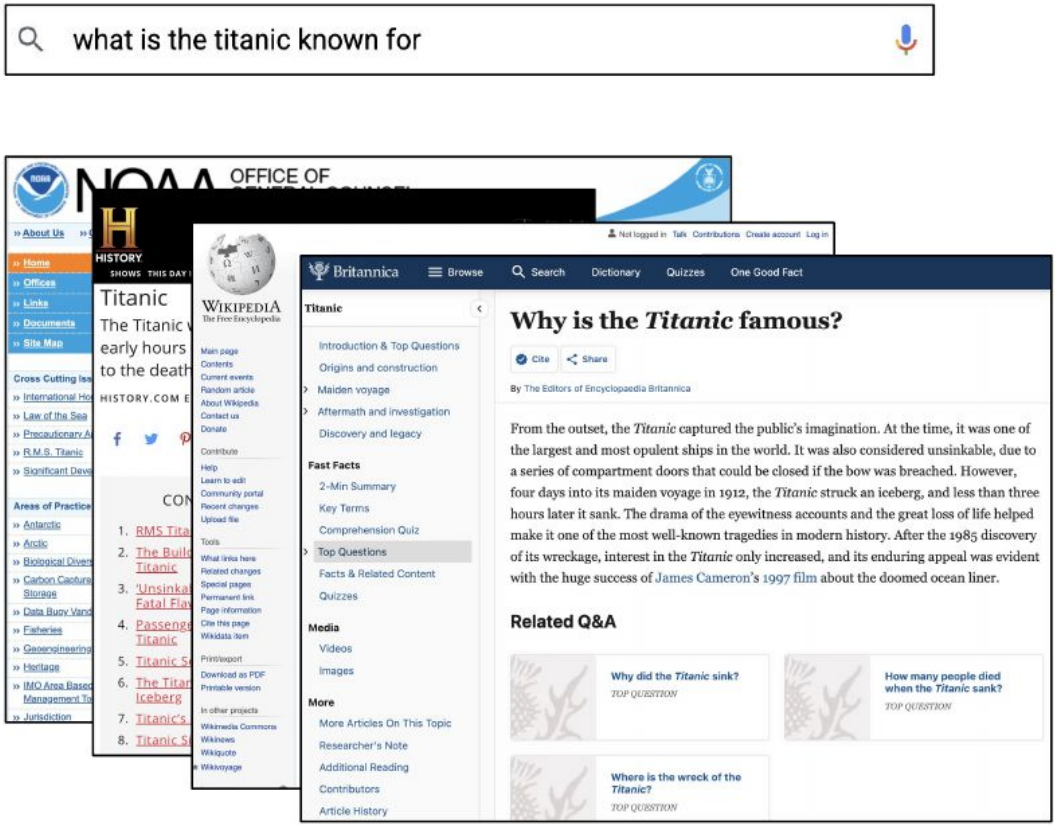
Conditional Generation with a Question-Answering Blueprint

Mirella Lapata
School of Informatics
University of Edinburgh

Conditional Generation: The Wish List

- Generate natural language towards a communicative goal
- which is faithful and can be attributed to is sources
- while users explicitly control generation outcome
- without being sensitive to the modality of the input

Conditional Generation: Long-form QA



- Some queries have only long-form answers
- based on input from multiple documents
- where hallucinations and attribution can be more problematic

Conditional Generation: Dialogue Summarization

The Justice League Recombination, The Big Bang Theory, S4.E11

Penny starts dating Zack again. Used to being the target of bullying, Leonard, Sheldon, Howard and Raj instead bully Zack through their implications to him that he's stupid. Not fully realizing what they're doing until Penny points it out to them, the guys decide to apologize to Zack in the only way they can. Accepting their apology, Zack decides to start hanging out with them, much to Penny's chagrin. This hanging out leads to the idea that Zack should replace Leonard as Superman in their group Justice League of America costume for the New Year's Eve party at the comic book store, with Penny the sixth wheel as Wonder Woman. These actions lead to questions from Leonard to Penny about why she's back together with Zack.

Conditional Generation: Visual Storytelling

Image Sequence:





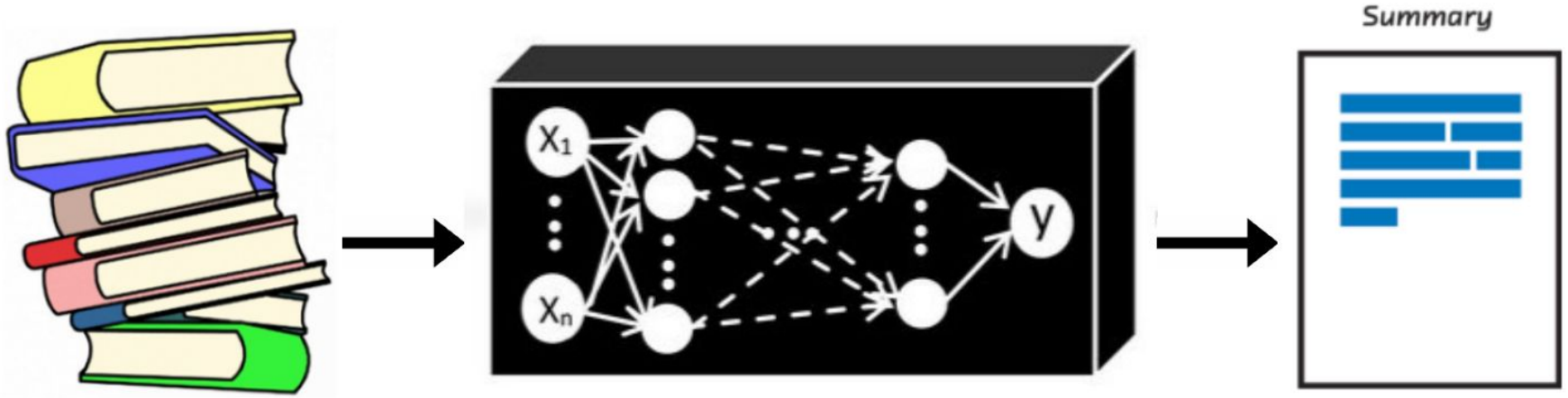
Gold-standard Story:

It was customary to give big brass a full tour. Of course, the mess hall would usually serve something a little better than average on "Big Brass" days. He looked like he was okay with it. After lunch the tour continued. He was introduced to the Commanding officers.

Story Generated by KE-VIST:

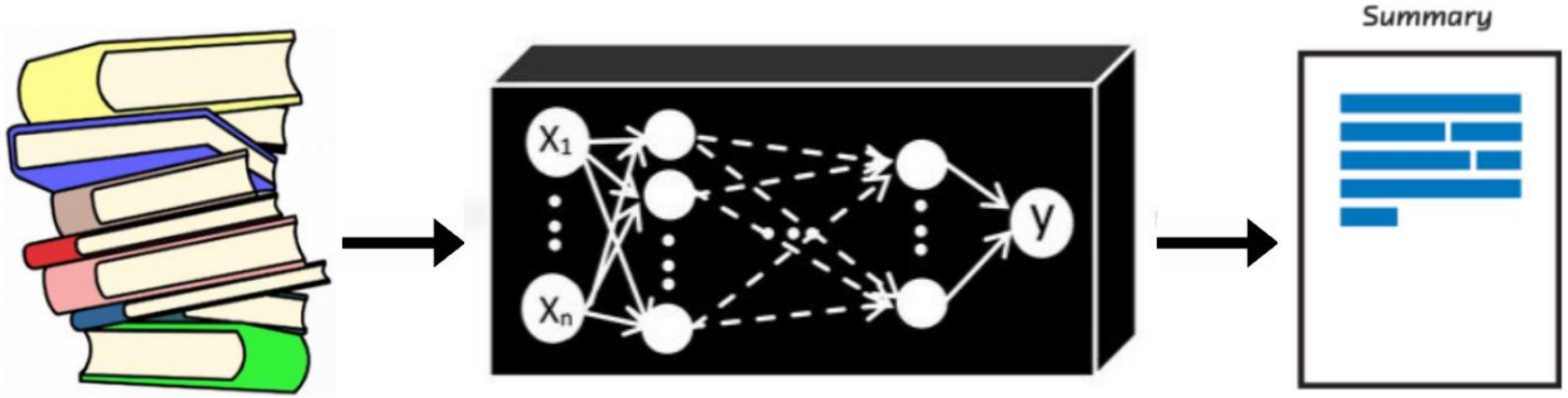
We all met up with the men. They walked a lot in the man walking. We had a lot of food. I gave a

