

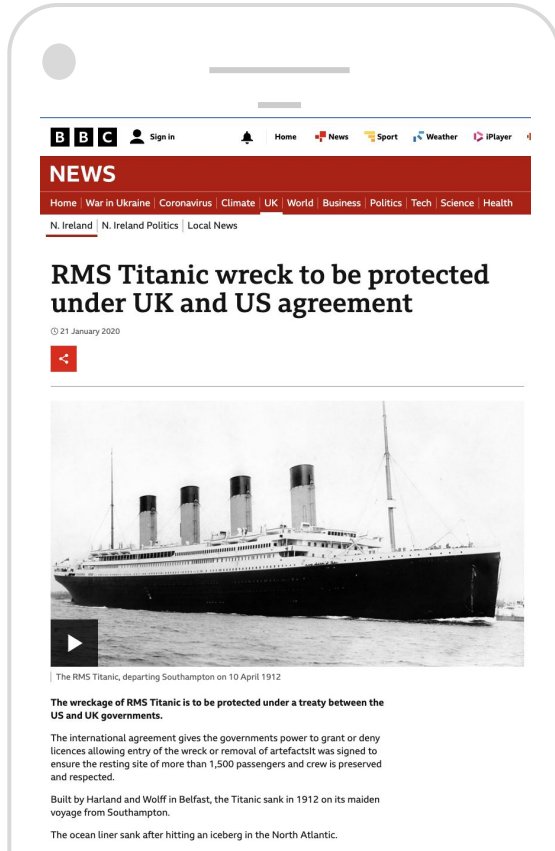
Conditional Generation with a Question-Answering Blueprint

Shashi Narayan
Google Brain

Gardent Fest @ Loria

Conditional Generation Task and Challenges

Communicative goal: Summarization



The screenshot shows a BBC News article titled "RMS Titanic wreck to be protected under UK and US agreement" dated 21 January 2020. The article features a video player showing the RMS Titanic departing Southampton on 10 April 1912. The text below the video states: "The wreckage of RMS Titanic is to be protected under a treaty between the US and UK governments. The international agreement gives the governments power to grant or deny licences allowing entry of the wreck or removal of artefacts it was signed to ensure the resting site of more than 1,500 passengers and crew is preserved and respected. Built by Harland and Wolff in Belfast, the Titanic sank in 1912 on its maiden voyage from Southampton. The ocean liner sank after hitting an iceberg in the North Atlantic."

Possible Summaries

The UK and US have signed a treaty to protect the Titanic wreck.

The Titanic is to be given full international protection after **the US** signed a treaty to protect the wreck.

Challenges

- **Hallucinations**: Generated text contains information that is not consistent with the input.
- **Lack of controllability and explainability**: Blackbox nature of sequence-to-sequence models
- **Lack of Personalization**: Not surfacing text catered to user interests

Task 1: AQuaMuSe: Query-focused Multi-document Summarization

what is the titanic known for

The screenshot displays a search engine interface with several search results for the query 'Titanic'. The top result is from Britannica, titled 'Why is the Titanic famous?'. Below it are results from History.com and Wikipedia. The Britannica article snippet reads: 'From the outset, the *Titanic* captured the public's imagination. At the time, it was one of the largest and most opulent ships in the world. It was also considered unsinkable, due to a series of compartment doors that could be closed if the bow was breached. However, four days into its maiden voyage in 1912, the *Titanic* struck an iceberg, and less than three hours later it sank. The drama of the eyewitness accounts and the great loss of life helped make it one of the most well-known tragedies in modern history. After the 1985 discovery of its wreck, interest in the *Titanic* only increased, and its enduring appeal was evident with the huge success of James Cameron's 1997 film about the doomed ocean liner.'

Answer Summary: RMS Titanic was a British passenger liner, operated by the White Star Line, which sank in the North Atlantic Ocean after striking an iceberg during its maiden voyage from Southampton, UK, to New York City.

