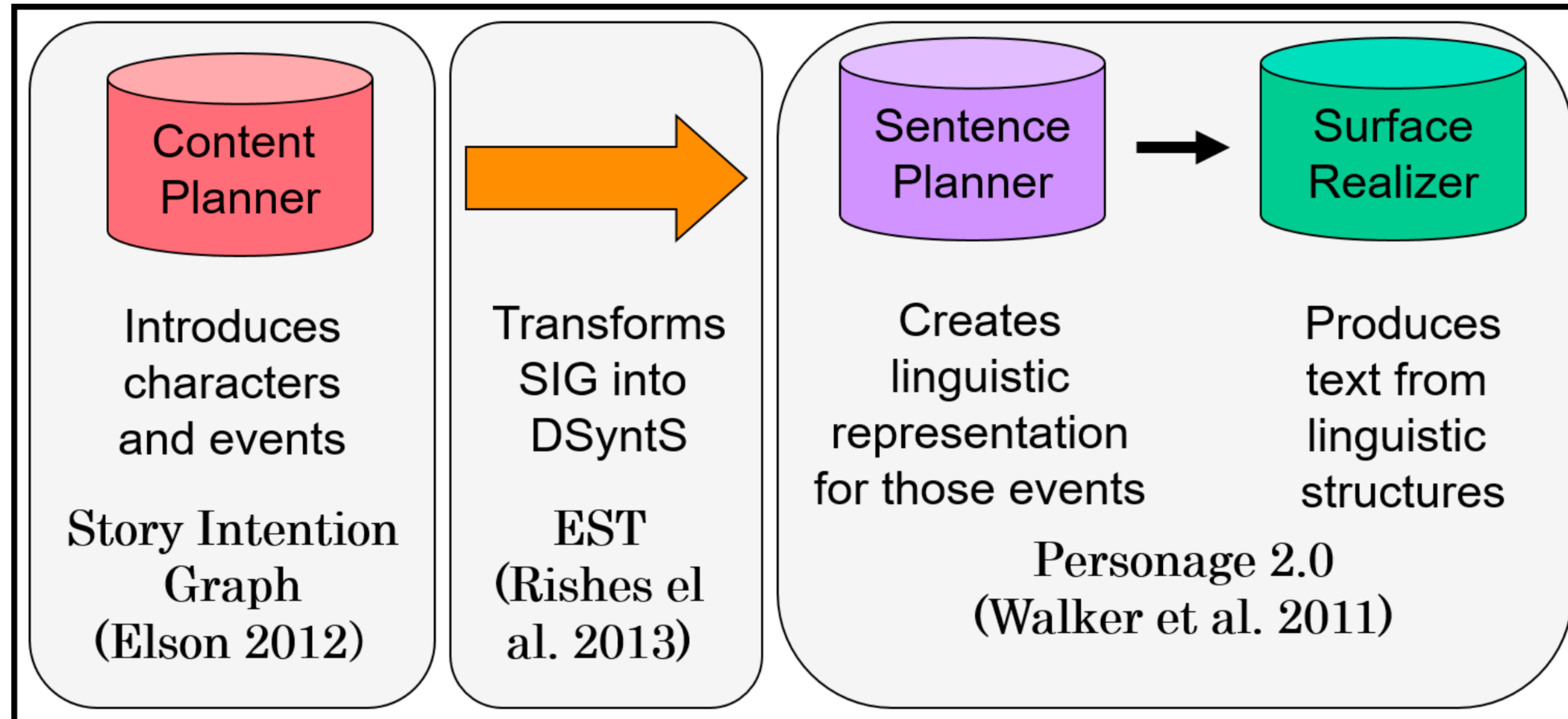
Discourse is (was?) useful!

Summarization



Discourse is (was?) useful!

Story generation



Discourse is (was?) useful!

- Text classification (Bhatia et al 2014, Nejat et al 2017, Ferracane et al 2017, Ji and Smith 2017)
- Machine translation and its evaluation (Marcu 2000, Li et al 2014, Joty et al 2017)
- Question answering and dialogue QA (Verberne et al 2007, Jansen et al 2014, Li et al 2021)
- Misinformation & fake news detection (Karimi and Tang 2019, Vargas et al 2022)

But what's the catch?

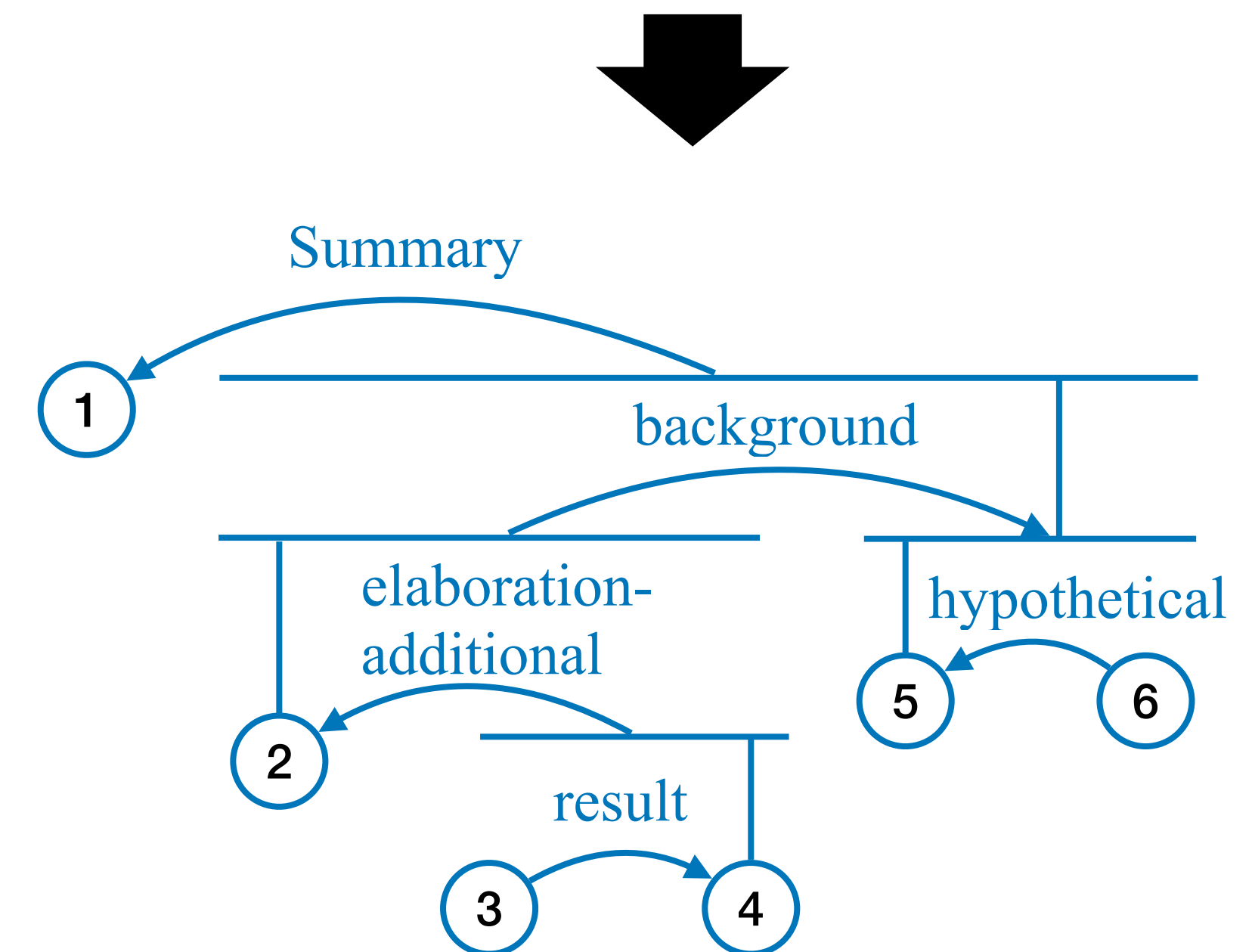
- “more than 10 years after the introduction of the PDTB, performance for English implicit discourse relation recognition has gone from 40.2 F-1 to 47.8, less than 8 percentage points” (Katherine Atwell et al, **Where Are We in Discourse Relation Recognition?, 2021**)
- “most of the heterogeneity in performance across RST parsers arises from differences in their evaluation procedures.” (Morey et al 2017)

Discourse annotation is hard to scale

e.g., RST...

- **Attribution:** attribution, attribution-negative
- **Background:** background, circumstance
- **Cause:** cause, result, consequence
- **Comparison:** comparison, preference, analogy, proportion
- **Condition:** condition, hypothetical, contingency, otherwise
- **Contrast:** contrast, concession, antithesis
- **Elaboration:** elaboration-additional, elaboration-general-specific, elaboration-part-whole, elaboration-process-step, elaboration-object-attribute, elaboration-set-member, example, definition
- **Enablement:** purpose, enablement
- **Evaluation:** evaluation, interpretation, conclusion, comment
- **Explanation:** evidence, explanation-argumentative, reason
- **Joint:** list, disjunction
- **Manner-Means:** manner, means
- **Topic-Comment:** problem-solution, question-answer, statement-response, topic-comment, comment-topic, rhetorical-question
- **Summary:** summary, restatement
- **Temporal:** temporal-before, temporal-after, temporal-same-time, sequence, inverted-sequence
- **Topic Change:** topic-shift, topic-drift

+ full document inference for each EDU



Now: Age of the LLMs

News Summarization and Evaluation in the Era of GPT-3

Tanya Goyal¹ Junyi Jessy Li² Greg Durrett¹

¹ Department of Computer Science ² Department of Linguistics

The University of Texas at Austin

tanyagoyal@utexas.edu

Summarization is (Almost) Dead

Xiao Pu*, Mingqi Gao*, Xiaojun Wan

Wangxuan Institute of Computer Technology, Peking University

puxiao@stu.pku.edu.cn

{gaomingqi, wanxiaojun}@pku.edu.cn

Classic benchmark domain of news summariza-
tion. First, we investigate how GPT-3 compares

assertions are implausible or unsupported.













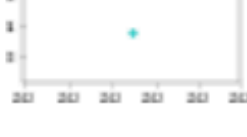


Now: Age of the LLMs

Story generation leaderboard, all results prior to 2022

Benchmarks

Add a Result

These leaderboards are used to track progress in Story Generation

Trend	Dataset	Best Model	Paper	Code	Compare
	Fandom dev	Everything here predates 2022...			See all
	Fandom test				See all
	TVMegaSite dev	(NN) Oracle plot + summary + oracle char. desc.			See all
	TVMegaSite test	(NN) Oracle plot + summary + oracle char. desc.			See all
	WritingPrompts	BART (TextBox 2.0)			See all

... and here come the LLMs

- Generated text is coherent. (Or, “at the level of an average college student essay”)
- SOTA performance on most tasks mentioned above (QA, summarization, dialogue generation...)
- How can we harness the generative ability of LLMs for models of discourse?
- Can we find a discourse model that are useful for a new range of tasks?

**Yes! Questions Under
Discussion**

This talk: discourse & generation

What is QUD and what does QUD structures look like?

- Discourse Comprehension: A Question Answering Framework to Represent Sentence Connections [Wei-Jen Ko et al, EMNLP'22]
- Discourse Analysis via Questions and Answers: Parsing Dependency Structures of Questions Under Discussion [Wei-Jen Ko, Yating Wu et al, ACL Findings'23]
- QUDeval: The Evaluation of Questions Under Discussion Discourse Parsing [Yating Wu et al., EMNLP'23]

How do we ask (generate) questions?

- Inquisitive Question Generation for High Level Text Comprehension [Wei-Jen Ko et al, EMNLP'20]
- Which questions should I answer? Saliency Prediction of Inquisitive Questions [Yating Wu, Ritika Mangla et al, EMNLP '24]

What can discourse do for generation?

- Elaborative Simplification as Implicit Questions Under Discussion [Yating Wu, William Sheffield et al, EMNLP'23]
- InfoLossQA: Characterizing and Recovering Information Loss in Text Simplification [Jan Trienes et al., ACL '24]

This talk: discourse & generation

What is QUD and what does QUD structures look like?

- Discourse Comprehension: A Question Answering Framework to Represent Sentence Connections [Wei-Jen Ko et al, EMNLP'22]
- Discourse Analysis via Questions and Answers: Parsing Dependency Structures of Questions Under Discussion [Wei-Jen Ko, Yating Wu et al, ACL Findings'23]
- QUDeval: The Evaluation of Questions Under Discussion Discourse Parsing [Yating Wu et al., EMNLP'23]

How do we ask (generate) questions?

- Inquisitive Question Generation for High Level Text Comprehension [Wei-Jen Ko et al, EMNLP'20]
- Which questions should I answer? Saliency Prediction of Inquisitive Questions [Yating Wu, Ritika Mangla et al, EMNLP '24]

What can discourse do for generation?

- Elaborative Simplification as Implicit Questions Under Discussion [Yating Wu, William Sheffield et al, EMNLP'23]
- InfoLossQA: Characterizing and Recovering Information Loss in Text Simplification [Jan Trienes et al., ACL '24]

Let's take a step back...

... and think about how **YOU** read...

[1] California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax. [2] The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts. [3] The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure... [4] That leaves the state roughly \$2 billion to \$4 billion short. [5] A sales tax increase appears to be the fastest and easiest to raise funds in a hurry. [6] According to the state department of finance, a one-penny increase in the state's six-cent per dollar sales tax could raise \$3 billion.



Let's take a step back...

... and think about how **YOU** read...

[1] California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax. [2] The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts. [3] The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure... [4] That leaves the state roughly \$2 billion to \$4 billion short. [5] A sales tax increase appears to be the fastest and easiest to raise funds in a hurry. [6] According to the state department of finance, a one-penny increase in the state's six-cent per dollar sales tax could raise \$3 billion.



What is prompting the California legislators to seek alternative forms of relief?

Let's take a step back...

... and think about how **YOU** read...

[1] California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax. [2] The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts. [3] The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure... [4] That leaves the state roughly \$2 billion to \$4 billion short. [5] A sales tax increase appears to be the fastest and easiest to raise funds in a hurry. [6] According to the state department of finance, a one-penny increase in the state's six-cent per dollar sales tax could raise \$3 billion.



What is prompting the California legislators to seek alternative forms of relief?

Answered in sentence 2

Let's take a step back...

... and think about how **YOU** read...