The Blackbox Nature of Conditional Generation



- Change the way entities are represented (Puduppully et al., 2019; Iso et al., 2019)
- The decoder skips low-confidence tokens (Tian et al., 2019)
- Encode documents hierarchically (Rhode et al., 2021)
- Adopt sparse attention mechanisms (Child et al., 2019; Beltagy et al., 2020)

The Blackbox Nature of Conditional Generation



Introduce intermediate planning components (Wiseman et al., 2018; Moryossef et al., 2019; Puduppully et al., 2022; Narayan et al., 2022)

The QUD Model of Discourse

Questions under Discussion (QUD) theory of discourse structure (Carlson, 1983; Ginzburg, 1994; Van Kuppevelt, 1995; Larson, 2002; Roberts, 2012; Riester, 2019).

A **QUD** is a partially structured set of questions which discourse participants are mutually committed to resolving at a given point in time.

A **question** is a set of alternative possibilities (corresponding in some way to the possible answers to the question).

Discourse contains **implicit questions** for each of the assertions made which are turned into answers.

Question-Answering Blueprints as Content Plans

what is the titanic known for

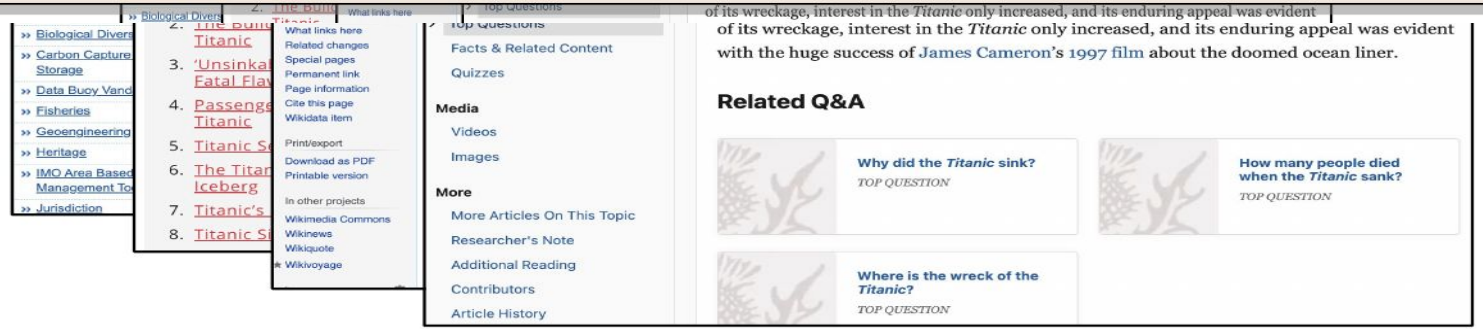
What kind of ship was the RMS Titanic? British passenger liner

What line operated the RMS Titanic? the White Star Line

How many ships did the White Star Line make? three ships

In what ocean did the RMS Titanic sink? Pacific Ocean

When did the RMS Titanic hit an iceberg? On its maiden voyage



Question-Answering Blueprints as Content Plans

what is the titanic known for

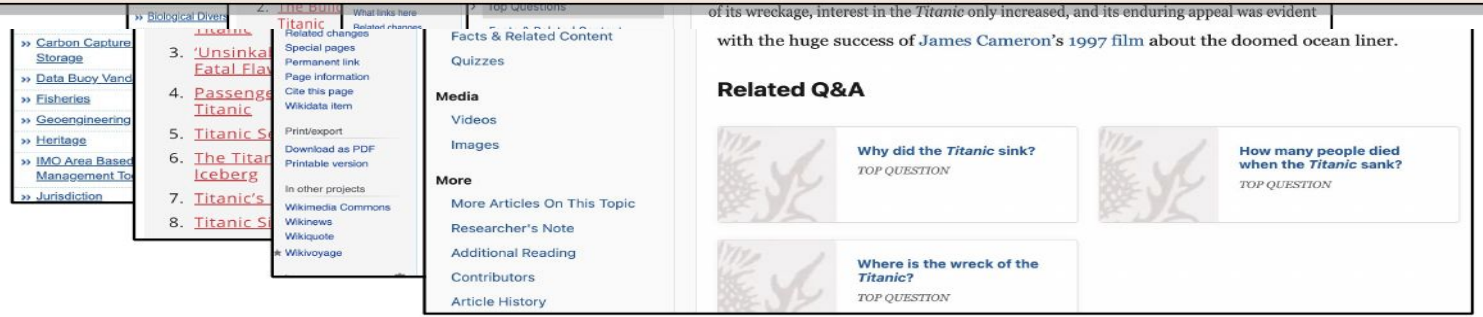
What kind of ship was the RMS Titanic? British passenger liner