Task 2: WikiCatSum: Generating Wikipedia Articles

Topic: Titanic

Wikipedia Reference Articles

The screenshot shows a web browser with multiple search results for the topic 'Titanic'. The top result is from NOAA, titled 'The Titanic early hours to the death'. Below it is a result from HISTORY.COM. The main content area shows a Wikipedia article titled 'Why is the Titanic famous?'. The article text reads: 'From the outset, the Titanic captured the public's imagination. At the time, it was one of the largest and most opulent ships in the world. It was also considered unlikely, due to a series of compartment doors that could be closed if the bow was breached. However, four days into its maiden voyage in 1912, the Titanic struck an iceberg, and less than three hours later it sank. The dozens of eyewitness accounts and the great loss of life helped make it one of the most well-known tragedies in modern history. After the 1985 discovery of its wreckage, interest in the Titanic only increased, and its enduring appeal was evident with the huge success of James Cameron's 1997 film about the doomed ocean liner.'

Titanic

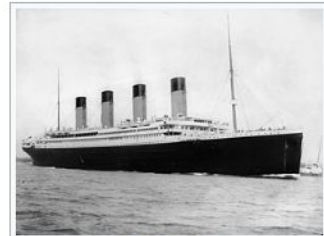
From Wikipedia, the free encyclopedia

Coordinates: 41°43′57″N 49°56′49″W﻿ / ﻿41.73250°N 49.94694°W﻿ / 41.73250; -49.94694

For the film by James Cameron, see [Titanic \(1997 film\)](#). For other uses, see [Titanic \(disambiguation\)](#).

RMS *Titanic* was a British passenger liner, operated by the **White Star Line**, which sank in the North Atlantic Ocean on 15 April 1912 after striking an iceberg during her maiden voyage from Southampton, UK, to New York City. Of the estimated 2,224 passengers and crew aboard, more than 1,500 died, which made the sinking one of the deadliest for a single ship up to that time.^[a] It remains the deadliest peacetime sinking of a superliner or cruise ship.^[4] The disaster drew public attention, provided foundational material for the **disaster film** genre, and has inspired many artistic works.

RMS *Titanic* was the largest ship afloat at the time she entered service and the second of three *Olympic*-class ocean liners operated by the White Star Line. She was built by the Harland and Wolff shipyard in Belfast. Thomas Andrews, the chief naval architect of the shipyard, died in the disaster.^[5] *Titanic* was under the command of Captain Edward Smith,^[6] who went down with the ship. The ocean liner carried some of the wealthiest people in the world, as well as hundreds of emigrants from Great Britain and Ireland, Scandinavia, and elsewhere throughout Europe, who were seeking a new life in the United States and Canada.



Titanic departing Southampton on 10 April 1912

Task 3: SummScreen: Summarizing Multi-turn Conversations/Transcripts

Series 03 Episode 23 – The Lunar Excitation

Scene: On the roof of the apartment building.

Leonard: Okay, we've got power to the laser.

Sheldon: I should've brought an umbrella.

Leonard: What for? It's not going to rain.

Sheldon: I know that, but with skin as fair as mine, moon burn is a real possibility.

Howard: That's a bazinga, right?

Sheldon: One of my best, don't you think?

Leonard: Howard, do you want to double-check the equatorial mount on the laser? We need it locked onto the Sea of Tranquility.

Howard: You got it. Oh, Raj, no. Billions of dollars have gone into inventing the Internet and filling it with pictures of naked women, so we don't have to peep through windows.

Raj: It's not like that, I'm watching someone's TV. The Good Wife is on. I tell you, this is my new Grey's Anatomy.

Sheldon: Leonard, Leonard. What is that? What is that?

Leonard: Relax, it's just a dirty sock.

Sheldon: How on earth can you say dirty sock and relax in the same sentence?

.....

(10 to 14 pages long)

Output: 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.