[1] California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax. [2] The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts. [3] The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure... [4] That leaves the state roughly \$2 billion to \$4 billion short. [5] A sales tax increase appears to be the fastest and easiest to raise funds in a hurry. [6] According to the state department of finance, a one-penny increase in the state's six-cent per dollar sales tax could raise \$3 billion.



What would be the advantage of the proposed tax?

Let's take a step back...

... and think about how **YOU** read...

[1] California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax. [2] The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts. [3] The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure... [4] That leaves the state roughly \$2 billion to \$4 billion short. [5] A sales tax increase appears to be the fastest and easiest to raise funds in a hurry. [6] According to the state department of finance, a one-penny increase in the state's six-cent per dollar sales tax could raise \$3 billion.



What would be the advantage of the proposed tax?

Answered in sentence 5

Questions Under Discussion (QUD)

A well-known linguistic framework in discourse & pragmatics

Each sentence in discourse addresses a (often implicit) QUD either by answering it, or by bringing up another question that can help answering that QUD. The linguistic form and the interpretation of a sentence, in turn, may depend on the QUD it addresses.

Benz and Jasinskaja, Questions under discussion: From sentence to discourse. 2017

Velleman and Beaver. Question-based models of information structure. 2016.

Roberts, Information structure: Towards an integrated formal theory of pragmatics. 2012

Roberts. Context in dynamic interpretation. 2004

Jonathan Ginzburg. Dynamics and the semantics of dialogue. 1996

Jan Van Kuppevelt. Discourse structure, topicality and questioning. 1995

Two ways to think about questions...

Way #1: QUD recovery

S₁: California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax.

S₂: The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts.

Question generation:
how does this sentence elaborate on prior context?

answers →

What is prompting the California legislators to seek alternative forms of relief?

Question linking:
which prior sentence is the anchor?

Two ways to think about questions...

Way #1: QUD recovery

Question generation:
how does this sentence elaborate on prior context?

S1: California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax.

S2: The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts.

answers → What is prompting the California legislators to seek alternative forms of relief?

S3: The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure...

answers → What type of relief is being considered?

S4: That leaves the state roughly \$2 billion to \$4 billion short.

answers → What would the shortfall be?

S5: A sales tax increase appears to be the fastest and easiest to raise funds in a hurry.

answers → What would be the advantage of the proposed tax?

Question linking:
which prior sentence is the anchor?

Introducing the **DCQA** dataset

Discourse **C**omprehension by **Q**uestion **A**nswering

- Ask about the focus of the answer sentence
- Ask open-ended questions
- Not use information first introduced in the new sentence
- Question-answer pairs are compatible

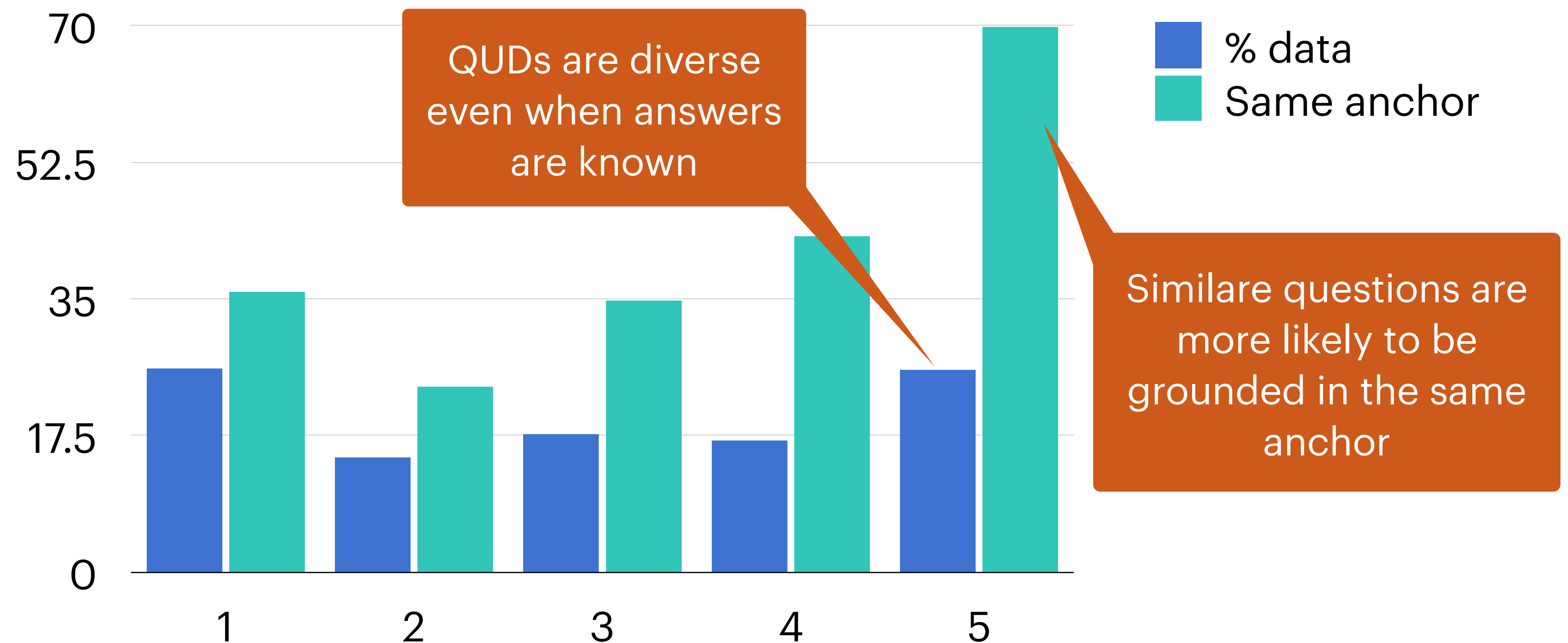
- Source articles: 607 articles WSJ, AP, Newsela
- Question collection:
 - Qualified MTurk annotators
 - 2 workers per article
 - ~22k question-answer pairs in total

Easier, crowdsource-able data collection framework

Analysis: subjectivity

Do annotators ask the same questions given the same answer sentence?

- Human rating of question similarity: 261 question pairs from the same answer sentences



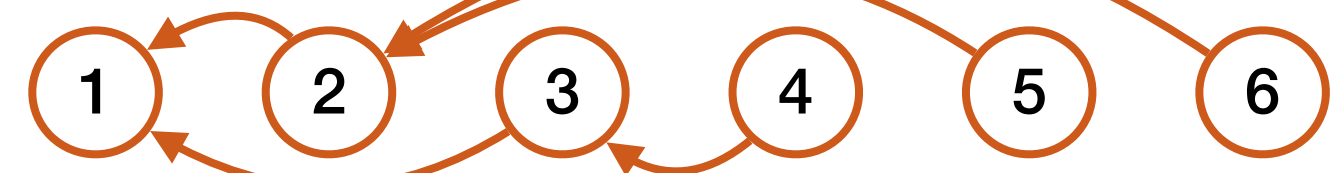
Dependency Parsing of QUD

[1] California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax. [2] The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts. [3] The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure... [4] That leaves the state roughly \$2 billion to \$4 billion short. [5] A sales tax increase appears to be the fastest and easiest to raise funds in a hurry. [6] According to the state department of finance, a one-penny increase in the state's six-cent per dollar sales tax could raise \$3 billion.

1-2 What is prompting the California legislators to seek alternative forms of relief?

2-6 What would be the impact of the proposed increase?

2-5 What would be the advantage of the proposed tax?



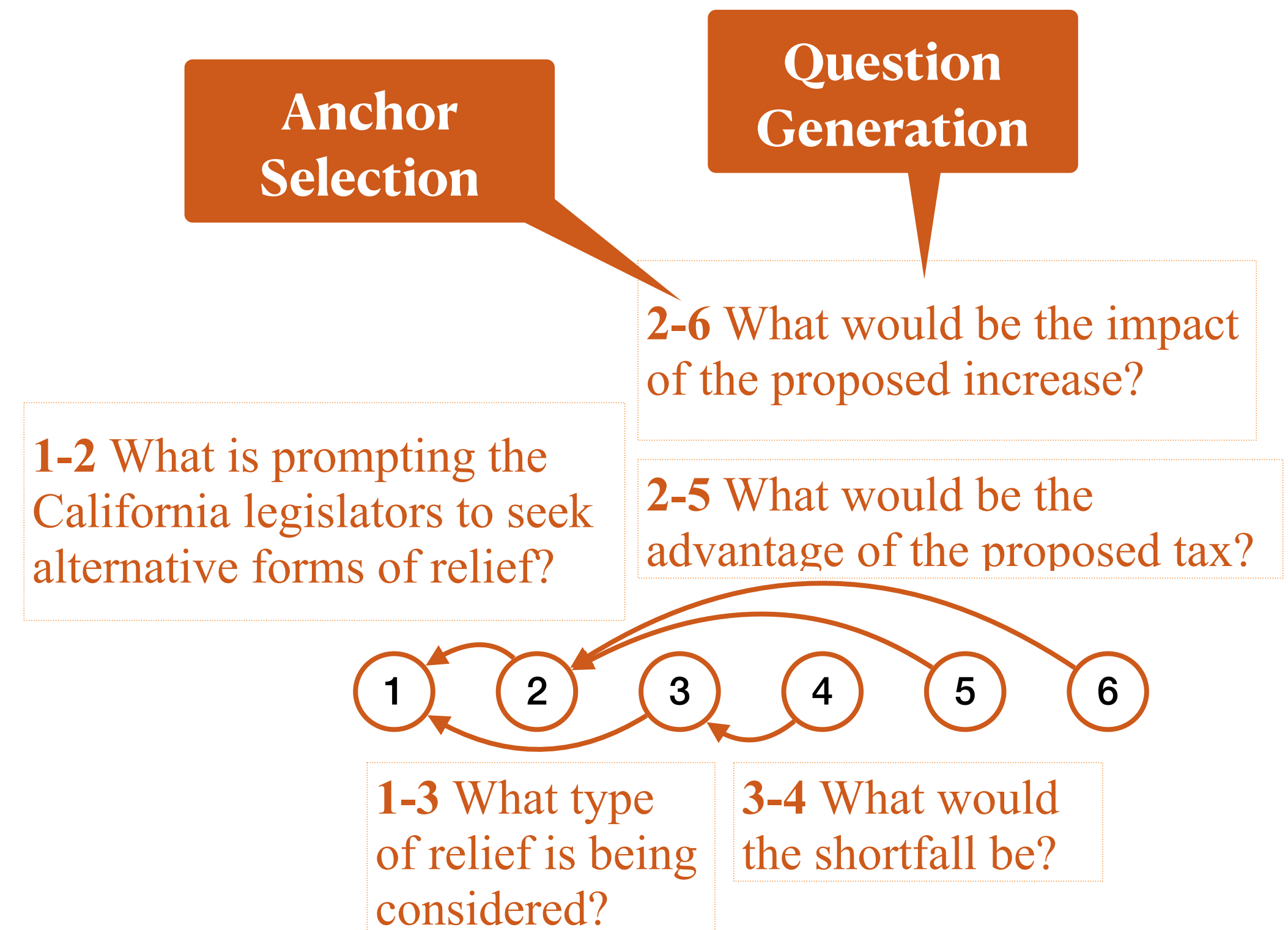
1-3 What type of relief is being considered?

3-4 What would the shortfall be?

Dependency Parsing of QUD

DCQA provides training data!

[1] California legislators, searching for ways to pay for ... damages from last week's earthquake, are laying the groundwork for a temporary increase in the state's sales tax. [2] The talk of a sales tax rise follows a rebuff from Congress on the question of how much the federal government is willing to spend to aid in California's earthquake relief efforts. [3] The state had sought as much as \$4.1 billion in relief, but yesterday the House approved a more general scaled-back measure... [4] That leaves the state roughly \$2 billion to \$4 billion short. [5] A sales tax increase appears to be the fastest and easiest to raise funds in a hurry. [6] According to the state department of finance, a one-penny increase in the state's six-cent per dollar sales tax could raise \$3 billion.



Parsing Model

Fine-tuned on DCQA

$$P(\mathbf{T} | D) = \prod_{i=1}^n [P_a(a_i | D, i) P_q(\mathbf{q}_i | D, i, a_i)]$$

Anchor selection model $P_a(a_i | D, i)$ takes in:

- the answer sentence s_i
 - the article D ,
- to find the most likely sentence where a QUD can be generated, such that s_i is the answer.

Longformer

Question generator $P_q(\mathbf{q}_i | D, i, a_i)$ takes in:

- the answer sentence s_i
 - the anchor sentence at a_i
 - the article D ,
- and aims to generate an appropriate QUD.

Language Model

Parsing Model

Few-shot prompting LLMs

Step 1: question generation

Examples for this question generation are:

Context: {context up to anchor}

Target answer: {answer sentence}

Question: {question}

By reading the context, generate a question that indicates how the Target Answer elaborates on earlier sentences. The Target Answer given should be the answer to the generated question. The question should reflect the main purpose of the Target Answer. It should not use information first introduced in the Target Answer and shouldn't copy-paste phrases newly introduced in the Target Answer.

Step 2: anchor selection

Context: {context prior to anchor}

Question: {question}

Anchor sentence: {anchor sentence}

By reading the Context, pick a sentence from the Context such that the above Question arises from it. An Anchor Sentence is a sentence from the Context that the Question is most related to. The words and concepts from the Anchor Sentence are used to generate the Question. The Target Answer cannot be the Anchor Sentence.

But how do we evaluate the question and anchors?

- Most of the existing metrics for question generation does not look beyond:
 - Whether the question and answer pairs are compatible
 - Whether the question is related to the article
 - Whether the question is similar to existing human-annotated questions
- But these don't quite capture QUDs...

But how do we evaluate the question and anchors?

Linguistically grounded criteria

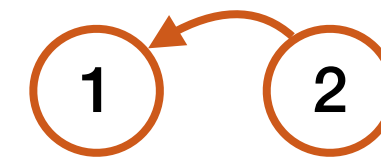
[1] U.S. exports of nuclear material cannot be adequately traced from country to country, according to a congressional report.

[2] ‘Scarcely a day goes by without a report of a new black market deal,’ said Sen. John Glenn ... reacting to the report.

[3] ‘Given the staggering amount of nuclear materials we have exported, it could only be a matter of time before some of this deadly contraband proves to be of U.S. origin.’ ...