What line operated the RMS Titanic? the White Star Line

How many ships did the White Star Line make? three ships

In what ocean did the RMS Titanic sink? Pacific Ocean

When did the RMS Titanic hit an iceberg? On its maiden voyage



Question-Answering Blueprints as Content Plans

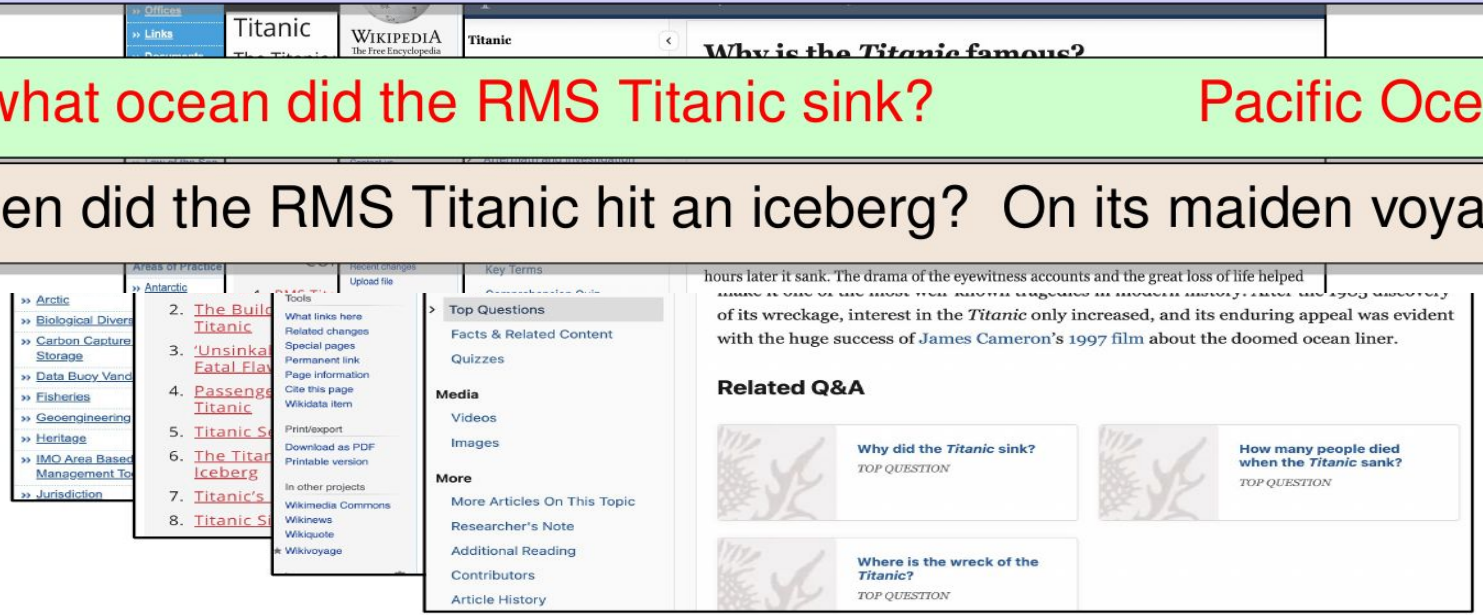
what is the titanic known for

What kind of ship was the RMS Titanic? British passenger liner

What line operated the RMS Titanic? the White Star Line

In what ocean did the RMS Titanic sink? Pacific Ocean

When did the RMS Titanic hit an iceberg? On its maiden voyage



Question-Answering Blueprints as Content Plans

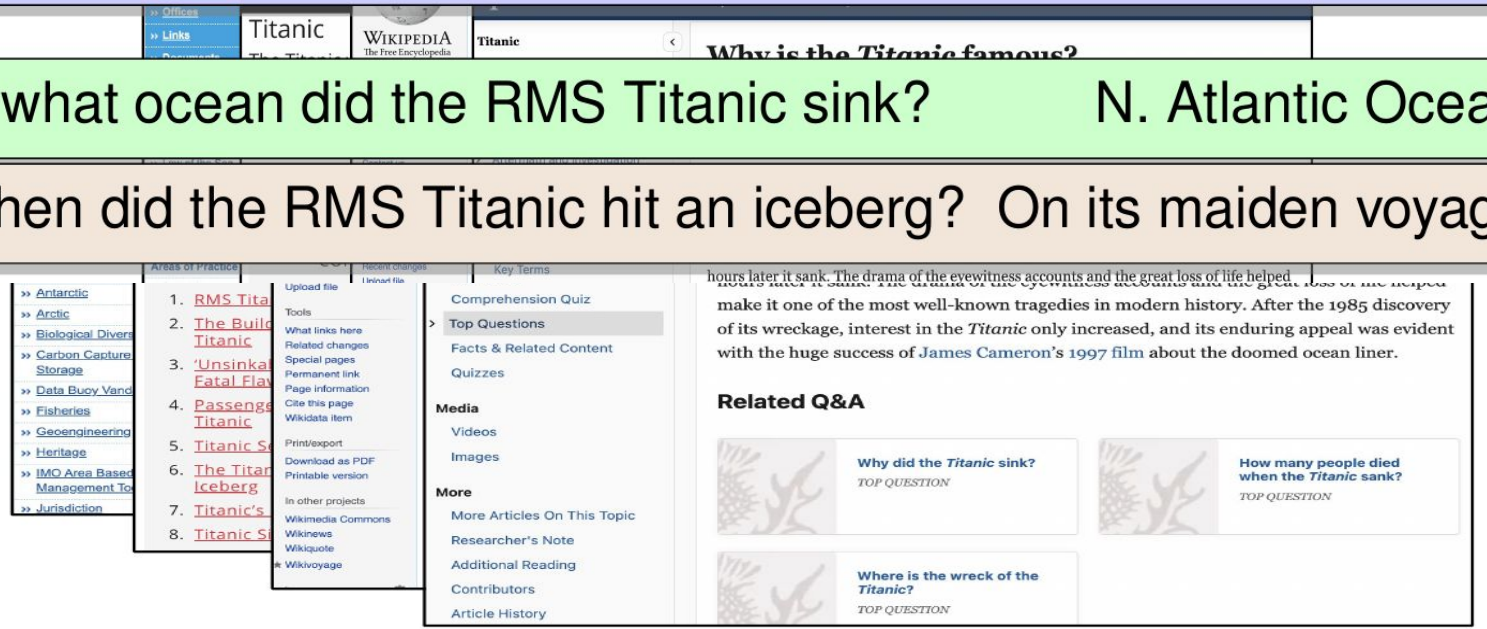
what is the titanic known for

What kind of ship was the RMS Titanic? British passenger liner

What line operated the RMS Titanic? the White Star Line

In what ocean did the RMS Titanic sink? N. Atlantic Ocean

When did the RMS Titanic hit an iceberg? On its maiden voyage



Question-Answering Blueprints as Content Plans

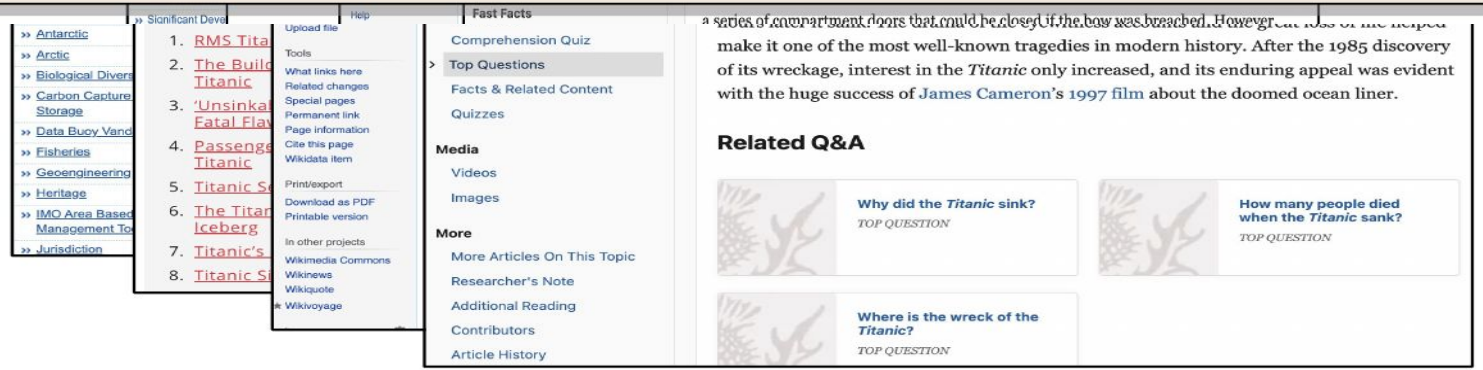
what is the titanic known for

What kind of ship was the RMS Titanic? British passenger liner

What line operated the RMS Titanic? the White Star Line

In what ocean did the RMS Titanic sink? N. Atlantic Ocean

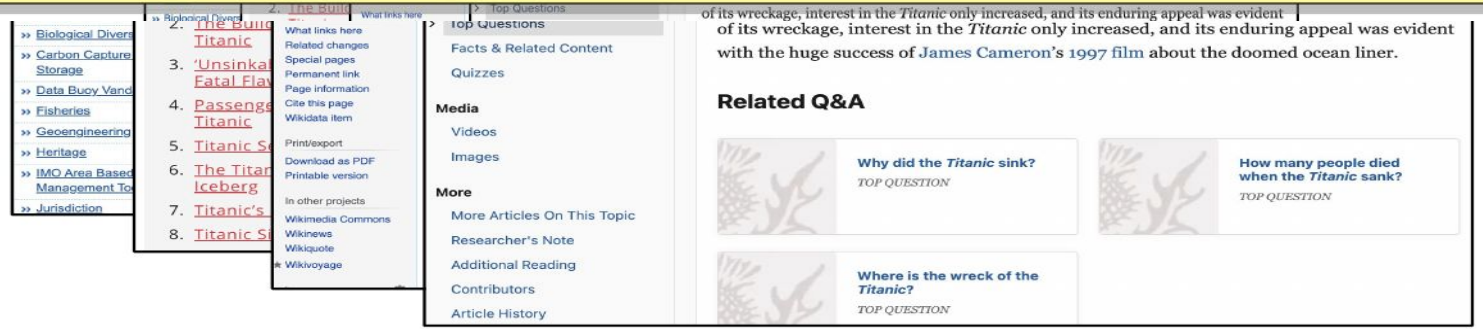
When did the RMS Titanic hit an iceberg? On its maiden voyage



Question-Answering Blueprints as Content Plans

what is the titanic known for

What kind of ship was the RMS Titanic?	British passenger liner
What line operated the RMS Titanic?	the White Star Line
In what ocean did the RMS Titanic sink?	N. Atlantic Ocean
When did the RMS Titanic hit an iceberg?	On its maiden voyage
When did the DiCaprio movie come out?	1997



Question-Answering Blueprints as Content Plans

what is the titanic known for

What kind of ship was the RMS Titanic?	British passenger liner
What line operated the RMS Titanic?	the White Star Line
In what ocean did the RMS Titanic sink?	N. Atlantic Ocean
When did the RMS Titanic hit an iceberg?	On its maiden voyage
When did the DiCaprio movie come out?	1997

RMS Titanic was a British passenger liner, operated by the White Star Line. It sank in the North Atlantic Ocean after striking an iceberg during its maiden voyage. The 1997 film Titanic starring Leonardo Di Caprio has popularized the disaster and its aftermath.