[SummScreen: Chen et al, ACL 2022.](#)

Conditional Generation Task and Challenges

	AQuaMuSe	WikiCatSum	SummScreen
Inputs			
# docs	6.5	135.6	1.0
# words	12,986.9	7455.8	8051.7
# sentences	339.6	307.8	804.0
Outputs			
# words	114.1	115.6	126.7
Novel trigrams	0.24	0.78	0.92
Task Description			
	Generative Question Answering Query-focused Multi-source Summarization	Wikipedia Generation Topic-focused Multi-source Summarization	Dialogue/Meeting Summarization

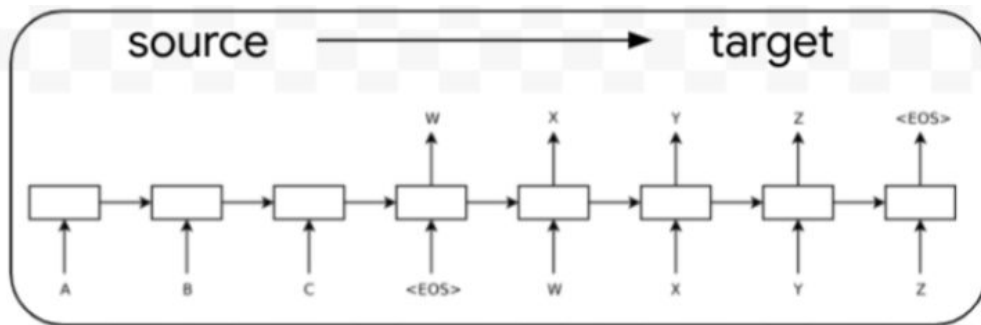
Conditional Generation Task and Challenges

	AQuaMuSe	WikiCatSum	SummScreen
Inputs			
# docs	6.5	135.6	1.0
# words	12,986.9	7455.8	8051.7
# sentences	339.6	307.8	804.0
Outputs			
# words	114.1	115.6	126.7
Novel trigrams	0.24	0.78	0.92
Task Description			
	Generative Question Answering Query-focused Multi-source Summarization	Wikipedia Generation Topic-focused Multi-source Summarization	Dialogue/Meeting Summarization

Challenges

- **Long Inputs**: Requires handling very long inputs
- **Inference and Reasoning over Multiple Sources**: Generated text may require information consolidation over multiple documents
- **Long-form and Multifaceted Nature of Response**: Generated text requires to cover different aspects of the queries or topic.

Conditional Text Generation with Planning



Blackbox
Generation



**Actionable
explanations of
model predictions**

**Grounded generation
of outputs to their content
plans**

Controllable Entity-Based Planning and Grounding for Generation

Proposal: Break the generation step into two parts (**FROST**).



Pretraining with Entity Chain Planning

Can be used in pre-training:

Frozen is a 2013 computer-animated musical fantasy film produced by Walt Disney animation studios. The film is inspired by Hans Christian Anderson's fairy tale 'The Snow Queen'. It depicts fearless princess Anna who joins forces with mountaineer Kristoff and his reindeer sidekick to find estranged sister, Snow Queen Elsa, and break her icy spell. Their epic journey leads them to encounters with mystical trolls, a comedic snowman, harsh conditions, and magic at every turn. Anna and Kristoff bravely push onward in a race to save their kingdom from winter's cold grip.

Article (*Source*)

[MASK] The film is inspired by Hans Christian Anderson's fairy tale 'The Snow Queen'.
[MASK] Their epic journey leads them to encounters with mystical trolls, a comedic snowman, harsh conditions, and magic at every turn. Anna and Kristoff bravely push onward in a race to save their kingdom from winter's cold grip.