[5] The report says hundreds of tons of plutonium and highly enriched uranium have accumulated worldwide...

[6] It does not include figures on U.S. nuclear exports but says 71 export licenses for nuclear materials were granted in 1993.



1-2 What do lawmakers think about this issue?

✓ Language quality

But how do we evaluate the question and anchors?

Linguistically grounded criteria

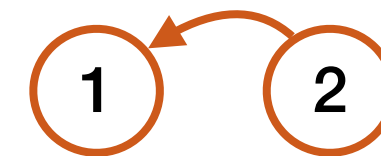
[1] U.S. exports of nuclear material cannot be adequately traced from country to country, according to a congressional report.

[2] ‘Scarcely a day goes by without a report of a new black market deal,’ said Sen. John Glenn ... reacting to the report.

[3] ‘Given the staggering amount of nuclear materials we have exported, it could only be a matter of time before some of this deadly contraband proves to be of U.S. origin.’ ...

[5] The report says hundreds of tons of plutonium and highly enriched uranium have accumulated worldwide...

[6] It does not include figures on U.S. nuclear exports but says 71 export licenses for nuclear materials were granted in 1993.



1-2 What do lawmakers think about this issue?

✓ Language quality

✓ Answer compatible

But how do we evaluate the question and anchors?

Linguistically grounded criteria

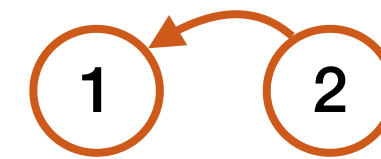
[1] U.S. exports of nuclear material cannot be adequately traced from country to country, according to a congressional report.

[2] ‘Scarcely a day goes by without a report of a new black market deal,’ said Sen. John Glenn ... reacting to the report.

[3] ‘Given the staggering amount of nuclear materials we have exported, it could only be a matter of time before some of this deadly contraband proves to be of U.S. origin.’ ...

[5] The report says hundreds of tons of plutonium and highly enriched uranium have accumulated worldwide...

[6] It does not include figures on U.S. nuclear exports but says 71 export licenses for nuclear materials were granted in 1993.



1-2 What do lawmakers think about this issue?

✓ Language quality

✓ Answer compatible

✓ No new concepts

✓ Grounded in S1

But how do we evaluate the question and anchors?

Linguistically grounded criteria

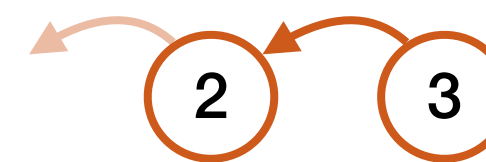
[1] U.S. exports of nuclear material cannot be adequately traced from country to country, according to a congressional report.

[2] ‘Scarcely a day goes by without a report of a new black market deal,’ said Sen. John Glenn ... reacting to the report.

[3] ‘Given the staggering amount of nuclear materials we have exported, it could only be a matter of time before some of this deadly contraband proves to be of U.S. origin.’ ...

[5] The report says hundreds of tons of plutonium and highly enriched uranium have accumulated worldwide...

[6] It does not include figures on U.S. nuclear exports but says 71 export licenses for nuclear materials were granted in 1993.



2-3 What does Glenn say about the situation?

✓ Language quality

✓ Answer compatible

✓ No new concepts

✓ Grounded in S1

But how do we evaluate the question and anchors?

Linguistically grounded criteria

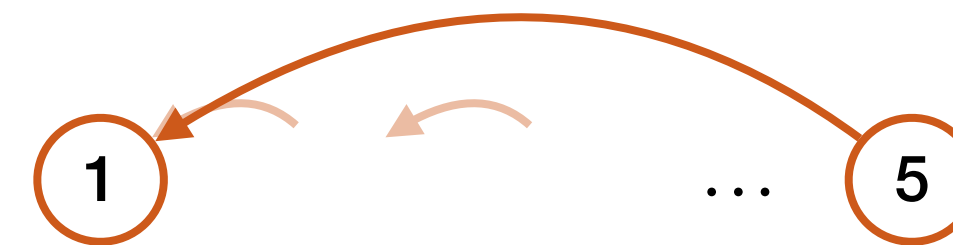
[1] U.S. exports of nuclear material cannot be adequately traced from country to country, according to a congressional report.

[2] ‘Scarcely a day goes by without a report of a new black market deal,’ said Sen. John Glenn ... reacting to the report.

[3] ‘Given the staggering amount of nuclear materials we have exported, it could only be a matter of time before some of this deadly contraband proves to be of U.S. origin.’ ...

[5] The report says hundreds of tons of plutonium and highly enriched uranium have accumulated worldwide...

[6] It does not include figures on U.S. nuclear exports but says 71 export licenses for nuclear materials were granted in 1993.



1-5 What does the report say is the reason for the export ban?

✓ Language quality

But how do we evaluate the question and anchors?

Linguistically grounded criteria

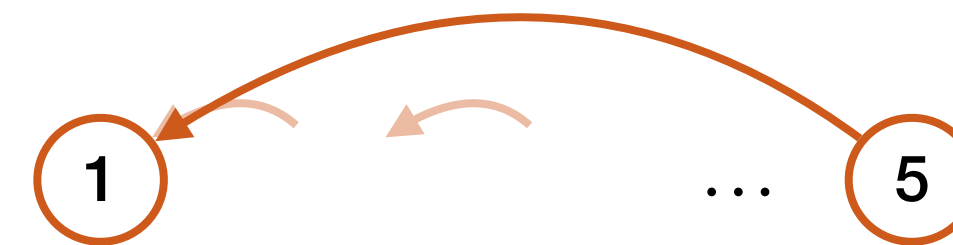
[1] U.S. exports of nuclear material cannot be adequately traced from country to country, according to a congressional report.

[2] ‘Scarcely a day goes by without a report of a new black market deal,’ said Sen. John Glenn ... reacting to the report.

[3] ‘Given the staggering amount of nuclear materials we have exported, it could only be a matter of time before some of this deadly contraband proves to be of U.S. origin.’ ...

[5] The report says hundreds of tons of plutonium and highly enriched uranium have accumulated worldwide...

[6] It does not include figures on U.S. nuclear exports but says 71 export licenses for nuclear materials were granted in 1993.



1-5 What does the report say is the reason for the export ban?

✓ Language quality

But how do we evaluate the question and anchors?

Linguistically grounded criteria

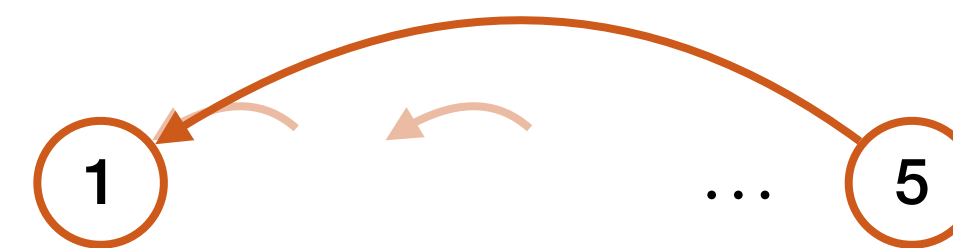
[1] U.S. exports of nuclear material cannot be adequately traced from country to country, according to a congressional report.

[2] ‘Scarcely a day goes by without a report of a new black market deal,’ said Sen. John Glenn ... reacting to the report.

[3] ‘Given the staggering amount of nuclear materials we have exported, it could only be a matter of time before some of this deadly contraband proves to be of U.S. origin.’ ...

[5] The report says hundreds of tons of plutonium and highly enriched uranium have accumulated worldwide...

[6] It does not include figures on U.S. nuclear exports but says 71 export licenses for nuclear materials were granted in 1993.



1-5 What does the report say is the reason for the export ban?

✓ Language quality

× Answer not compatible

But how do we evaluate the question and anchors?

Linguistically grounded criteria

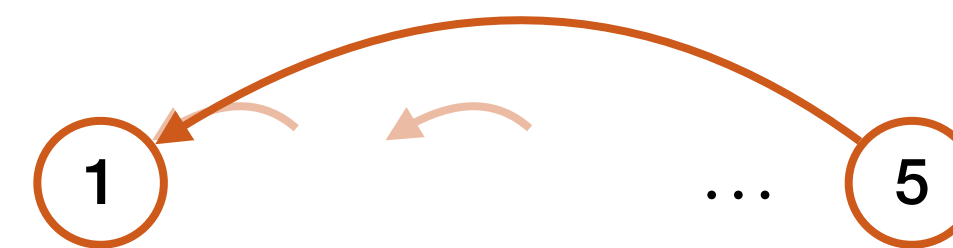
[1] U.S. exports of nuclear material cannot be adequately traced from country to country, according to a congressional report.

[2] ‘Scarcely a day goes by without a report of a new black market deal,’ said Sen. John Glenn ... reacting to the report.

[3] ‘Given the staggering amount of nuclear materials we have exported, it could only be a matter of time before some of this deadly contraband proves to be of U.S. origin.’ ...

[5] The report says hundreds of tons of plutonium and highly enriched uranium have accumulated worldwide...

[6] It does not include figures on U.S. nuclear exports but says 71 export licenses for nuclear materials were granted in 1993.



1-5 What does the report say is the reason for the export ban?

✓ Language quality

× Answer not compatible

× Hallucination in question

~ Somewhat grounded in **S1**

But how do we evaluate the question and anchors?

Linguistically grounded criteria

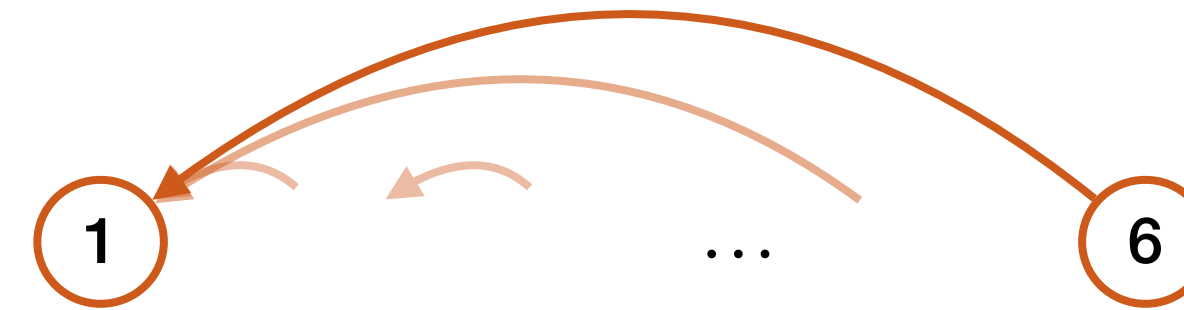
[1] U.S. exports of nuclear material cannot be adequately traced from country to country, according to a congressional report.

[2] ‘Scarcely a day goes by without a report of a new black market deal,’ said Sen. John Glenn ... reacting to the report.

[3] ‘Given the staggering amount of nuclear materials we have exported, it could only be a matter of time before some of this deadly contraband proves to be of U.S. origin.’ ...

[5] The report says hundreds of tons of plutonium and highly enriched uranium have accumulated worldwide...

[6] It does not include figures on U.S. nuclear exports but says 71 export licenses for nuclear materials were granted in 1993.



1-6 What is missing in the report?

✓ Language quality

✓ Answer compatible

× Hallucination in question

~ Somewhat grounded in **S1**

But how do we evaluate the question and anchors?

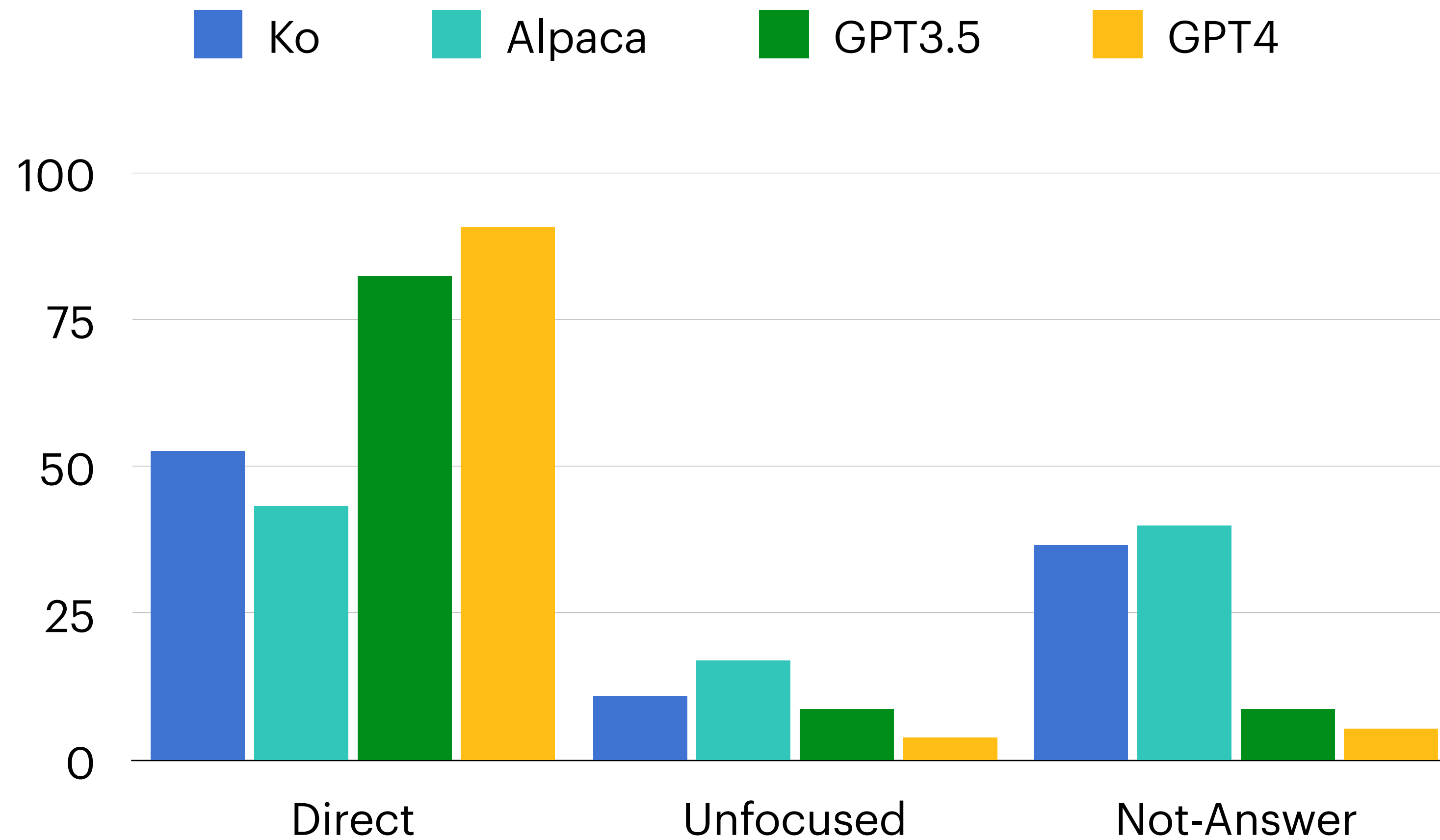
Linguistically grounded criteria

- Linguistic quality
- Answer compatibility
 - Direct and explicit, unfocused, not-an-answer
- Givenness
 - No new concepts, answer leakage, hallucination
- Anchor relevance
 - Fully grounded, partially grounded, ungrounded