Question-Answering Blueprints as Content Plans

- Blueprints as intermediate (discrete) planning stage for conditional generation
- Reduce faithfulness errors and increase controllability
- Are better for long-form inputs and outputs

Where do the blueprints come from?

Datasets as we know them consist of document(s)-output pairs!

Training Data Annotation with Blueprints

Shelby Mustang

19 languages

Article Talk | Read Edit View history Tools

From Wikipedia, the free encyclopedia

The **Shelby Mustang** is a high-performance variant of the Ford Mustang built by Shelby American from 1965 to 1967 and by the Ford Motor Company from 1968 to 1970.

In 2005, Ford revived the Shelby nameplate for a high-performance model of the fifth-generation Ford Mustang.[1]

1st generation (1965–1970) [edit]

1965–1966 Shelby GT350 [edit]

The 1965–1966 cars were the smallest and lightest of the GT 350 models. Production started when 289 AC Cobra production was wrapping up and 427 AC Cobra production was beginning, both of which were Ford-powered AC-based two-seat sports car also produced by Shelby American during the same period. Both models use the Cobra emblem, similar paint scheme, and the optional "Cobra" valve covers on many GT350s that were part of a marketing tie-in by Shelby, as well as one of his iconic symbols. All 1965–66 cars had the Windsor 289 cu in (4.7 L) HiPo K-Code 271 hp (202 kW; 275 PS) V8 engine, modified with a large 4-barrel Holley 715 CFM carburetor to produce 306 bhp (310 PS; 228 kW) at 6,000 rpm and

Shelby Mustang

1968 Shelby Cobra GT350

Overview

Manufacturer	Ford Performance Shelby SVT
Also called	Ford Mustang Shelby Ford Shelby Mustang Mustang Shelby Mustang GT500

The Shelby Mustang is a high performance variant of the Ford Mustang which was built by Shelby American from 1965 to 1968, and from 1969 to 1970 by Ford. Following the introduction of the fifth generation Ford Mustang in 2005, the Shelby nameplate was revived as a new high-performance model, this time designed and built by Ford.

Training Data Annotation with Blueprints

The Shelby Mustang is a high performance variant of the Ford Mustang which was built by Shelby American from 1965 to 1968, and from 1969 to 1970 by Ford. Following the introduction of the fifth generation Ford Mustang in 2005, the Shelby nameplate was revived as a new high-performance model, this time designed and built by Ford.

Question-Answering Blueprints as Content Plans

The Shelby Mustang is a high performance variant of the Ford Mustang which was built by Shelby American from 1965 to 1968, and from 1969 to 1970 by Ford. Following the introduction of the fifth generation Ford Mustang in 2005, the Shelby nameplate was revived as a new high-performance model, this time designed and built by Ford.

What is a high performance variant of the Ford Mustang?	Shelby
Who built the Shelby Mustang from 1969 to 1970?	Ford
Q2: What is the high performance variant of the Ford Mustang called?	Shelby
What is a high performance variant of the Ford Mustang?	Shelby Mustang
What is a high performance variant of the Ford Mustang?	Shelby
What is a Shelby Mustang?	A high performance variant
The Shelby Mustang is a high performance variant of what?	Ford Mustang
During what years was the Shelby Mustang built by Shelby American?	1965 to 1968
Who built the Shelby Mustang from 1965 to 1968?	Shelby American
In what year was the fifth generation of the for Mustang introduced?	2005
What generation of Mustang was introduced in 2005?	the fifth generation
What was the Shelby Mustang revived as?	a new high-performance model

Question-Answering Blueprints as Content Plans

The Shelby Mustang is a high performance variant of the Ford Mustang which was built by Shelby American from 1965 to 1968, and from 1969 to 1970 by Ford. Following the introduction of the fifth generation Ford Mustang in 2005, the Shelby nameplate was revived as a new high-performance model, this time designed and built by Ford.

What is a high performance variant of the Ford Mustang?	Shelby
Who built the Shelby Mustang from 1969 to 1970?	Ford
Q2: What is the high performance variant of the Ford Mustang called?	Shelby
What is a high performance variant of the Ford Mustang?	Shelby Mustang
What is a high performance variant of the Ford Mustang?	Shelby
What is a Shelby Mustang?	A high performance variant
The Shelby Mustang is a high performance variant of what?	Ford Mustang
During what years was the Shelby Mustang built by Shelby American?	1965 to 1968
Who built the Shelby Mustang from 1965 to 1968?	Shelby American
In what year was the fifth generation of the for Mustang introduced?	2005
What generation of Mustang was introduced in 2005?	the fifth generation
What was the Shelby Mustang revived as?	a new high-performance model

Question-Answering Blueprints as Content Plans

The Shelby Mustang is a high performance variant of the Ford Mustang which was built by Shelby American from 1965 to 1968, and from 1969 to 1970 by Ford. Following the introduction of the fifth generation Ford Mustang in 2005, the Shelby nameplate was revived as a new high-performance model, this time designed and built by Ford.

Who built the Shelby Mustang from 1969 to 1970?	Ford
During what years was the Shelby Mustang built by Shelby American?	1965 to 1968
In what year was the fifth generation of the for Mustang introduced?	2005
What was the Shelby Mustang revived as?	a new high-performance model

Question-Answering Blueprints as Content Plans

Shelby Mustang

19 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

The **Shelby Mustang** is a high-performance variant of the Ford Mustang built by Shelby American from 1965 to 1967 and by the Ford Motor Company from 1968 to 1970.

In 2005, Ford revived the Shelby nameplate for a high-performance model of the fifth-generation Ford Mustang.[1]

Shelby Mustang

1st generation (1965–1970) [edit]

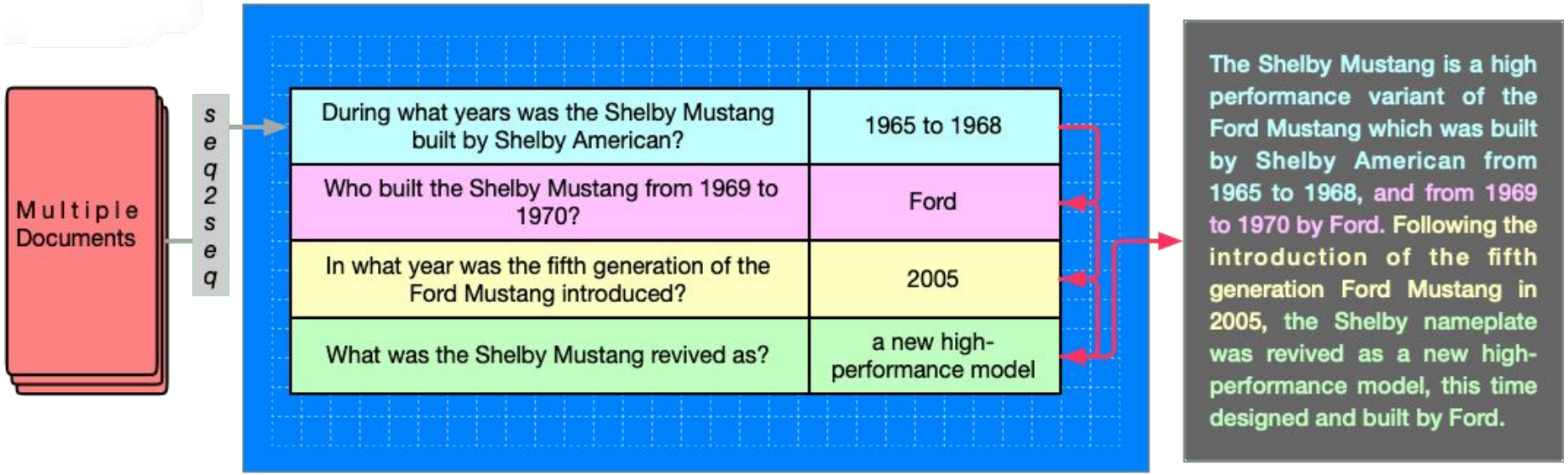
1965–1966 Shelby GT350 [edit]