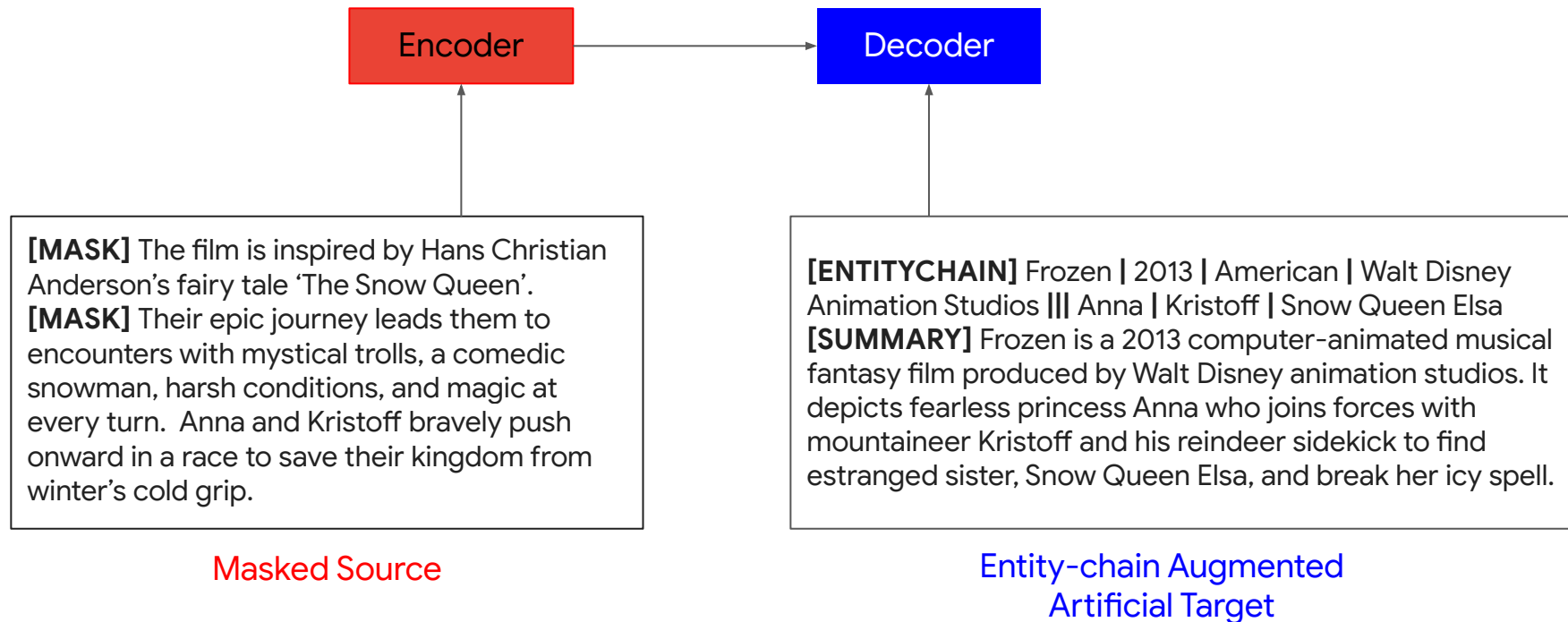
Masked Source

[ENTITYCHAIN] Frozen | 2013 | American | Walt Disney Animation Studios ||| Anna | Kristoff | Snow Queen Elsa
[SUMMARY] Frozen is a 2013 computer-animated musical fantasy film produced by Walt Disney animation studios. It depicts fearless princess Anna who joins forces with mountaineer Kristoff and his reindeer sidekick to find estranged sister, Snow Queen Elsa, and break her icy spell.

Entity-chain Augmented
Artificial Target

Pretraining with Entity Chain Planning

Can be used in pre-training:



Reducing Hallucination with Controlled and Grounded Generation

GOLD	A teenager who stabbed a 16-year-old to death in an Aberdeen school has been jailed for nine years.
PEGASUS ($d \rightarrow s$)	A teenager who killed a 15-year-old pupil at an Aberdeen school has been jailed for eight years.
FROST ($d \rightarrow c s$)	[Entity-Chain] 14 Aberdeen nine [Summary] A teenager who killed a 14-year-old pupil at an Aberdeen school has been jailed for nine years.
FROST ($d \rightarrow c_{\text{drop}} s$)	[Entity-Chain] Aberdeen nine [Summary] A teenager who killed a fellow pupil at an Aberdeen school has been jailed for nine years.
FROST ($d \rightarrow c_{\text{oracle}} s$)	[Entity-Chain] 16 Aberdeen nine [Summary] A teenager who killed a 16-year-old pupil at an Aberdeen school has been jailed for nine years.