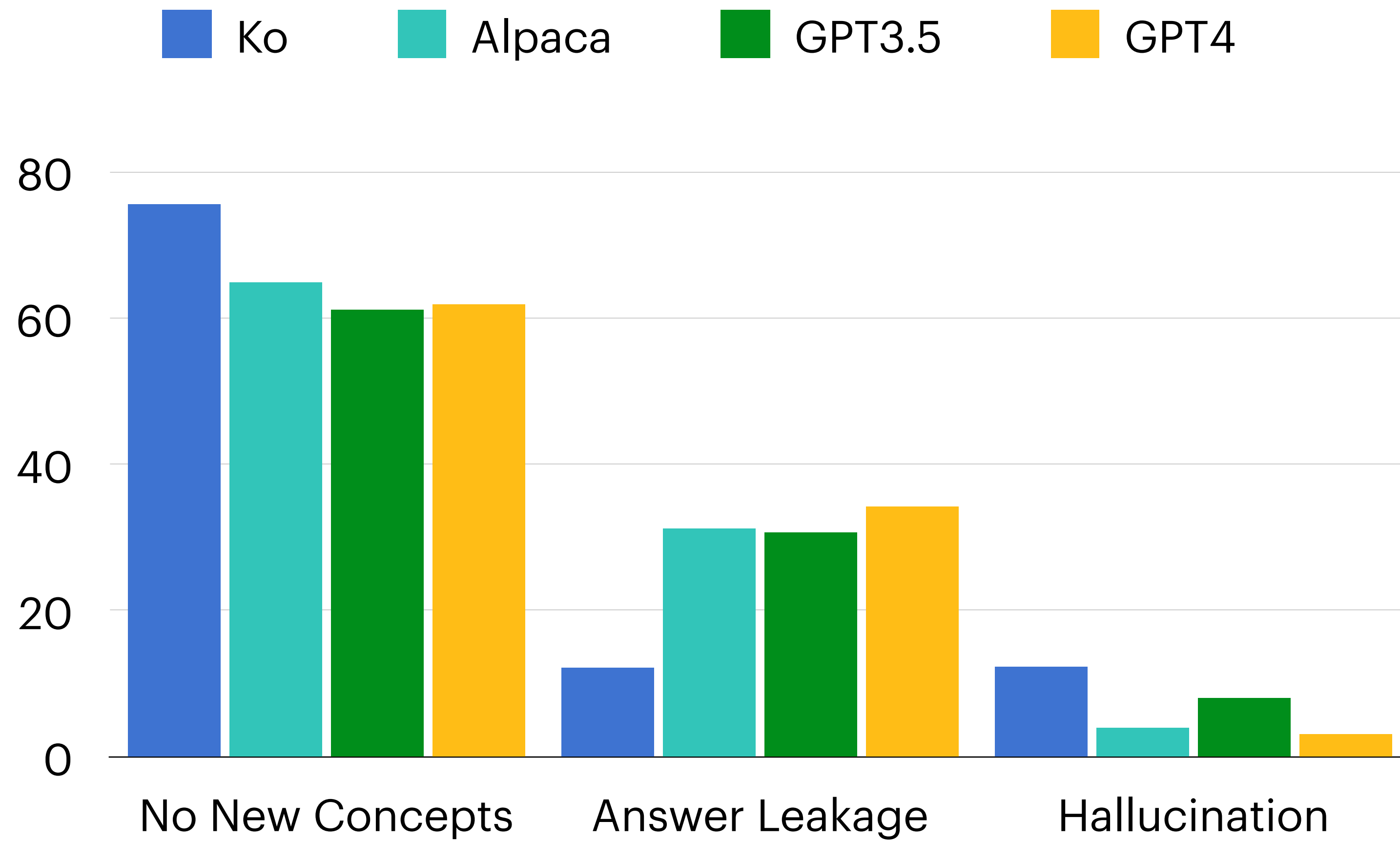
QUDeval Dataset

- 2,190 QUDs generated by 4 systems: Ko et al'23, GPT 3.5, Alpaca, GPT-4
- Annotated by 3 trained linguists

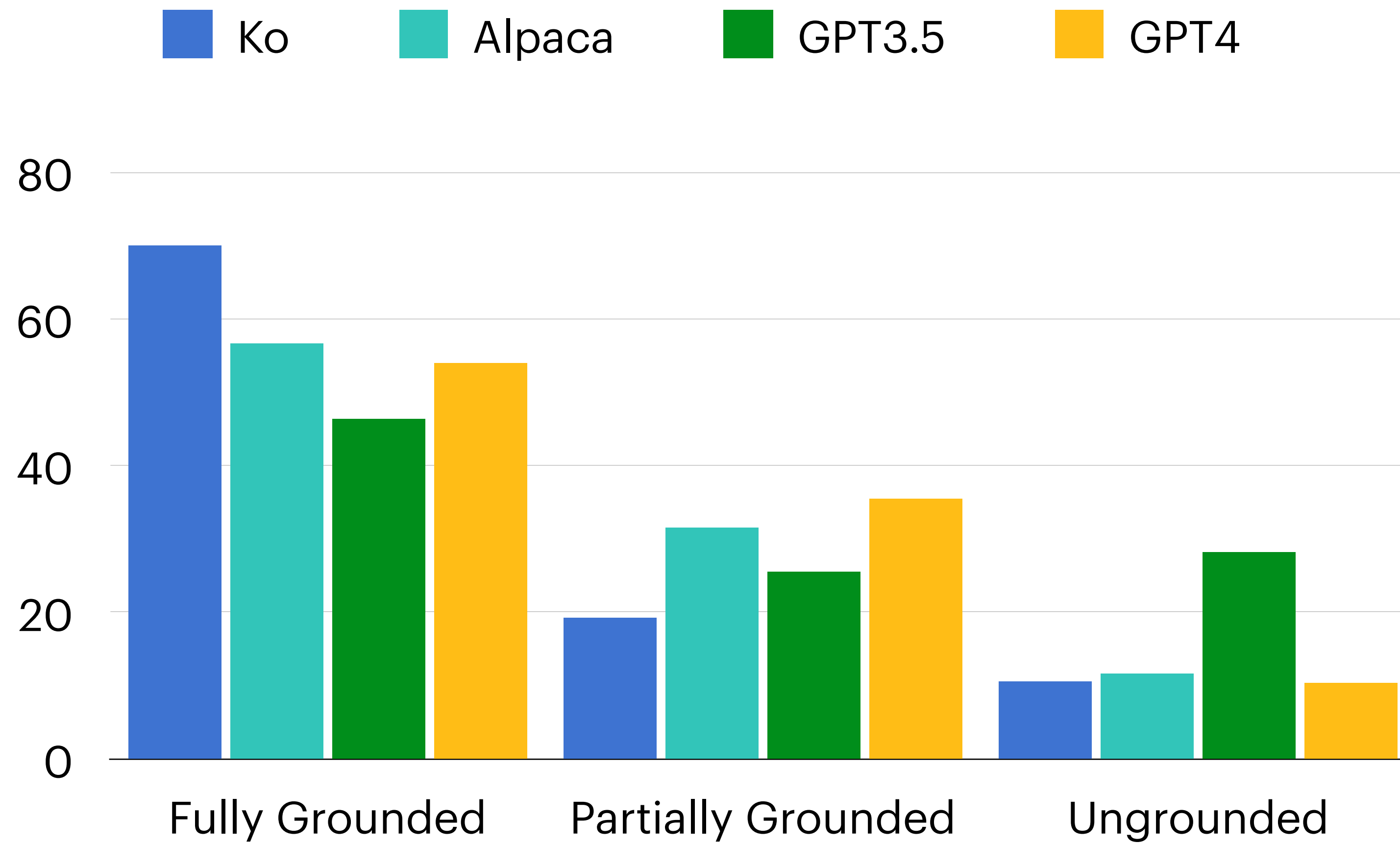
Answer Compatibility



Givenness



Relevance



Automatic Evaluation Metrics

... don't yet work on QUDs

Comp.	Direct Ans.	Unfocused	Not Ans.	Macro F1
GPT-Scr	0.26	0.09	0.24	0.42
GPT-Ans		0.42	0.18	0.32
Random	0.62	0.11	0.31	0.35
Givn.	No new cncpt.	Ans. leak.	Hallu-cinat.	Macro F1
Rule-based	0.52	0.40	0.19	0.37
Hou (2021)	0.77	0.09	0.10	0.32
GPT-Cls-zs	0.76	0.18	0.16	0.37
GPT-Cls-fs	0.45	0.27	0.19	0.30
Random	0.66	0.20	0.08	0.32
Relv.	Fully grnd.	Some grnd.	Not grnd.	Macro F1
Rule-based	0.31	0.35	0.37	0.34
GPT-Cls-zs	0.66	0.11	0.41	0.40
GPT-Cls-fs	0.62	0.23	0.44	0.43
GPT-Scr	0.72	0.34	0.51	0.53
BLEU1-sim	0.59	0.26	0.22	0.35
Random	0.58	0.28	0.15	0.34

Table 4: Reference-free metric assessment results (in F1). For Answer Compatibility, *GPT-Ans* does not differentiate between direct answer and unfocused answer.

Comp.	Direct Ans.	Unfocused	Not Ans.	Macro F1
BLEU-1	0.51	0.14	0.36	0.32
BERTScore	0.51	0.14	0.43	0.36
METEOR	0.63	0.07	0.36	0.36
ROUGE	0.61	0.03	0.37	0.34
QSTS	0.52	0.12	0.38	0.35
GPT-Scr	0.69	0	0.41	0.37
Givn.	No new cncpt.	Ans. leak.	Hallu-cinat.	Macro F1
BLEU-1	0.61	0.21	0.12	0.32
BERTScore	0.4	0.29	0.13	0.28
METEOR	0.45	0.33	0.08	0.29
ROUGE	0.73	0.13	0.10	0.32
QSTS	0.62	0.30	0.07	0.33
GPT-Scr	0.39	0.30	0.18	0.29
Givn.	No new cncpt.	Ans. leak.	Hallu-cinat.	Macro F1
BLEU-1	0.55	0.19	0.21	0.32
BERTScore	0.36	0.27	0.20	0.28
METEOR	0.42	0.27	0.24	0.32
ROUGE	0.63	0.19	0.18	0.34
QSTS	0.57	0.21	0.22	0.34
GPT-Scr	0.54	0.26	0.20	0.34

Table 5: Reference-based metric assessment results.

This talk: discourse & generation

What is QUD and what does QUD structures look like?

- Discourse Comprehension: A Question Answering Framework to Represent Sentence Connections [Wei-Jen Ko et al, EMNLP'22]
- Discourse Analysis via Questions and Answers: Parsing Dependency Structures of Questions Under Discussion [Wei-Jen Ko, Yating Wu et al, ACL Findings'23]
- QUDeval: The Evaluation of Questions Under Discussion Discourse Parsing [Yating Wu et al., EMNLP'23]

How do we ask (generate) questions?

- Inquisitive Question Generation for High Level Text Comprehension [Wei-Jen Ko et al, EMNLP'20]
- Which questions should I answer? Saliency Prediction of Inquisitive Questions [Yating Wu, Ritika Mangla et al, EMNLP '24]

What can discourse do for generation?

- Elaborative Simplification as Implicit Questions Under Discussion [Yating Wu, William Sheffield et al, EMNLP'23]
- InfoLossQA: Characterizing and Recovering Information Loss in Text Simplification [Jan Trienes et al., ACL '24]

This talk: discourse & generation

What is QUD and what does QUD structures look like?

- Discourse Comprehension: A Question Answering Framework to Represent Sentence Connections [Wei-Jen Ko et al, EMNLP'22]
- Discourse Analysis via Questions and Answers: Parsing Dependency Structures of Questions Under Discussion [Wei-Jen Ko, Yating Wu et al, ACL Findings'23]
- QUDeval: The Evaluation of Questions Under Discussion Discourse Parsing [Yating Wu et al., EMNLP'23]

How do we ask (generate) questions?

- Inquisitive Question Generation for High Level Text Comprehension [Wei-Jen Ko et al, EMNLP'20]
- Which questions should I answer? Saliency Prediction of Inquisitive Questions [Yating Wu, Ritika Mangla et al, EMNLP '24]

What can discourse do for generation?

- Elaborative Simplification as Implicit Questions Under Discussion [Yating Wu, William Sheffield et al, EMNLP'23]
- InfoLossQA: Characterizing and Recovering Information Loss in Text Simplification [Jan Trienes et al., ACL '24]

Two ways to think about questions...

Way #2: Potential Questions (Onea, 2016)



Inquisitive
reader

Prior context of the article...

This gritty city of 7.6 million rarely gets respect. It often ranks low in “livability” surveys and near the top of the hemisphere’s ugliest capitals.

Two ways to think about questions...

Way #2: Potential Questions (Onea, 2016)



Why is it referred
to as gritty?

Why does it rarely
get respect?

This **gritty** city of 7.6 million **rarely gets respect**. It often ranks low in
“**livability**” surveys and near the top of the hemisphere’s **ugliest capitals**.

What constitutes
“livability”?

Which are the
other ones?

But which questions **should** be answered?

Empirical (NLP)

- Generating these questions can be useful!
- But we don't want an over-generation of questions.
- We want a question salience prediction model.
- But is question salience a “real” notion here?