The 1965–1966 cars were the smallest and lightest of the GT 350 models. Production started when 289 AC Cobra production was wrapping up and 427 AC Cobra production was beginning, both of which were Ford-powered AC-based two-seat sports car also produced by Shelby American during the same period. Both models use the Cobra emblem, similar paint scheme, and the optional "Cobra" valv covers on many GT350s that were part of a marketing tie-in by Shelby, as well as one of his iconic symbols. All 1965–66 cars had the Windsor 289 cu in (4.7 L) HiPc K-Code 271 hp (202 kW; 275 PS) V8 engine, modified with a large 4-barrel Holley

Who built the Shelby Mustang from 1969 to 1970?	Ford
During what years was the Shelby Mustang built by Shelby American?	1965 to 1968
In what year was the fifth generation of the for Mustang introduced?	2005
What was the Shelby Mustang revived as?	a new high-performance model

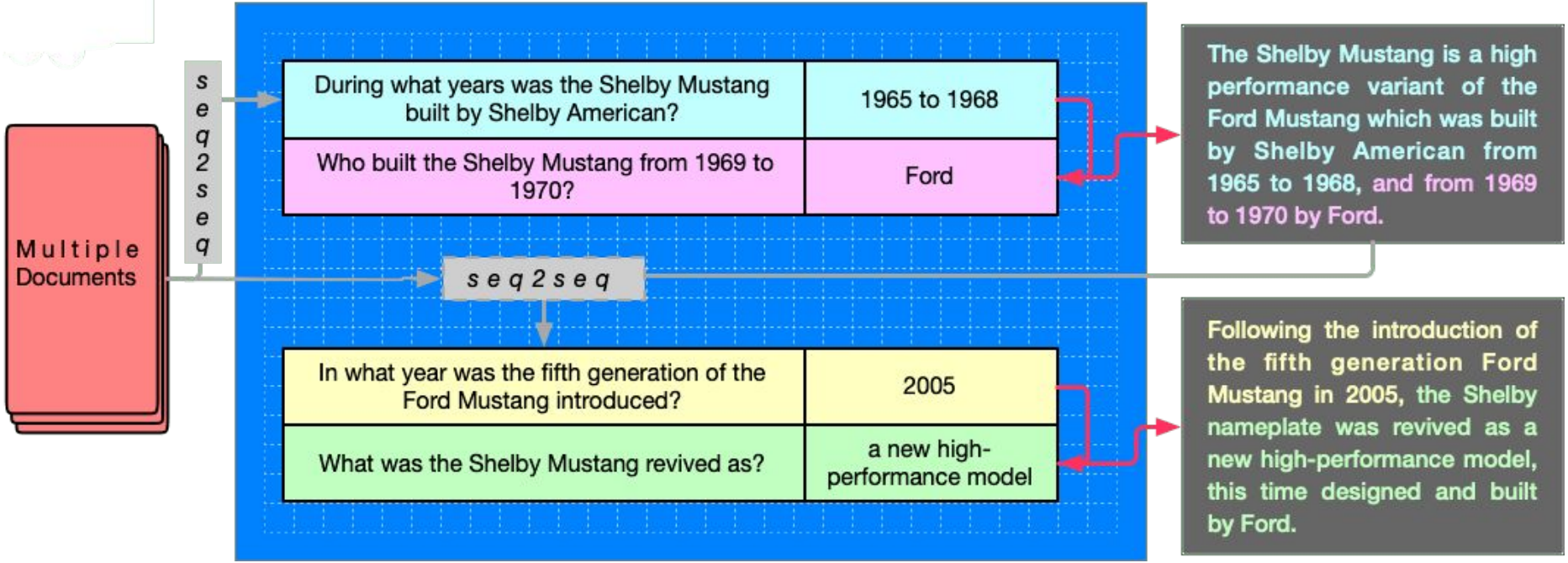
The Shelby Mustang is a high performance variant of the Ford Mustang which was built by Shelby American from 1965 to 1968, and from 1969 to 1970 by Ford. Following the introduction of the fifth generation Ford Mustang in 2005, the Shelby nameplate was revived as a new high-performance model, this time designed and built by Ford.

End-to-End Blueprint Model



Generates blueprint and output sequence in one go.

Iterative Blueprint Model



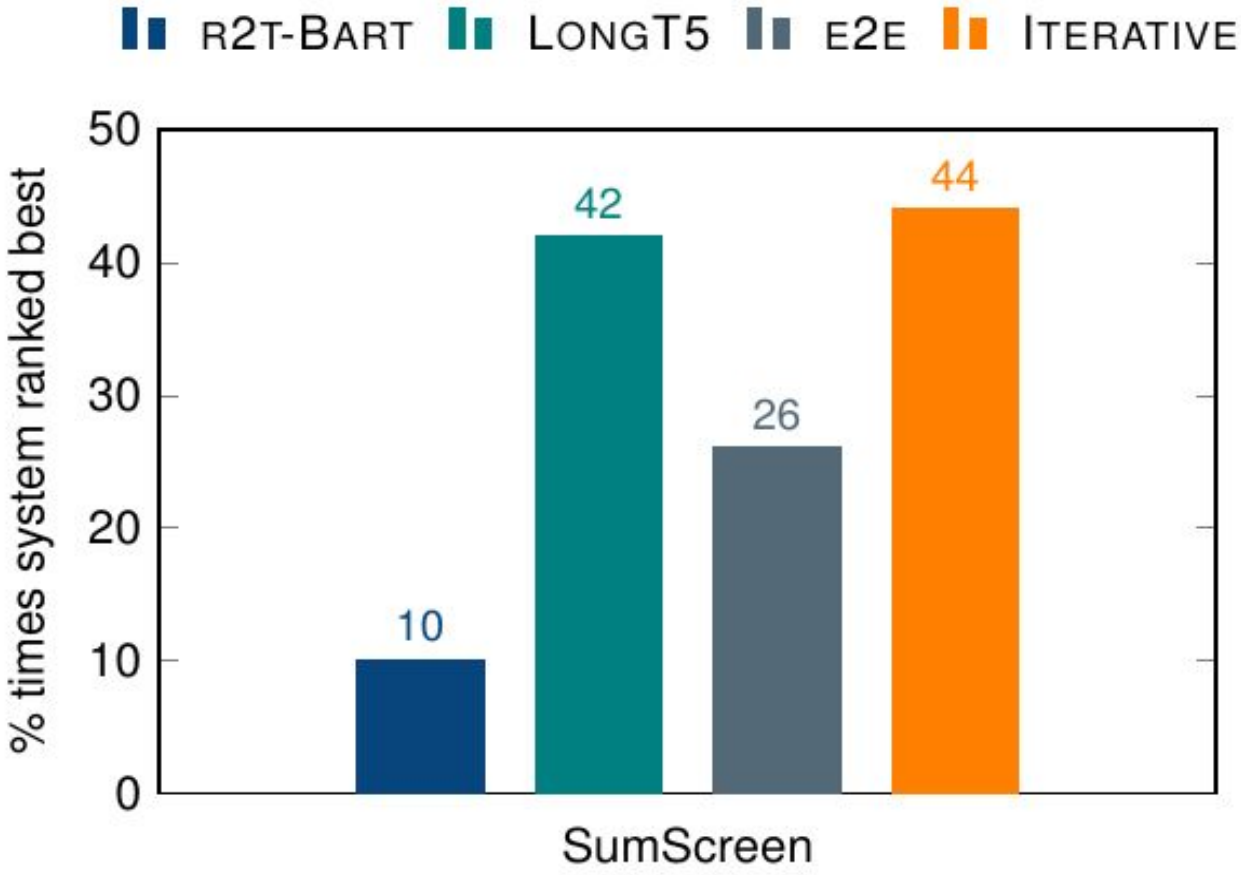
Iteratively plans and generates one sentence at a time.