Frost models are Better Microplanner and Produce more Concise Summaries

Reference

Raheem Sterling could leave Liverpool in the summer transfer window .
Arsenal and Manchester City are both interested in signing him .
Chelsea could make a big money move if Sterling becomes available .
Bayern Munich boss Pep Guardiola is believed to be an admirer .

Pegasus (d -> t)


Raheem Sterling has attracted interest from a number of big clubs .
Liverpool boss Brendan Rodgers has insisted he will not be sold this summer .
Sterling has rejected a contract worth £100,000-a-week at Anfield .
[Click here for Liverpool transfer news](#) .

Frost

[CONTENT]

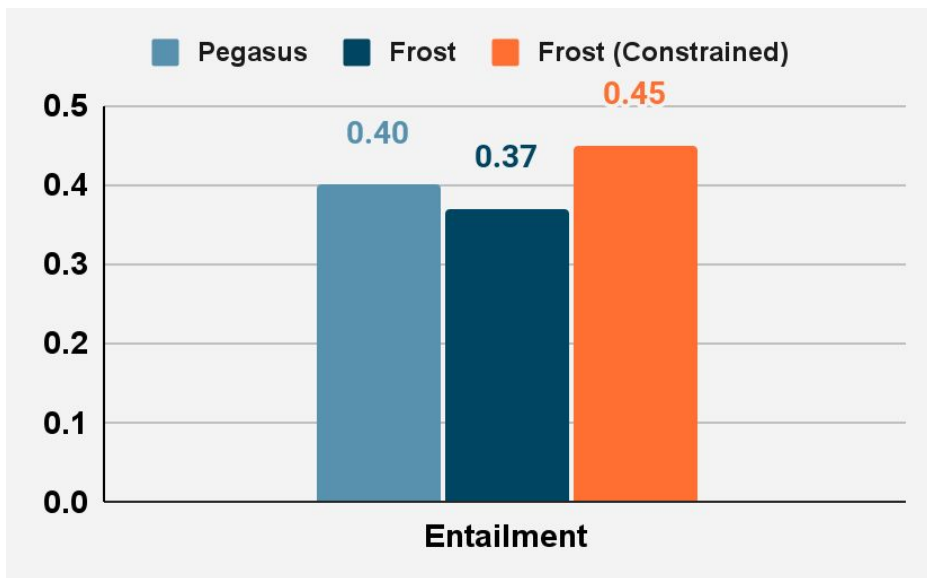
Raheem Sterling | £100,000 | Liverpool |||
Arsenal | Chelsea | Manchester City | Manchester United | Bayern Munich |||
Liverpool | Brendan Rodgers | Sterling

[SUMMARY]

**Raheem Sterling has rejected a £100,000-a-week contract at Liverpool .
Arsenal, Chelsea, Manchester City, Manchester United and Bayern Munich all interested .
Liverpool boss Brendan Rodgers has insisted that Sterling will not be sold .** 

Average lengths: target (60.7), Pegasus (68.6) and Frost (65.0)

XSum: Planning allows Controllable and Faithful Generation



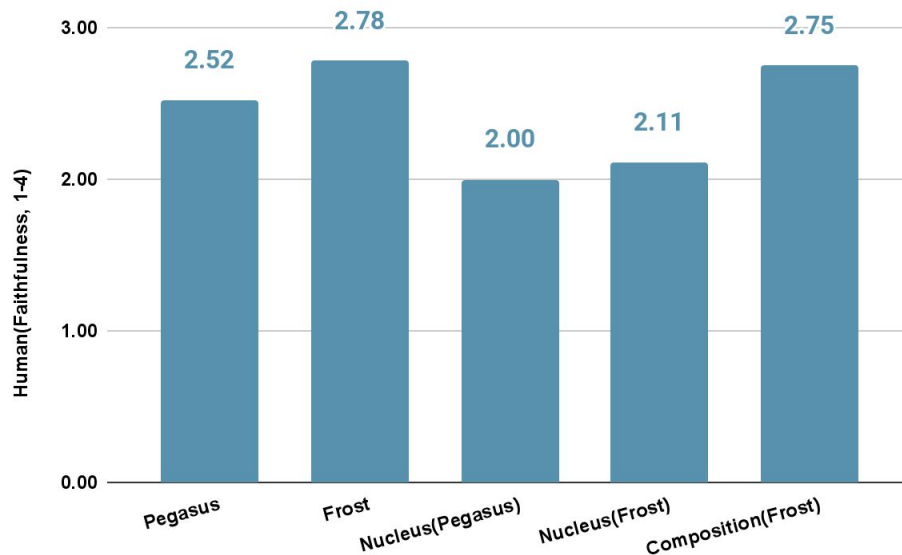
Composition Sampling: Hybrid Decoding with Planned Generation