Theoretical

- Onea (2016): only described hypotheses, mostly related to QUD validity.
- Expectation-driven QUDs (Kehler and Rohde, 2017): readers form expectations how discourse would progress.
- Van Rooy (2003): answer utility!
- Do readers form mental models of salience?
- Is there a bridge between potential questions and QUDs?

Let's take a look at salience!

[1] Amid skepticism that Russia's war in Chechnya can be ended across a negotiating table, peace talks were set to resume Wednesday in neighboring Ingushetia. **[2]** The scheduled resumption of talks in the town of Sleptsovsk came two days after agreement on a limited cease-fire, calling for both sides to stop using heavy artillery Tuesday. **[3]** They also agreed in principle to work out a mechanism for exchanging prisoners of war and the dead. **[4]** Despite the pact, artillery fire sounded in the Grozny on Tuesday, and there were reports of Chechen missile attacks southwest of the Chechen capital.

[Q1] What is the significance of the limited cease-fire agreement that was reached?

[Q2] What was the reason behind the artillery fire in Grozny on Tuesday despite the agreed cease-fire?

[Q3] What are the reports of Chechen missile attacks southwest of the Chechen capital?

[Q4] What is the source of the Chechen missile attacks?

QSalience data collection

- Question salience ratings: **1 (unrelated) — 5 (definitely should be answered); invalid**
- Natural language rationale (justification) of the ratings.

[1] Amid skepticism that Russia's war in Chechnya can be ended across a negotiating table, peace talks were set to resume Wednesday in neighboring Ingushetia. **[2]** The scheduled resumption of talks in the town of Sleptsovsk came two days after agreement on a limited cease-fire, calling for both sides to stop using heavy artillery Tuesday.

[Q1] What other progress has been made towards peace recently?

Salience: 4. Elaborates on what is being done to establish peace in the region

[3] They also agreed in principle to work out a mechanism for exchanging prisoners of war and the dead. **[4]** Despite the pact, artillery fire sounded in the Grozny on Tuesday, and there were reports of Chechen missile attacks southwest of the Chechen capital.

[Q2] What is the significance of the limited cease-fire agreement that was reached?

Invalid. Incorrect anchor

[Q3] What was the reason behind the artillery fire in Grozny on Tuesday despite the agreed cease-fire?

Salience: 5. This question would be useful in understanding why the cease-fire was broken, which could give insight into how optimistic the peace talks will be.

[Q4] What are the reports of Chechen missile attacks southwest of the Chechen capital?

Salience: 3. This question doesn't interest me; why there are missing attacks would help my understanding more

[Q5] What is the source of the Chechen missile attacks?

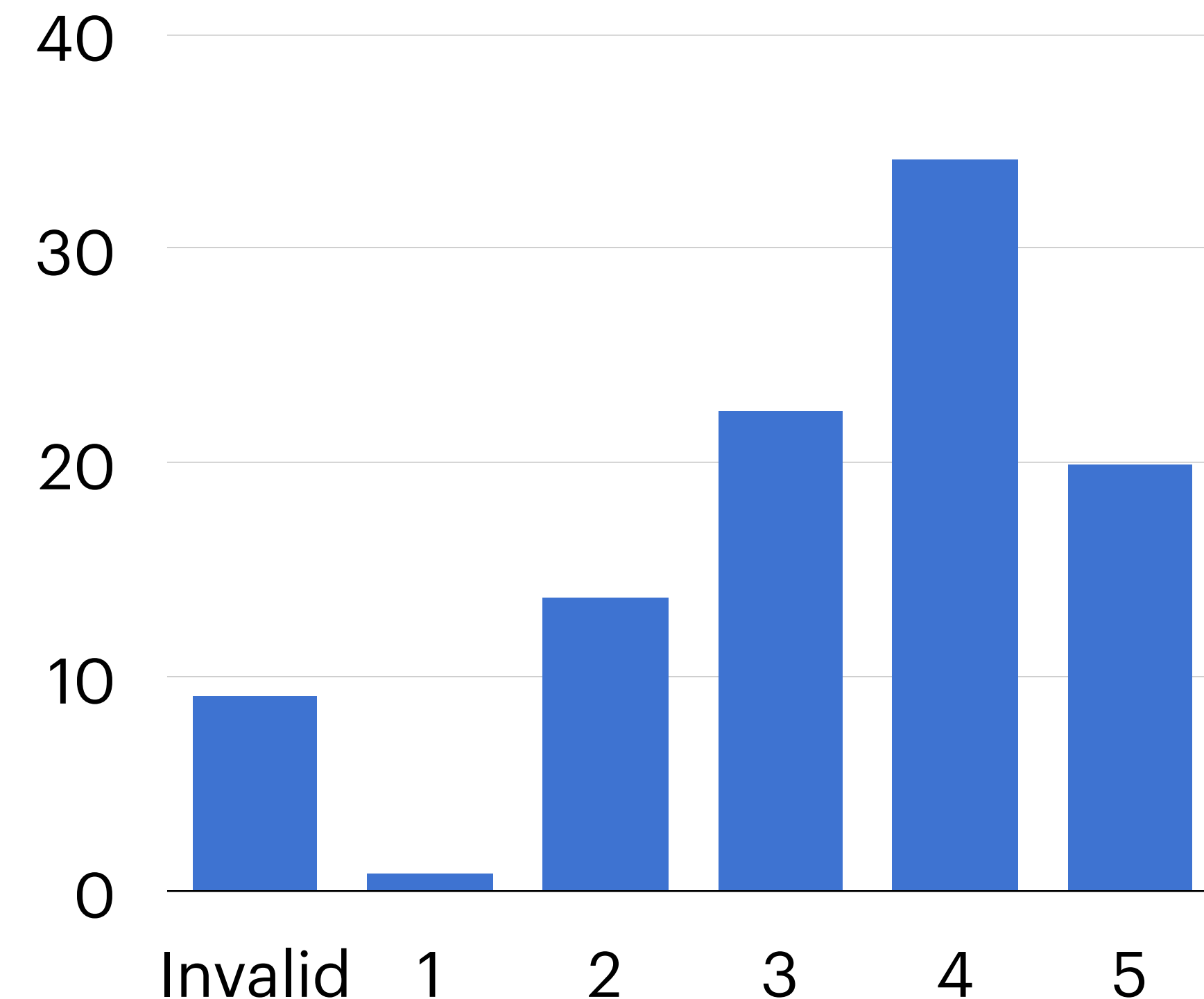
Salience: 2. Based on context, can be inferred that attack comes from Russia

[5] Many Chechens are fighting independently of the forces loyal to Chechen President Dzhokhar Dudayev, and Dudayev's representative at the peace talks, Aslan Maskhadov, has warned that he does not control them. [...]

QSalience data collection

- Source texts:
 - News texts, news summaries, TED-talks
- Question generation:
 - Llama-2-7b-chat, Mistral-7b-instruct, GPT-3.5-turbo, GPT-4-turbo
 - LLMs are shown context up to the anchor sentence

# articles	# questions	avg. length	std.dev
66	1766	14	4.6



Agreement (alpha): 0.63-0.75

Answerability annotation

- Answerability ratings: **unanswered (1), partially answered (2), fully answered (3)**
- Sentences that (partially) answer the question.

[1] Amid skepticism that Russia's war in Chechnya can be ended across a negotiating table, peace talks were set to resume Wednesday in neighboring Ingushetia. **[2]** The scheduled resumption of talks in the town of Sleptsovsk came two days after agreement on a limited cease-fire, calling for both sides to stop using heavy artillery Tuesday.

[Q1] What other progress has been made towards peace recently?

Salience: 4. Elaborates on what is being done to establish peace in the region

Answered in [3]

[3] They also agreed in principle to work out a mechanism for exchanging prisoners of war and the dead. **[4]** Despite the pact, artillery fire sounded in the Grozny on Tuesday, and there were reports of Chechen missile attacks southwest of the Chechen capital.

[Q2] What is the significance of the limited cease-fire agreement that was reached?

Invalid. Incorrect anchor

[Q3] What was the reason behind the artillery fire in Grozny on Tuesday despite the agreed cease-fire?

Salience: 5. This question would be useful in understanding why the cease-fire was broken, which could give insight into how optimistic the peace talks will be.

Answered in [5]

[Q4] What are the reports of Chechen missile attacks southwest of the Chechen capital?

Salience: 3. This question doesn't interest me; why there are missing attacks would help my understanding more

Not Answered

[Q5] What is the source of the Chechen missile attacks?

Salience: 2. Based on context, can be inferred that attack comes from Russia

Answered in context

[5] Many Chechens are fighting independently of the forces loyal to Chechen President Dzhokhar Dudayev, and Dudayev's representative at the peace talks, Aslan Maskhadov, has warned that he does not control them. [...]

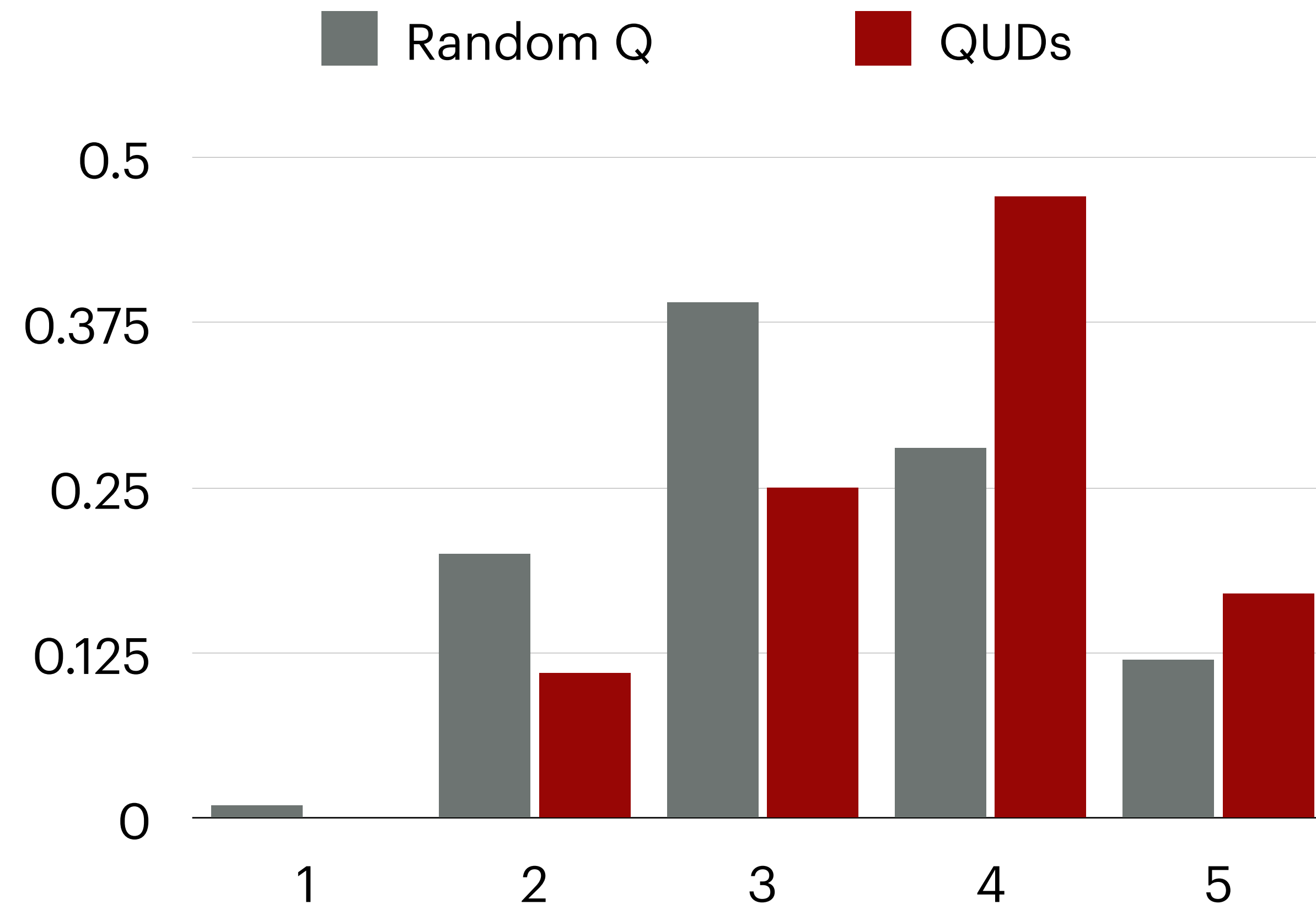
Do salience and answerability correlate?

- Spearman's rho between:
 - Annotated salience vs. answerability of a random question in the dataset
 - Annotated salience vs. answerability of the current question

	Human Salience
Random Questions	-0.02*
Answerability	0.65

Are QUDs salient potential questions?

- Salience of fandom questions vs human-annotated QUDs in DCQA



We believe there is strong evidence that salient questions are more likely to be QUDs

QSalience: building a saliience model

- Instruction fine-tuning:
 - Mistral-7b-instruct, Llama-2-7b-chat, TinyLlama-1.1b-chat, Flan-T5-base
 - Method: QLoRA
- LLM prompting and in-context learning:
 - Zero-shot, few shot
 - Few-shot with example retrieval
 - Chain-of-Thought (zero- and few-shot)
 - All with GPT-4

QSalience: results

This is a task that is difficult for LLMs to grasp!