Experimental Setting

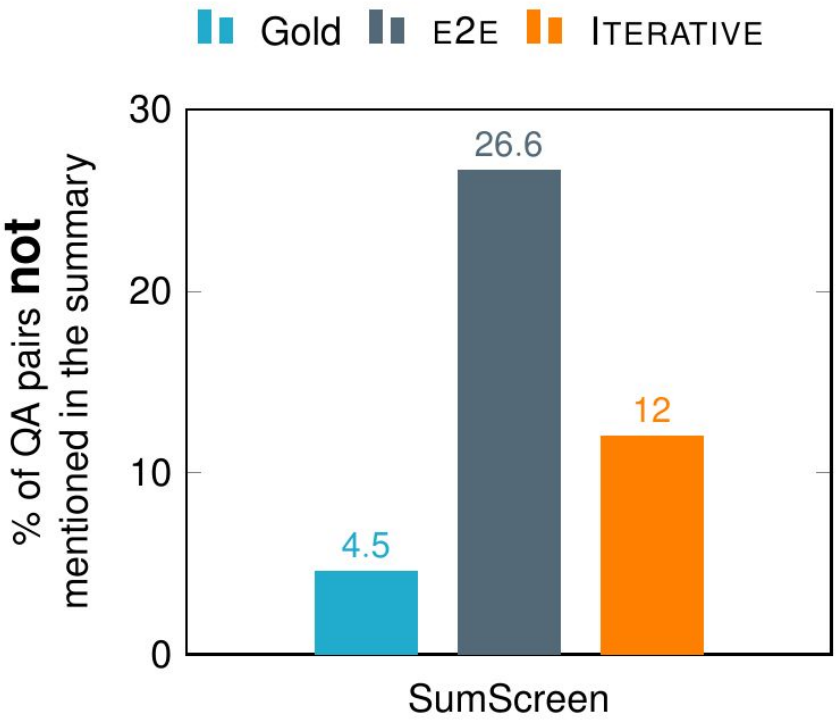
- **AQuaMuse** (Kulkarni et al., 2002, 2021): long-form question answering, simulates search engine, answer based on multiple retrieved documents.
- **WikiCatSum** (Perez-Beltrachini et al., 2019): topic-focused multi-document summarization, generate Wikipedia abstracts.
- **SummScreen** (Chen et al., 2022): dialogue summarization, generate summaries of TV episodes (e.g., CSI, The Bing Bang Theory).

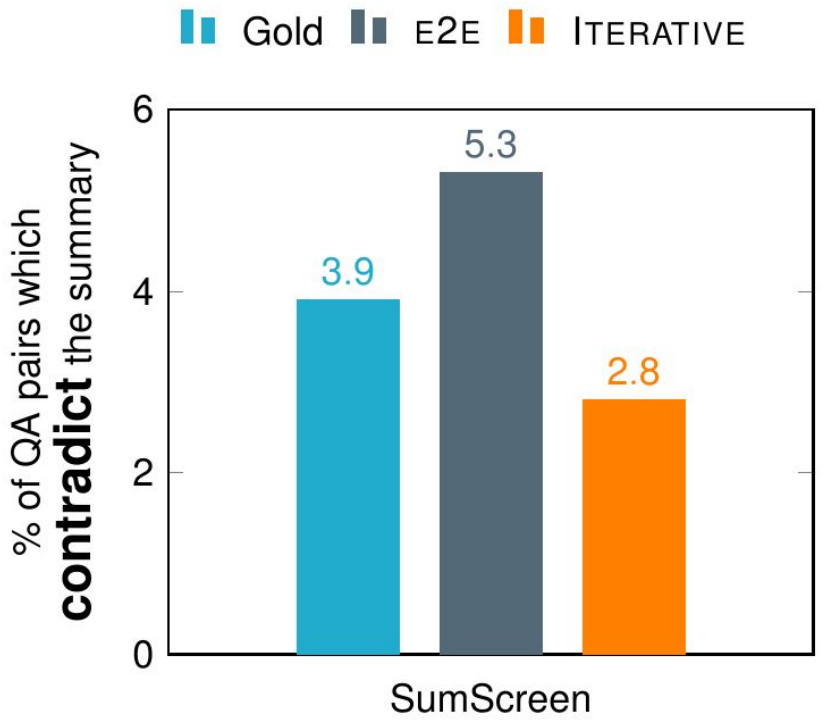
All experiments with a LongT5 (Guo et al., 2021) backbone model.

Evaluation: How Good are the Blueprint Summaries?



Evaluation: Are Summaries Grounded to their Blueprint?

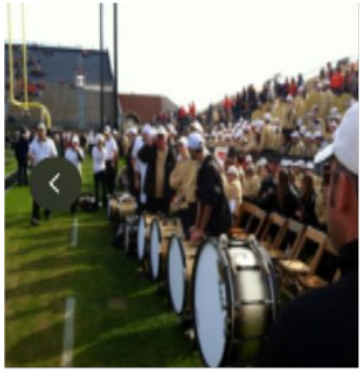
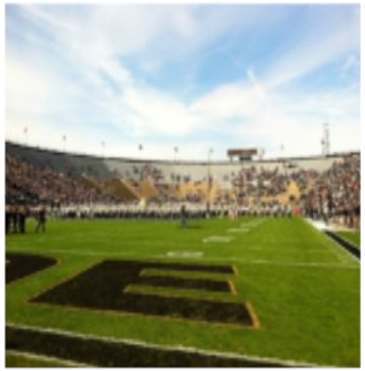


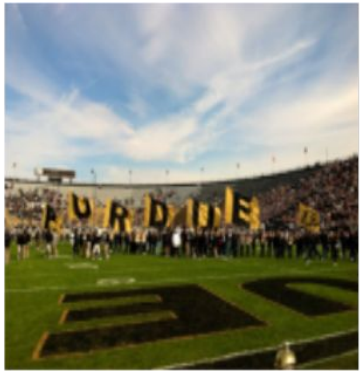


Interim Conclusions

- Bleuprint models are more informative than vanilla seq2seq models
- Errors can be examined and traced back to the blueprint
- QA pairs are intuitive and user-friendly, human-in-the-loop generation
- How do blueprints work across tasks and modalities?