Model	MAE ↓	Spearman ↑	Macro F1 ↑	krippendorff's α ↑
GPT4 zero-shot (vanilla)	1.314	0.229	0.193	-0.141
GPT4 few-shot (vanilla)	0.910	0.417	0.316	0.358
GPT4 few-shot (kNN)	1.063	0.359	0.245	0.215
GPT4 CoT zero-shot	1.144	0.366	0.197	0.058
GPT4 CoT few-shot	1.034	0.327	0.292	0.165

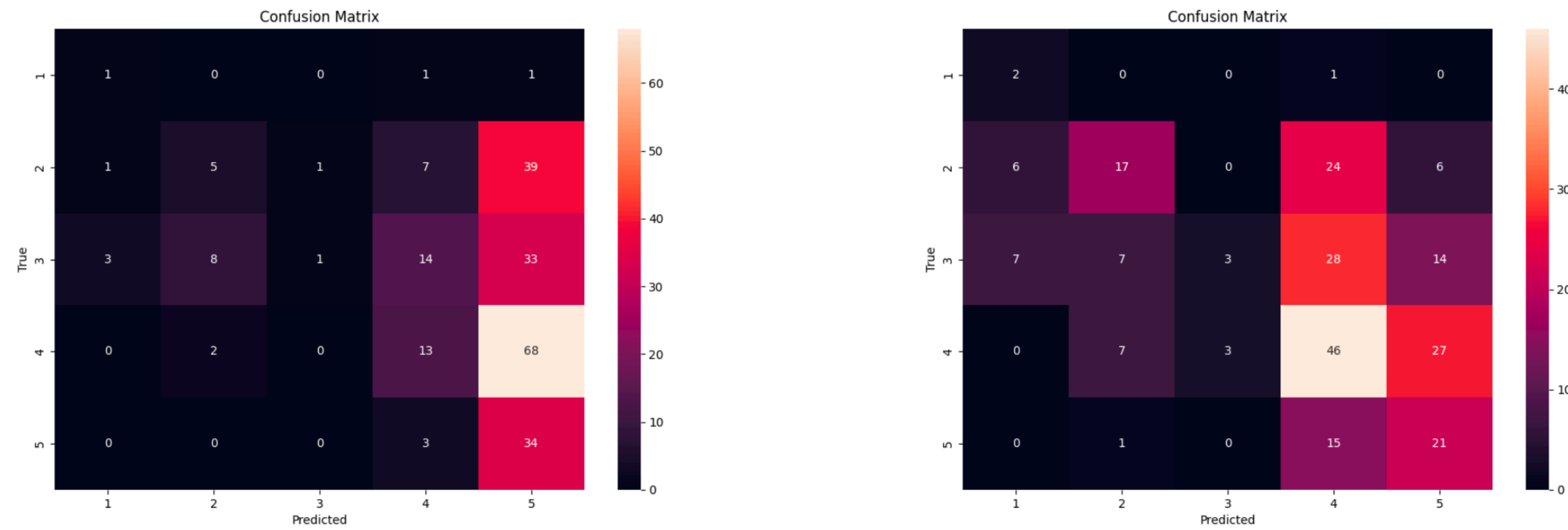
QSalience: results

This is a task that is difficult for LLMs to grasp!

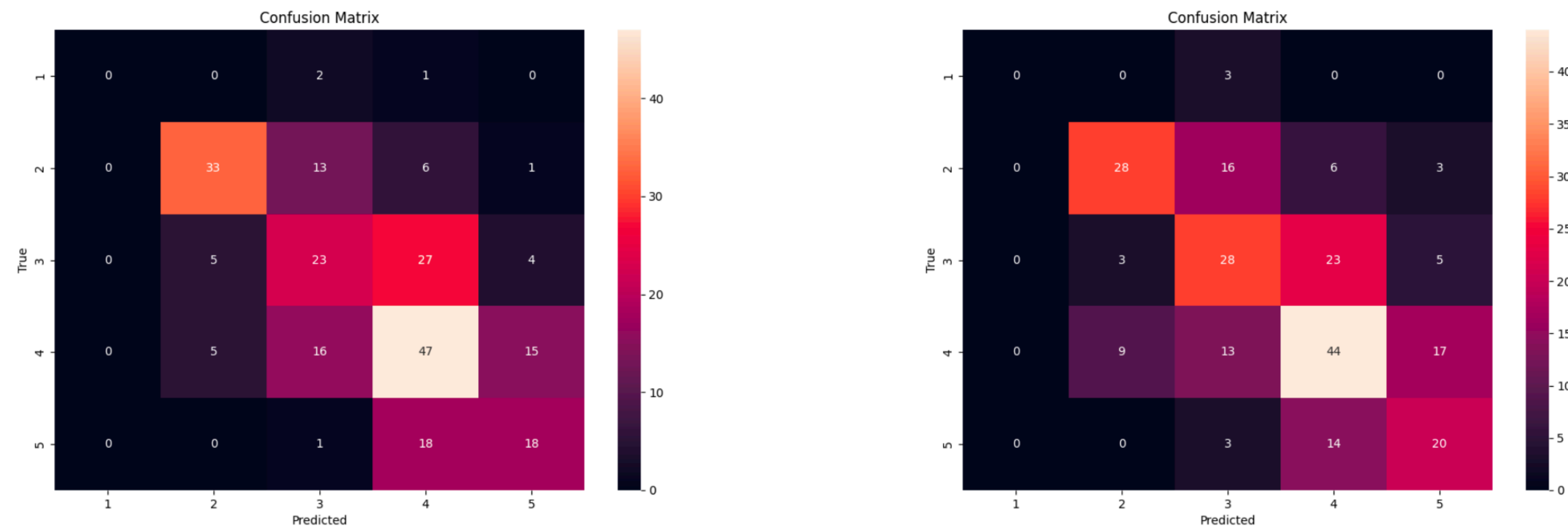
Model	MAE ↓	Spearman ↑	Macro F1 ↑	krippendorff's α ↑
GPT4 zero-shot (vanilla)	1.314	0.229	0.193	-0.141
GPT4 few-shot (vanilla)	0.910	0.417	0.316	0.358
GPT4 few-shot (kNN)	1.063	0.359	0.245	0.215
GPT4 CoT zero-shot	1.144	0.366	0.197	0.058
GPT4 CoT few-shot	1.034	0.327	0.292	0.165
QSALIENCE (Mistral-7B-instruct)	0.579	0.623	0.417	0.615
Llama-2-7B-chat	0.626	0.566	0.413	0.557
Flan-t5-base	0.706	0.542	0.370	0.526
TinyLlama-1.1B-chat	0.664	0.522	0.402	0.496

QSalience: results

This is a task that is difficult for LLMs to grasp!



(a) GPT-4-turbo zero-shot vanilla (left), GPT-4-turbo few-shot vanilla (right)



(b) Mistral-7B-Instruct (left), Llama-2-7B-chat (right)

Use case: summary expansion

Better expansions answer more salient questions?



TL;DR Hong Kong and Singaporean authorities welcome UBS's acquisition of Credit Suisse. Temasek plans to invest in the Philippines sovereign wealth fund, and CDPQ may buy part of ATC's Indian unit. Meanwhile, the US prepares to inspect Hong Kong auditors, and Taiwan's FSC says local institutions' Credit Suisse exposure isn't severe.

Generated Potential Questions

How will UBS's acquisition of Credit Suisse affect the banking industry in Hong Kong and Singapore?

Salience: 5

What is the reason behind the US's intent to inspect Hong Kong auditors?

Salience: 5

What specific part of ATC's Indian unit is CDPQ considering to buy?

Salience: 4

What does the Taiwan FSC's statement imply about the financial health of their local institutions in relation to Credit Suisse?

Salience: 3

Expanded Summary 1

... The Hong Kong Monetary Authority (HKMA) ... highlighted the minimal exposure of the local banking sector to Credit Suisse... Similarly, the Monetary Authority of Singapore (MAS) assured that the takeover would not impact the stability of Singapore's banking system... In investment news, Singapore's Temasek Holdings has shown interest in the Maharlika Investment Fund... Additionally, Caisse de depot et Placement du Quebec (CDPQ) is reportedly in talks to acquire a significant stake in the Indian subsidiary of American Tower Corp, indicating a keen interest in the telecommunications infrastructure sector in India. On the regulatory front, the US is set to begin a new round of inspections on Chinese companies' auditors in Hong Kong, reflecting ongoing efforts to ensure compliance and prevent delistings from the New York stock exchange. Meanwhile, Taiwan's Financial Supervisory Commission (FSC) has downplayed concerns over local institutions' exposure to Credit Suisse...

Expanded Summary 2

... In parallel developments, Temasek has shown interest in investing in the Philippines' sovereign wealth fund, while CDPQ is considering purchasing a portion of ATC's operations in India. This period also sees the United States gearing up to conduct inspections on Hong Kong auditors, indicating increased scrutiny in financial practices. In Taiwan, the Financial Supervisory Commission (FSC) has addressed concerns regarding local institutions' exposure to Credit Suisse, stating that the impact is not severe... This figure represents merely 0.35% of the total usable funds within the insurance sector, underscoring the limited risk Credit Suisse's situation poses to Taiwan's financial stability. The welcoming stance of the Hong Kong Monetary Authority (HKMA) and the Securities and Futures Commission (SFC) towards the UBS and Credit Suisse deal further underscores the regional support for consolidations that aim to stabilize the banking sector...

Use case: summary expansion

Better expansions answer more salient questions?



TL;DR Hong Kong and Singaporean authorities welcome UBS's acquisition of Credit Suisse. Temasek plans to invest in the Philippines sovereign wealth fund, and CDPQ may buy part of ATC's Indian unit. Meanwhile, the US prepares to inspect Hong Kong auditors, and Taiwan's FSC says local institutions' Credit Suisse exposure isn't severe.

Generated Potential Questions

How will UBS's acquisition of Credit Suisse affect the banking industry in Hong Kong and Singapore?

Salience: 5

What is the reason behind the US's intent to inspect Hong Kong auditors?

Salience: 5

What specific part of ATC's Indian unit is CDPQ considering to buy?

Salience: 4

What does the Taiwan FSC's statement imply about the financial health of their local institutions in relation to Credit Suisse?

Salience: 3

Expanded Summary 1

... The Hong Kong Monetary Authority (HKMA) ... highlighted the minimal exposure of the local banking sector to Credit Suisse... Similarly, the Monetary Authority of Singapore (MAS) assured that the takeover would not impact the stability of Singapore's banking system... In investment news, Singapore's Temasek Holdings has shown interest in the Maharlika Investment Fund... Additionally, Caisse de depot et Placement du Quebec (CDPQ) is reportedly in talks to acquire a significant stake in the Indian subsidiary of American Tower Corp, indicating a keen interest in the telecommunications infrastructure sector in India. On the regulatory front, the US is set to begin a new round of inspections on Chinese companies' auditors in Hong Kong, reflecting ongoing efforts to ensure compliance and prevent delistings from the New York stock exchange. Meanwhile, Taiwan's Financial Supervisory Commission (FSC) has downplayed concerns over local institutions' exposure to Credit Suisse...

better

Expanded Summary 2

... In parallel developments, Temasek has shown interest in investing in the Philippines' sovereign wealth fund, while CDPQ is considering purchasing a portion of ATC's operations in India. This period also sees the United States gearing up to conduct inspections on Hong Kong auditors, indicating increased scrutiny in financial practices. In Taiwan, the Financial Supervisory Commission (FSC) has addressed concerns regarding local institutions' exposure to Credit Suisse, stating that the impact is not severe... This figure represents merely 0.35% of the total usable funds within the insurance sector, underscoring the limited risk Credit Suisse's situation poses to Taiwan's financial stability. The welcoming stance of the Hong Kong Monetary Authority (HKMA) and the Securities and Futures Commission (SFC) towards the UBS and Credit Suisse deal further underscores the regional support for consolidations that aim to stabilize the banking sector...

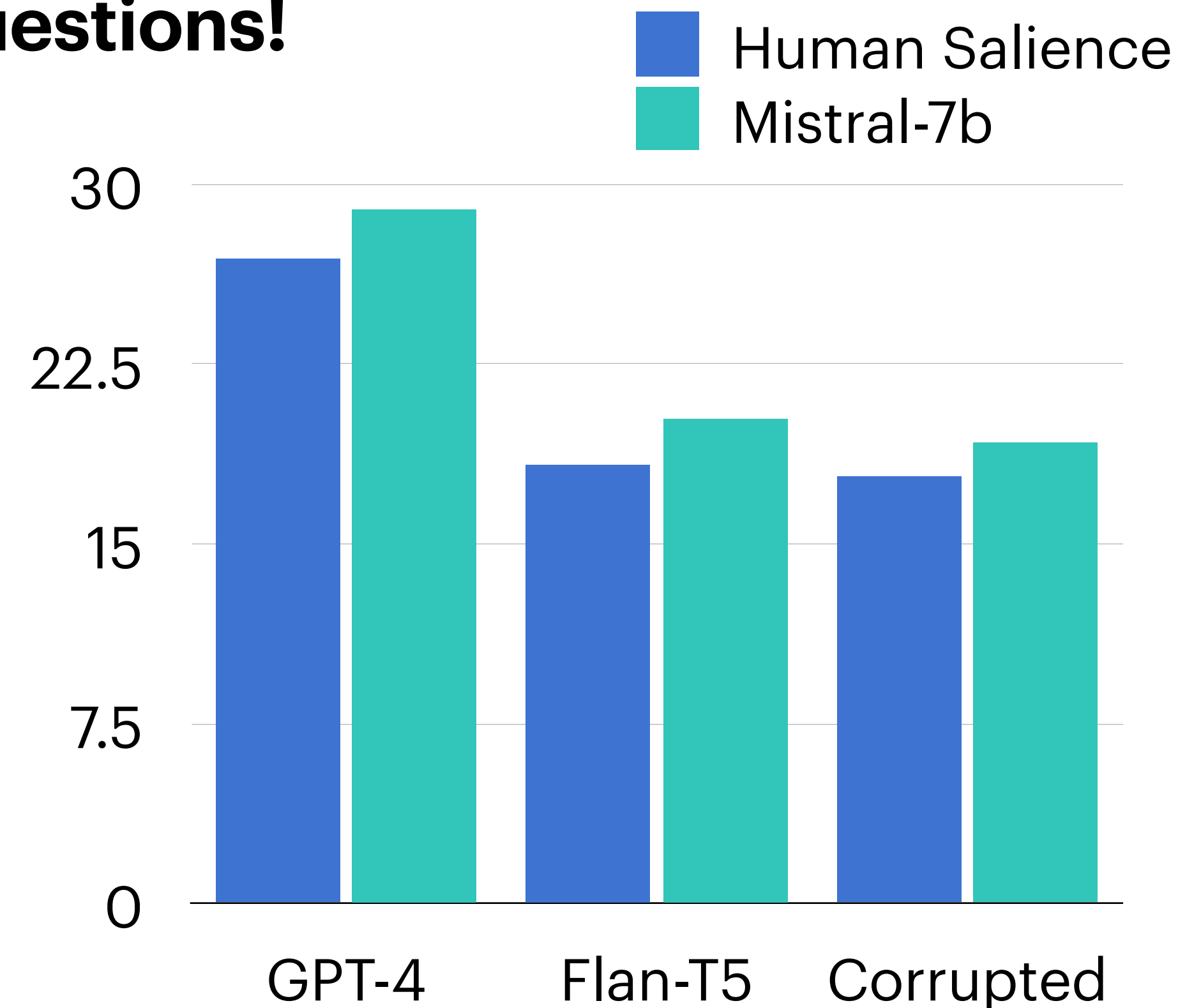
worse

Use case: summary expansion

Yes, better expansions answer more salient questions!

- Best expansion: GPT-4
- Medium quality expansion: Flan-T5 with GPT-4 touchups
- Poor expansion: Corrupted GPT-4

$$\text{Summscr} = \frac{1}{n} \sum_{i=1}^n \sum_{q \in Q_i} \textit{salience}(q)$$



This is recovering system-level rankings (from humans)

This talk: discourse & generation

What is QUD and what does QUD structures look like?

- Discourse Comprehension: A Question Answering Framework to Represent Sentence Connections [Wei-Jen Ko et al, EMNLP'22]
- Discourse Analysis via Questions and Answers: Parsing Dependency Structures of Questions Under Discussion [Wei-Jen Ko, Yating Wu et al, ACL Findings'23]
- QUDeval: The Evaluation of Questions Under Discussion Discourse Parsing [Yating Wu et al., EMNLP'23]

How do we ask (generate) questions?

- Inquisitive Question Generation for High Level Text Comprehension [Wei-Jen Ko et al, EMNLP'20]
- Which questions should I answer? Saliency Prediction of Inquisitive Questions [Yating Wu, Ritika Mangla et al, EMNLP '24]

What can discourse do for generation?

- Elaborative Simplification as Implicit Questions Under Discussion [Yating Wu, William Sheffield et al, EMNLP'23]
- InfoLossQA: Characterizing and Recovering Information Loss in Text Simplification [Jan Trienes et al., ACL '24]

This talk: discourse & generation

What is QUD and what does QUD structures look like?

- Discourse Comprehension: A Question Answering Framework to Represent Sentence Connections [Wei-Jen Ko et al, EMNLP'22]
- Discourse Analysis via Questions and Answers: Parsing Dependency Structures of Questions Under Discussion [Wei-Jen Ko, Yating Wu et al, ACL Findings'23]
- QUDeval: The Evaluation of Questions Under Discussion Discourse Parsing [Yating Wu et al., EMNLP'23]

How do we ask (generate) questions?

- Inquisitive Question Generation for High Level Text Comprehension [Wei-Jen Ko et al, EMNLP'20]
- Which questions should I answer? Saliency Prediction of Inquisitive Questions [Yating Wu, Ritika Mangla et al, EMNLP '24]

What can discourse do for generation?

- Elaborative Simplification as Implicit Questions Under Discussion [Yating Wu, William Sheffield et al, EMNLP'23]
- InfoLossQA: Characterizing and Recovering Information Loss in Text Simplification [Jan Trienes et al., ACL '24]

Document Simplification

...involves content *addition*



Results, she said, “could help the team better understand ancient Egyptian health” and, correspondingly, modern-day health.

For instance, some mummies still have arteries in their mummified remains, Miller-Thomas said.

And, sometimes, scientists can tell if those arteries had hardened.



The scans could help the team understand about ancient Egyptians’ health.

For example, some mummies still have arteries.

An artery is a tube that moves blood through the body.

The artery could show if the person had been healthy or not.

Define “artery”

Explain

Document Simplification

...involves content *addition*



Results, she said, “could help the team better understand ancient Egyptian health” and, correspondingly, modern-day health.

For instance, some mummies still have arteries in their mummified remains, Miller-Thomas said.

And, sometimes, scientists can tell if those arteries had hardened.



The scans could help the team understand about ancient Egyptians’ health.

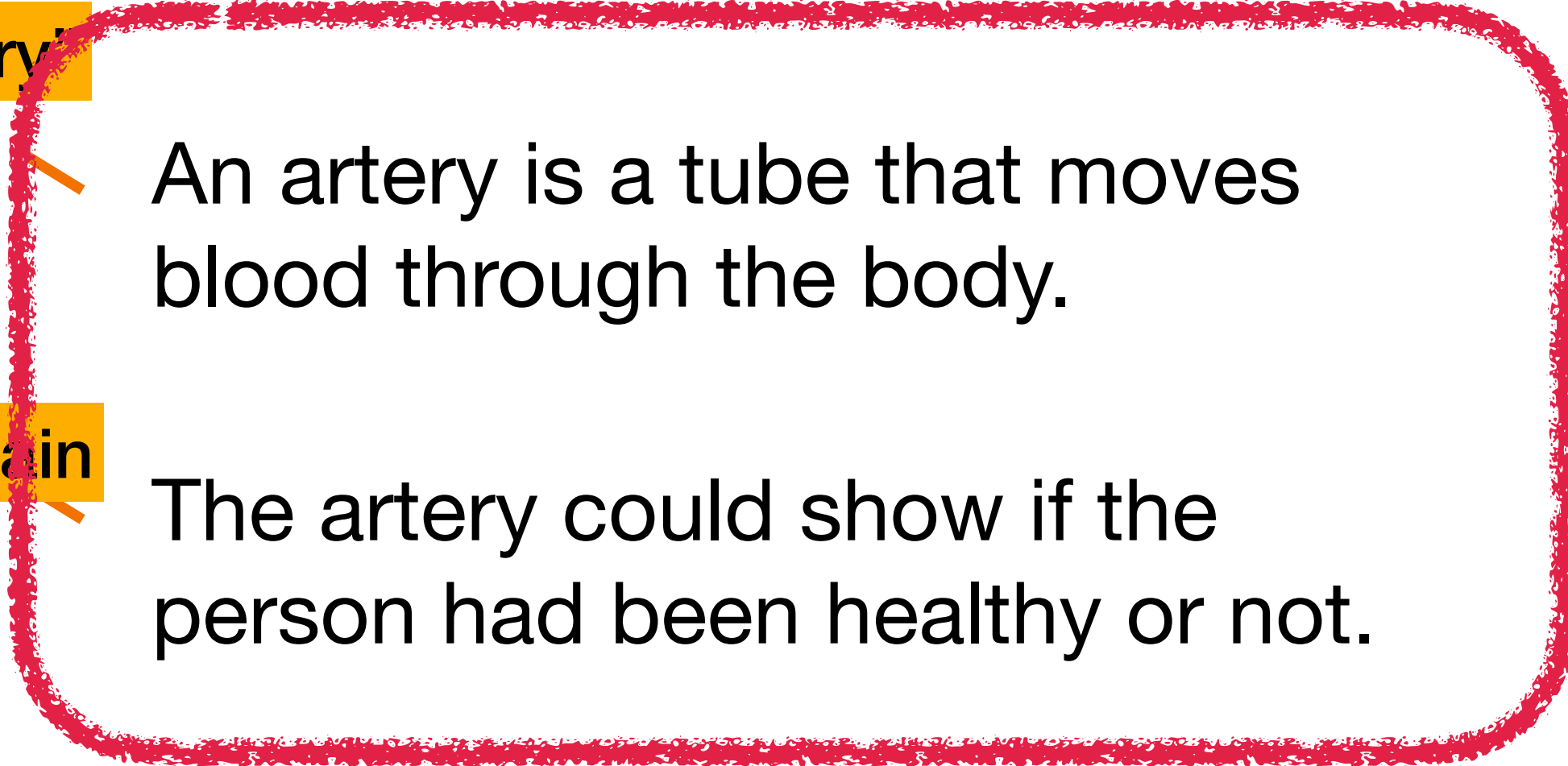
For example, some mummies still have arteries.

An artery is a tube that moves blood through the body.

The artery could show if the person had been healthy or not.

Define “artery”

Explain



Elaborative Simplification

Key idea: viewing elaborations as implicit QUDs

The scans could help the team understand about ancient Egyptians' health. For example, some mummies still have arteries.

Implicit QUD: What is an artery?

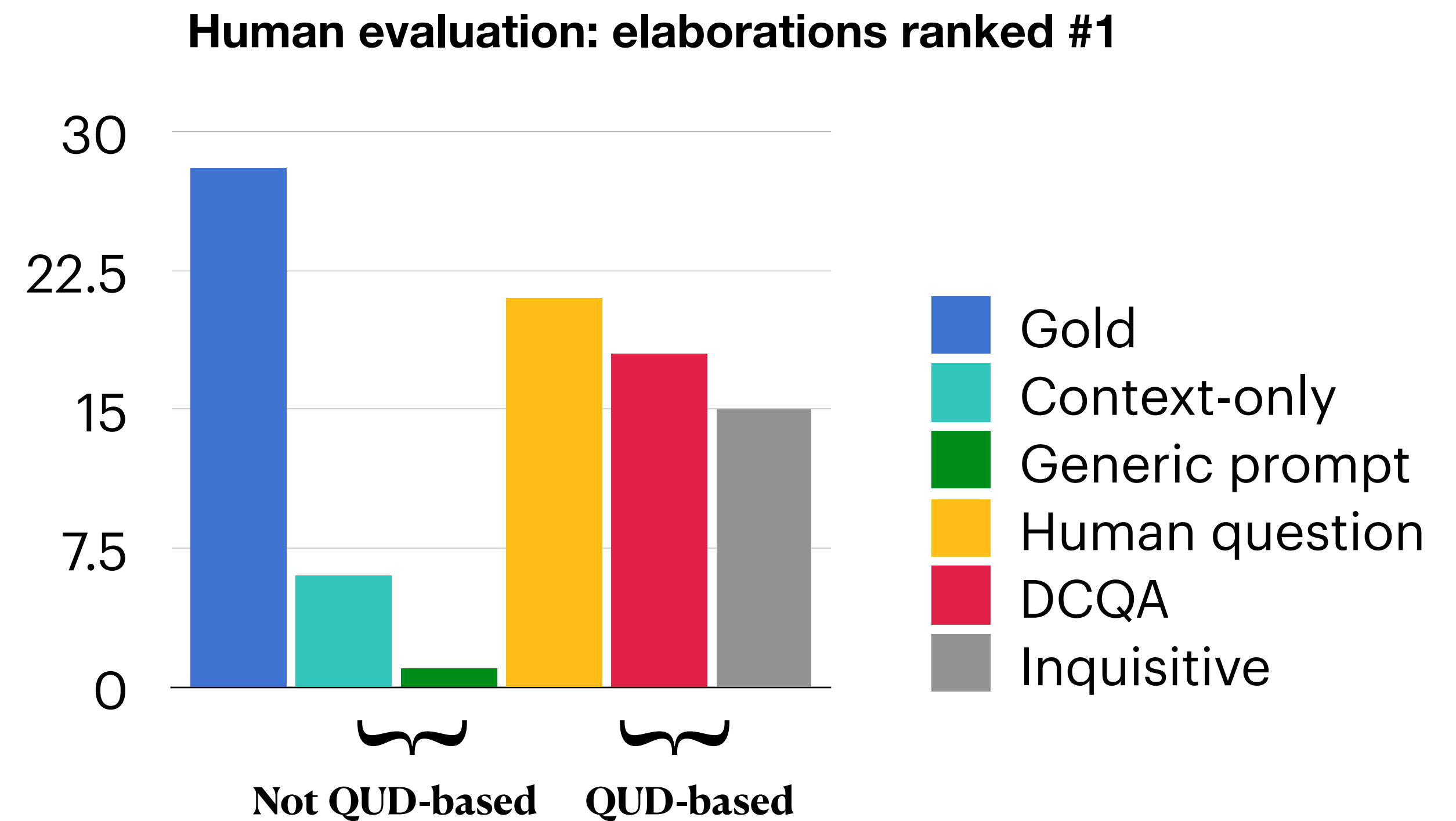
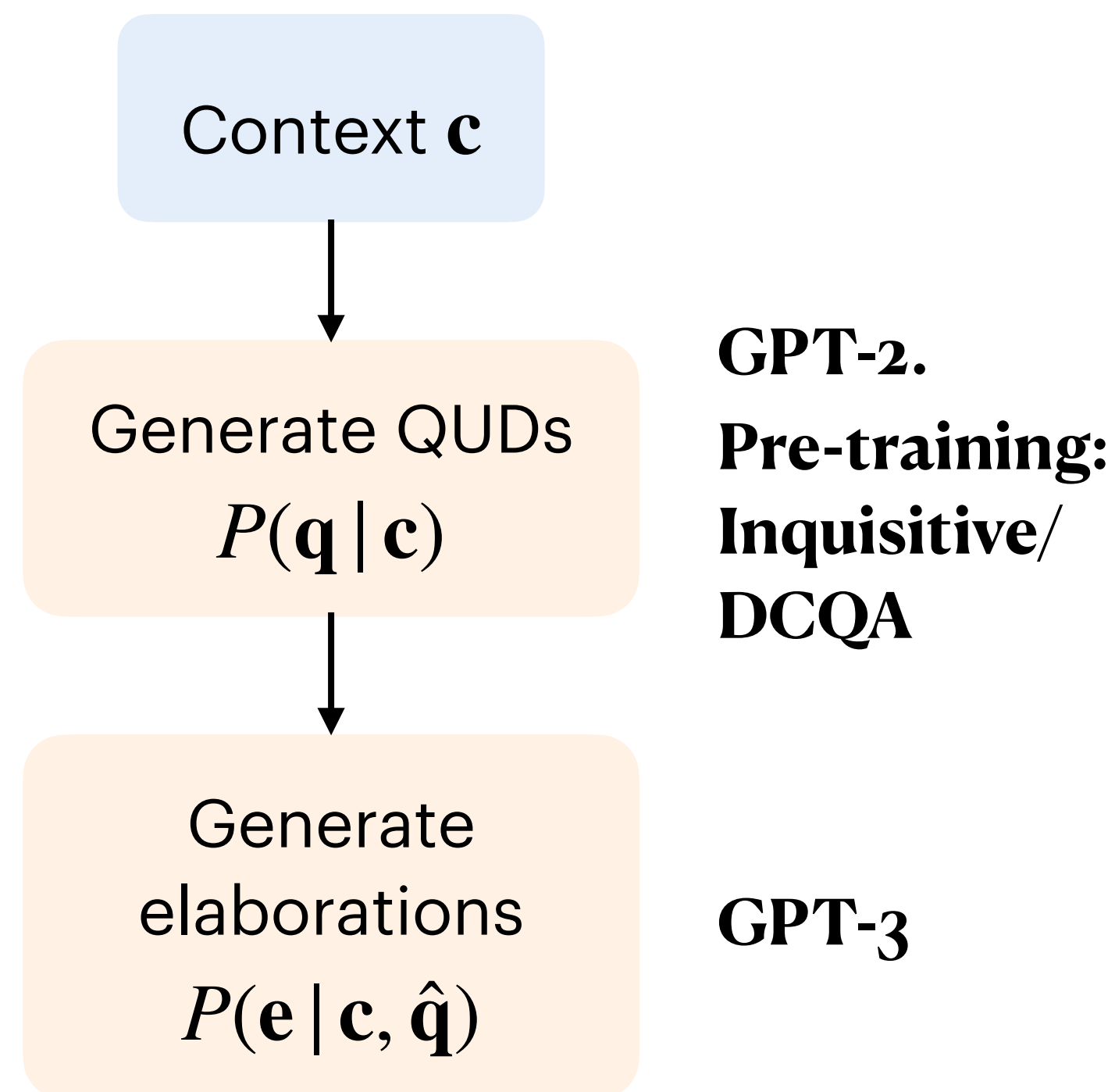
An artery is a tube that moves blood through the body.