Visual Story Telling: Blueprint Annotation

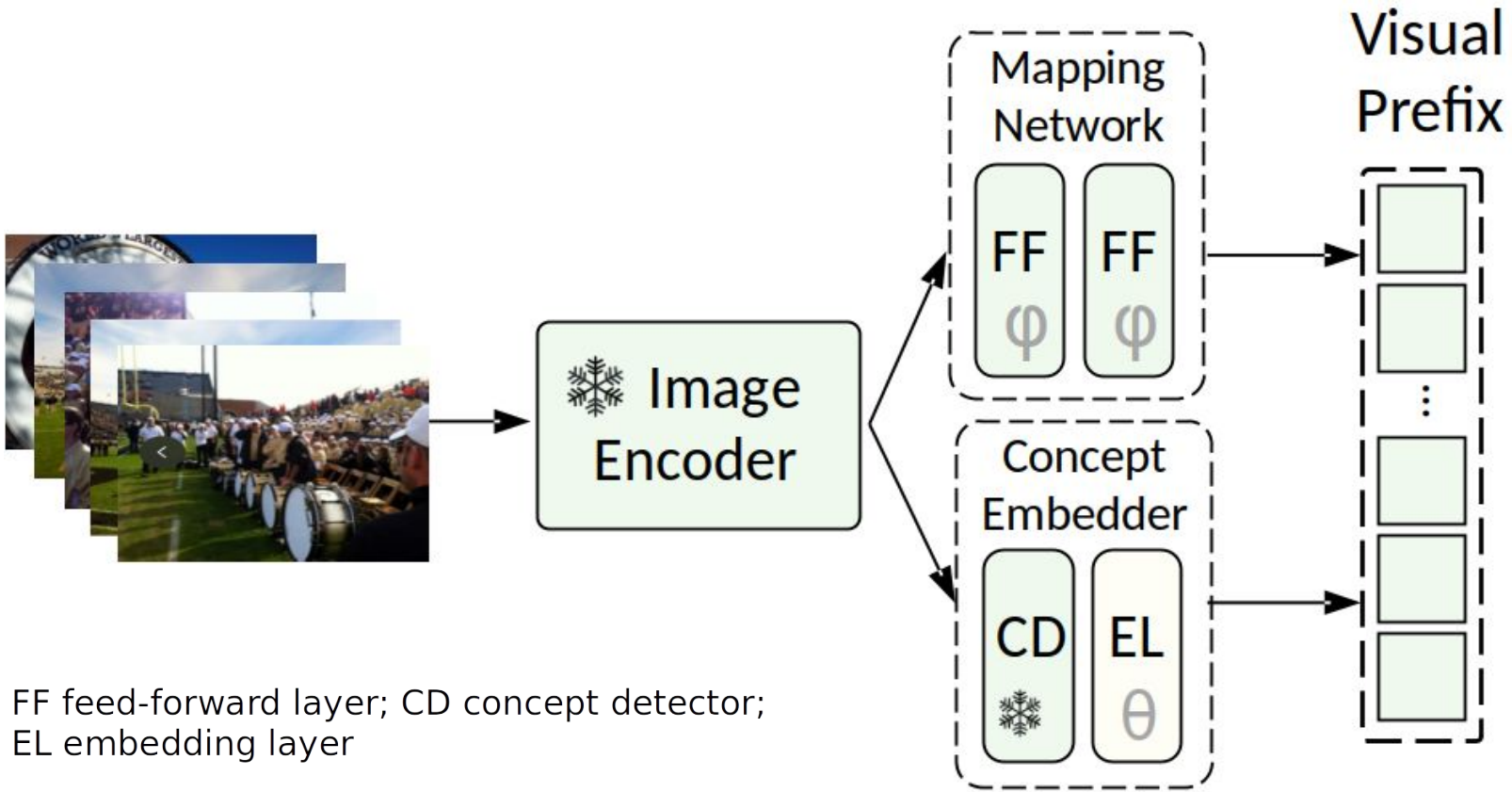






My Uncle Jack made us go to the Purdue game on Saturday. He went to Purdue and he thinks they are great no matter what else anyone thinks. We all had to sit in the bleachers and wear some ridiculous brown coats...

Q_0: Where did my uncle Jack made us go on Saturday?
A_0: the Purdue game
Q_1: When did my uncle Jack made us go to the Purdue game?
A_1: Saturday
Q_2: Where did we all sit?
A_2: the bleachers
Q_3: What did the fans of Purdue have to wear?
A_3: some ridiculous brown coats

Visual Prefixes

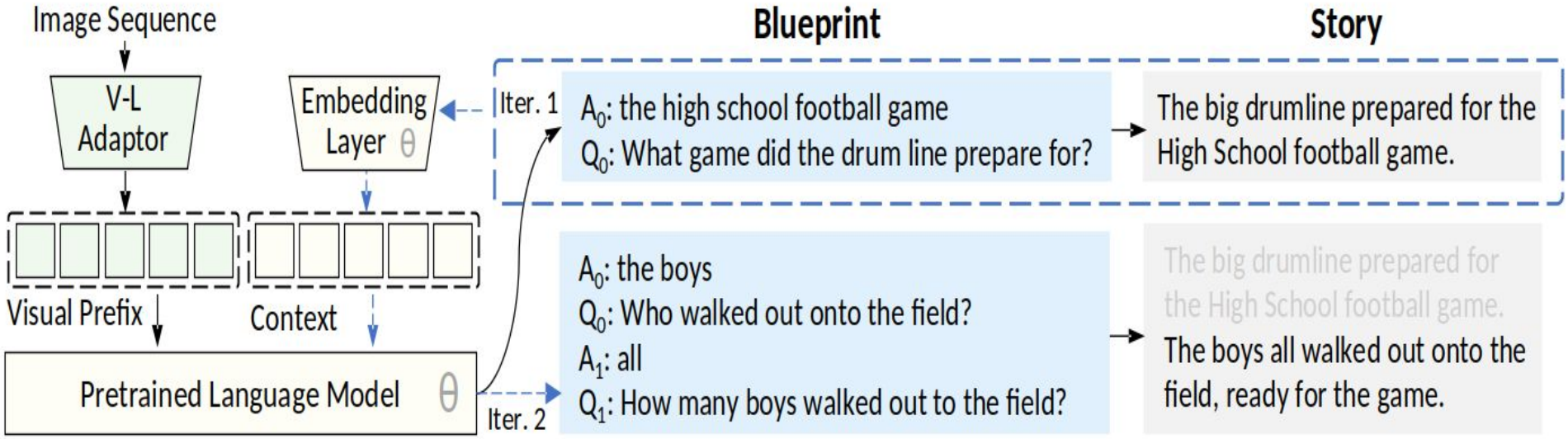


FF feed-forward layer; CD concept detector;
EL embedding layer

Blueprint Storytelling Model (Iterative)

- Use a generation model (based on T5) to turn VIST stories into blueprints
- Train a linear model to map image sequences into visual prefixes (based on ResNet-152)
- Fine-tune a pretrained model to generate blueprints and stories from visual prefixes
- Generation proceeds sentence-by-sentence, where previous sentences are context for the current one: iterative model
- An end-to-end model that generates all sentences at once from the visual prefix performs worse