Implicit QUD: What can the arteries tell us?

The artery could show if the person had been healthy or not.

- Data collection: QUDs for all 1.3K annotated elaborations from Newsela (*Srikanth & Li, Elaborative simplification: Content addition and explanation generation in text simplification, ACL Findings 2021*)

QUDs help elaborative simplification!



Information Loss in Simplification

InfoLossQA!

1 Text simplification

Lossy!

Original: In this study, the benefit of preoperative nutritional support was investigated for *non-small cell lung cancer patients* who underwent anatomic resection. [...] Patients who were malnourished, diabetic or who had undergone *bronchoplastic procedures or neoadjuvant therapy* were excluded from the study.

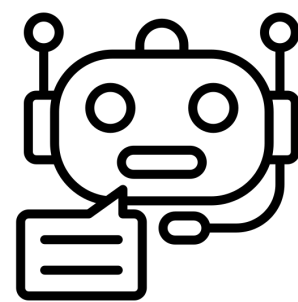
Simplification: This study looked at if eating a protein-rich diet before surgery could help *lung cancer patients* recover more quickly after surgery. [...] We didn't include patients who were already not eating well, had diabetes, or had received *other treatments* for their lung cancer.

- How do we help readers recover what's lost?
- How do we do this in a reader-centric way?

Information Loss in Simplification

InfoLossQA!

What kind of lung cancer do these patients have?

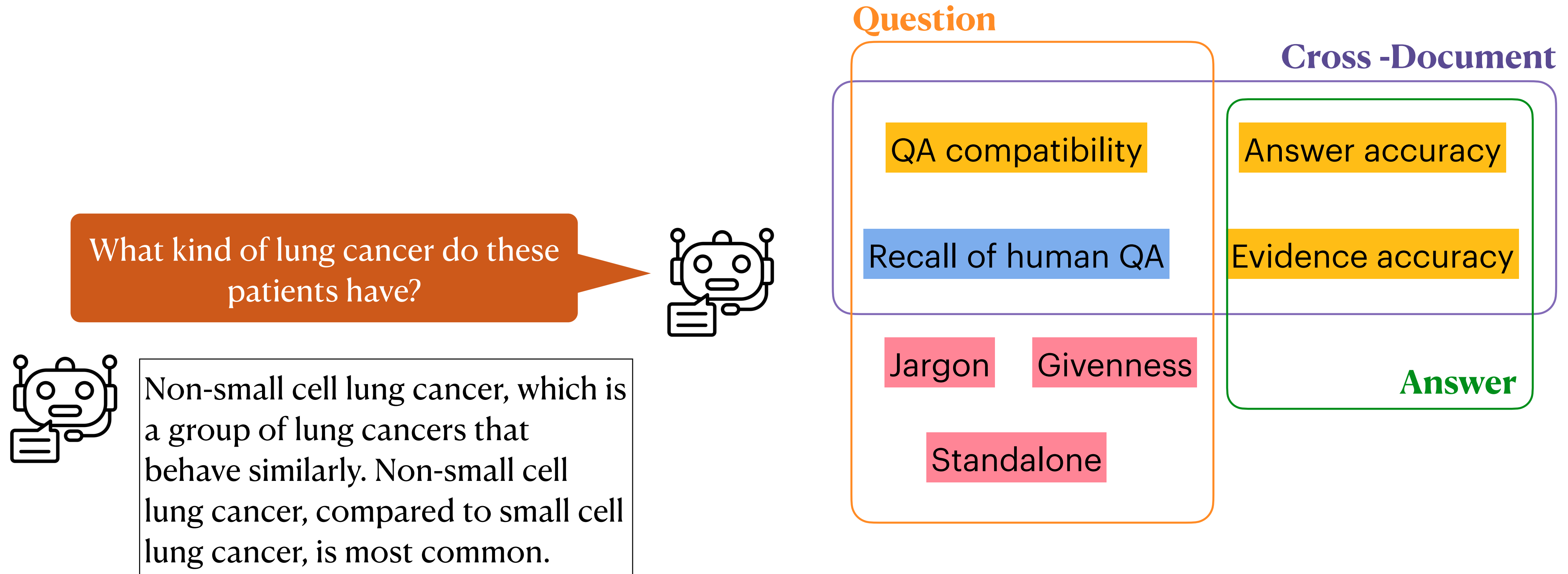


Non-small cell lung cancer, which is a group of lung cancers that behave similarly. Non-small cell lung cancer, compared to small cell lung cancer, is most common.

Simplification: This study looked at if eating a protein-rich diet before surgery could help *lung cancer patients* recover more quickly after surgery. [...] We didn't include patients who were already not eating well, had diabetes, or had received *other treatments* for their lung cancer.

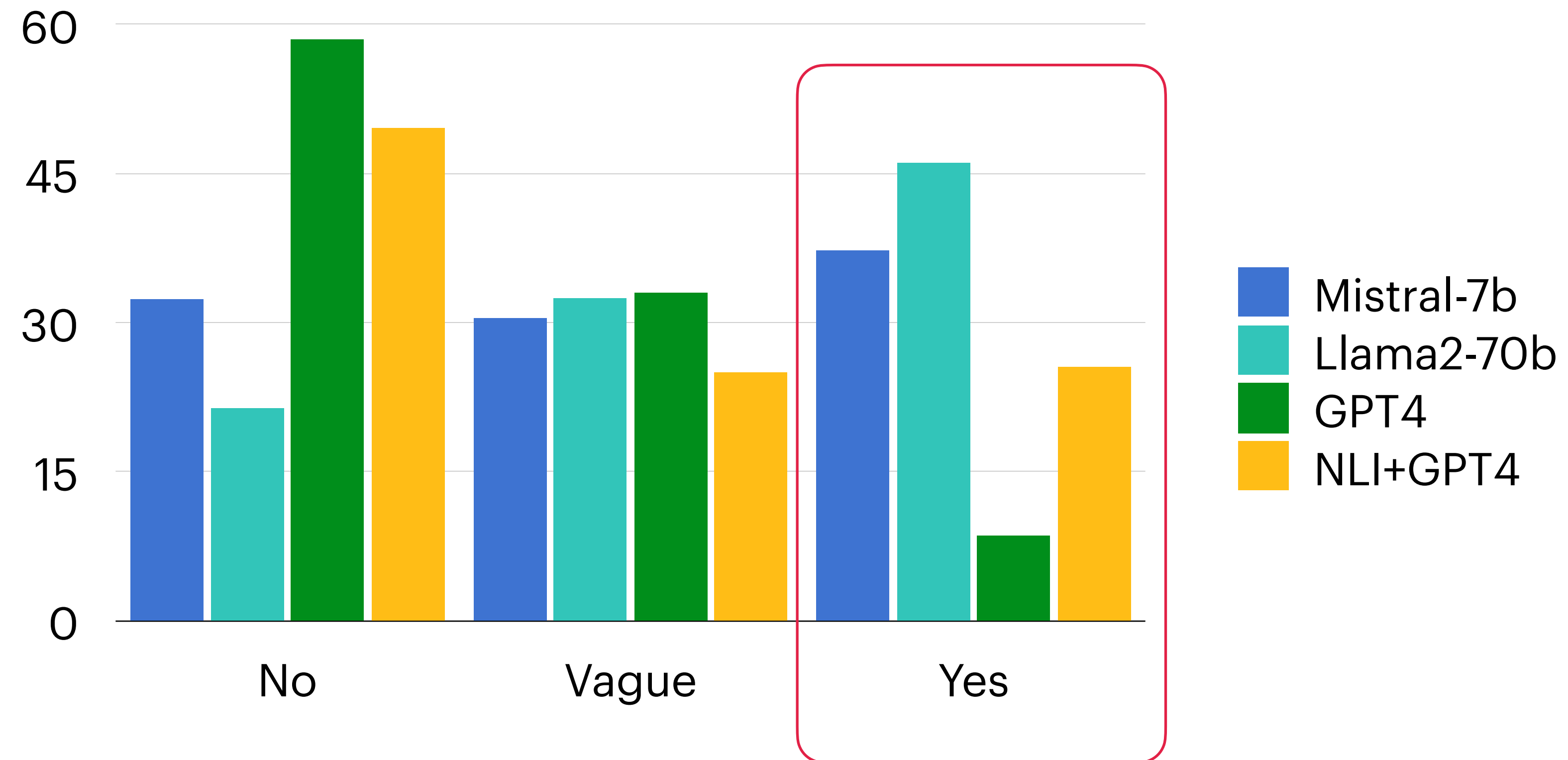
Information Loss in Simplification

InfoLossQA!



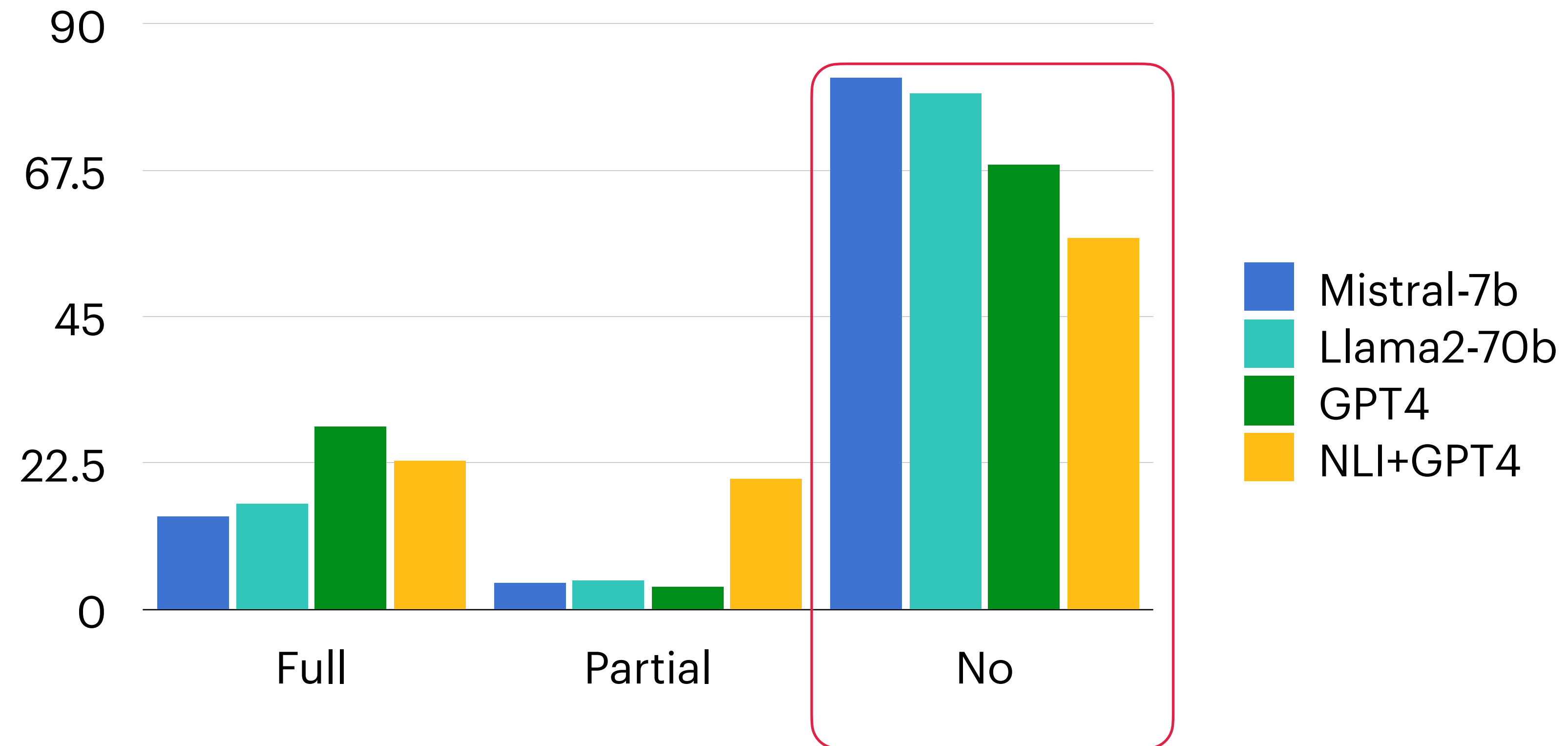
Information Loss in Simplification

LLMs still generate questions that are *answerable* by the target



Information Loss in Simplification

Models are not good at reproducing human-generated questions



Information Loss in Simplification

...More results

- LLMs, including GPT-4, are not very good at satisfying the Givenness criteria
- Neither are they good at always producing standalone questions
- Only GPT-4 can always do evidence grounding

QUD/Inquisitive QG has lots of applications!

- Characterizing factual differences between texts (Cole et al., 2023)
- Analyze diverse news perspectives (Laban et al., 2022)
- Expandable summarization (Fok et al., 2023)
- Asking follow-up questions (Meng et al., 2023)
- Decontextualization (Newman et al., 2023)
- Text planning (Narayan et al., 2023)
- Question generation in education (Cui et al., 2024)

Takeaways

- Questions Under Discussion: a generative discourse framework
 - Modeling discourse with questions and answers!
 - Salient inquisitive questions are more likely to become QUDs.
- QUD in generation tasks: new frontiers in text simplification