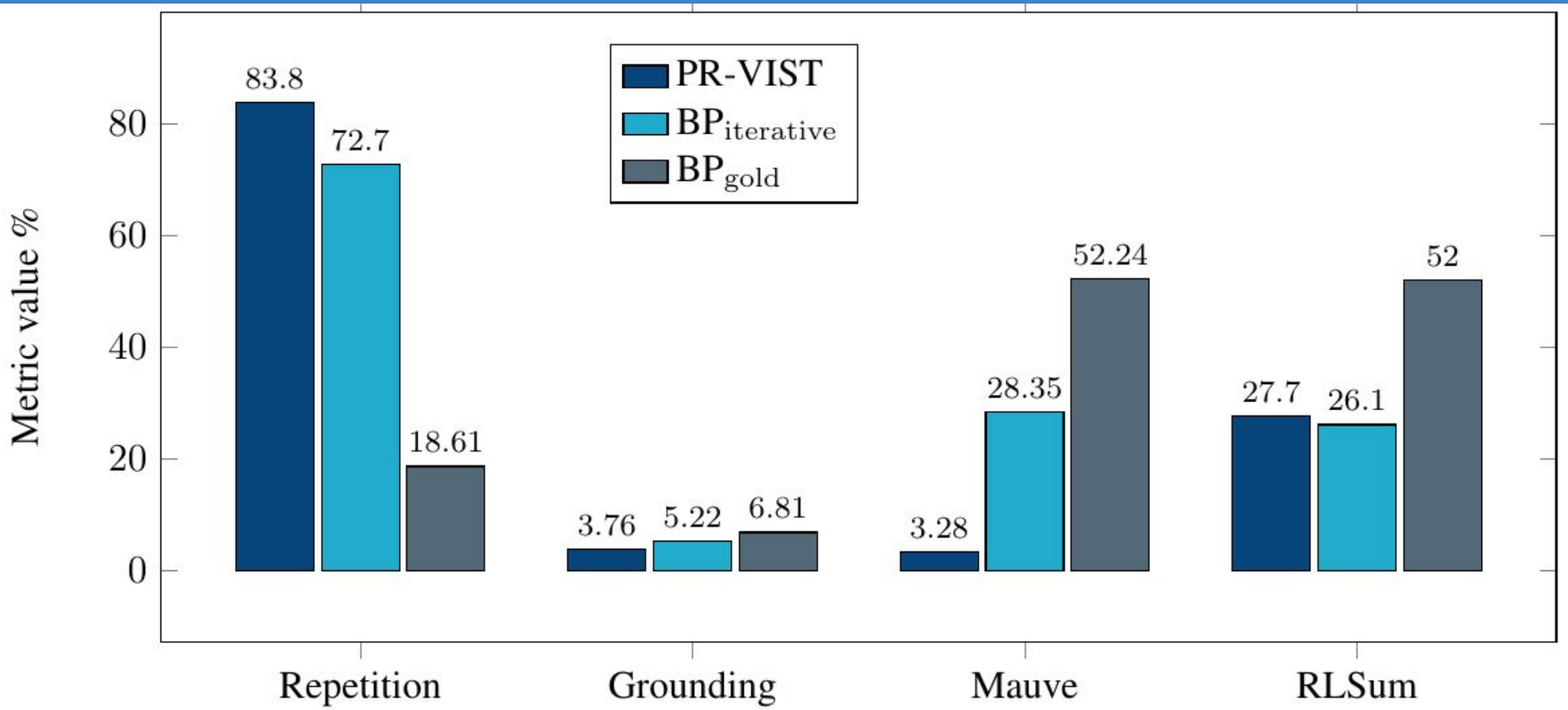
Iterative Blueprint Model



Evaluation

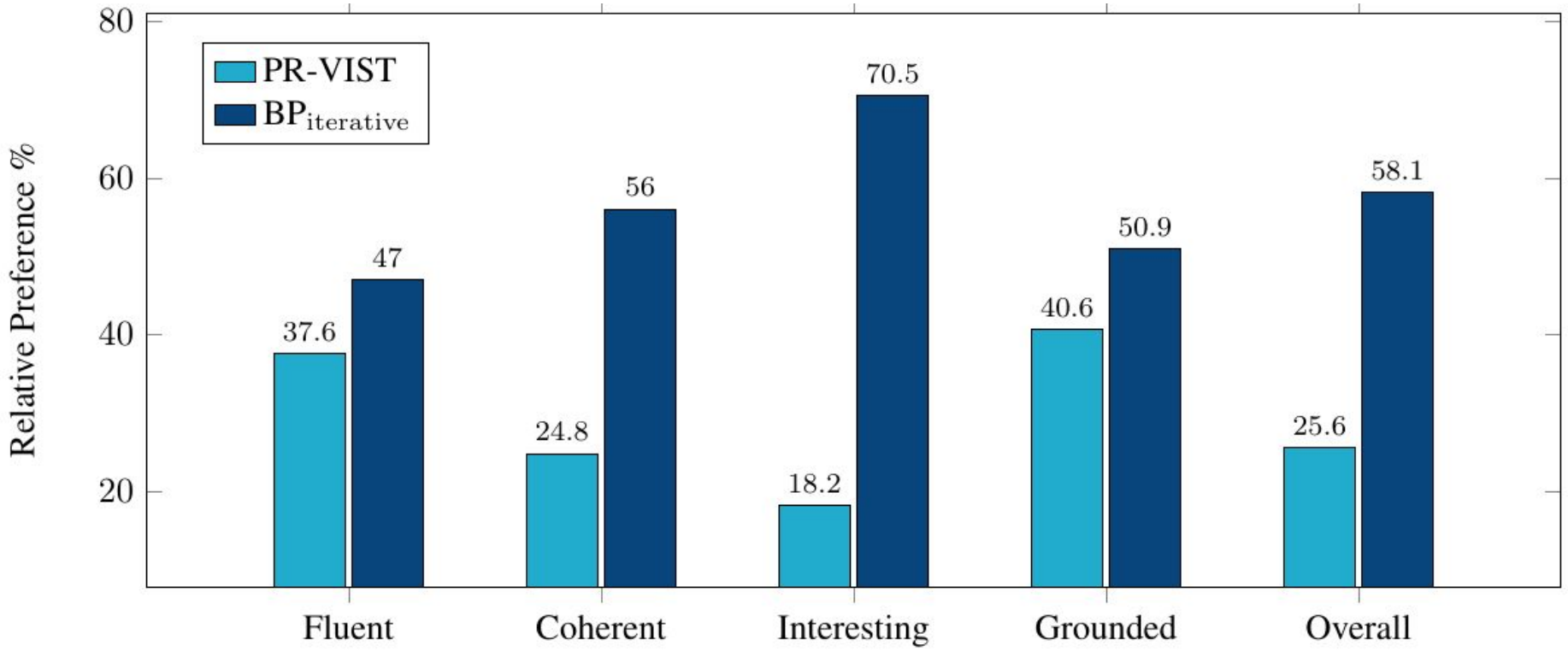
- Automatic evaluation using Blue, Rouge, Meteor is problematic – mostly measures fluency, and all modern LMs generate fluent output
- Also, a story may be interesting, coherent, and grounded in the images, but completely different from the reference
- We use **automatic evaluation** using repetition and grounding, plus Mauve for naturalness (similarity of distribution with human text)
- More importantly, we ask **human judge**s to evaluate fluency, coherence, interestingness and grounding
- We compare to a range of SotA models and to GPT 3.5 (with and without Blueprint)

Automatic Evaluation



VIST is a SotA model in the literature. It doesn't use pretrained LMs; see paper for other comparisons

Human Evaluation



Example Output





PR-VIST: I bought a cart to the market with some food. It looked great in the market. There were so many of their enthusiasm. They even had a sale with their bumper. The market was sold out of fresh vegetables. the [location] car is very nice and beautiful.

VP-BART: I went to the market yesterday. There were many different kinds of fruits there. I bought a lot of them. They were very expensive. Afterward I went back home.

Iterative: Today we decided to take a small shopping trip to the Market. The market had so many wonderful things to choose from that we looked at and bought so many of them. There was an array of different fruits that I could hardly resist buying. There was also a huge array of various types of pepper. Finally the day had come to an end and we piled in the taxi back to head home.

GPT-3.5 + BP: On Saturday morning, I visited the local market. It had a lot of fruits and vegetables being sold. The atmosphere was friendly and buzzing with many vendors ready to help you. The fresh produce was the best thing about the market. You can always find friendly vendors that sell the most delicious produce. The key to finding the best bargains at the market is bargaining. The market is always a great place to shop for fresh fruits and vegetables.

Human: Shoppers arrived early for the market. There was a variety of goods sold. Some carried away large bundles. Every item was fresh and colorful. The spot was a favorite among produce shoppers.

Analysis of Results

- Pretrained language models can produce better stories than specialized models trained from scratch
- The visual prefix is an effective interface between image and text; we don't need multimodal models
- Blueprint model output is most grounded, by automatic and human evaluation
- Iterative planning strategy that generates sentence by sentence works best
- GPT-3.5 struggles with Blueprints; they reduce its performance in most metrics

Conclusions

- Blueprint-based model generates output that is more **coherent**, **interesting**, and **grounded** than existing methods
- Blueprints aid in **selecting key concepts** and **guiding narrative construction**
- Blueprints are **controllable**: generate longer or shorter stories (more/less iterations), emphasize entities or characters (filter Blueprint), etc.
- Blueprints are **interpretable** and could enable human-in-the-loop and personalized storytelling

Narayan et al. Conditional Generation with a Question-Answering Blueprint. TACL,11, 2023. 974-996
Liu et al. Visual Storytelling with Question-Answer Plans. Findings of the Association for Computational Linguistics: EMNLP 2023