[CONTENT] Russia | Anna Pavlona | king William II | Netherlands

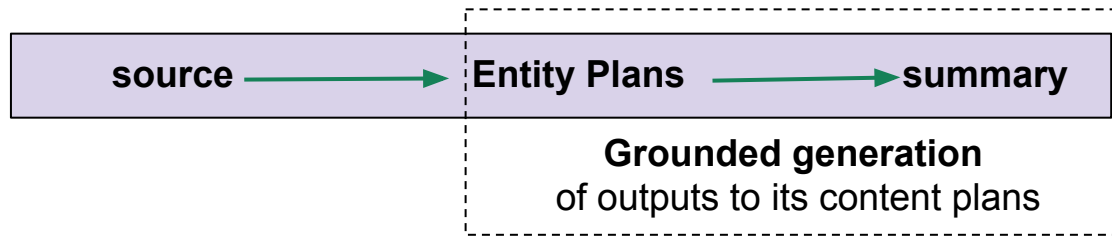
Sampled text
Composition

[SUMMARY] Russian Patriot Anna Pavlona was married to king William II, and the queen consort, of the Netherlands

Generated output grounded to
its composition



Planning with Entity Chains



FROST
Planning with Learned Entity Prompts for
Abstractive Summarization
TACL 2021.



A Well-Composed Text is Half Done!
Composition Sampling for Diverse Conditional Generation
ACL 2022.

<https://github.com/google-research/language/tree/master/language/frost>

Conditional Generation Task and Challenges

	AQuaMuSe	WikiCatSum	SummScreen
Inputs			
# docs	6.5	135.6	1.0
# words	12,986.9	7455.8	8051.7
# sentences	339.6	307.8	804.0
Outputs			
# words	114.1	115.6	126.7
Novel trigrams	0.24	0.78	0.92
Task Description			
	Generative Question Answering Query-focused Multi-source Summarization	Wikipedia Generation Topic-focused Multi-source Summarization	Dialogue/Meeting Summarization

Challenges

- **Long Inputs:** Requires handling very long inputs
- **Inference and Reasoning over Multiple Sources:** Generated text may require information consolidation over multiple documents
- **Long-form and Multifaceted Nature of Response:** Generated text requires to cover different aspects of the queries or topic.

Entity-plans are not expressive enough!

How can we generate a long-form
multifaceted response in a planned and
controllable fashion?

Question-Answering Blueprints as Content Plans

what is the titanic known for



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

Q: What line operated the RMS Titanic?
A: the White Star Line

Q: In what ocean did the RMS Titanic sink?
A: North Atlantic Ocean

Q: During what voyage did the RMS Titanic hit an iceberg?
A: its maiden voyage from Southampton, UK, to New York City



Idea: Use a **question-answering blueprint** as an intermediate planning stage for conditional text generation.

Connection to **Questions Under Discussion (QUD)** theory of discourse structure

Results:

- ✓ Reduced factuality errors
- ✓ Increased controllability and explainability
- ✓ Better long-form summaries

RMS Titanic was a British passenger liner, operated by the White Star Line, which sank in the North Atlantic Ocean after striking an iceberg during its maiden voyage from Southampton, UK, to New York City.

Answer Summary

Blueprint Models: Training and Inference

2-Step Training

Blueprint Annotation
(of gold answers)

Training Blueprint Models

Question Answer Annotations of
Gold Responses

Train encoder-decoder models to take
the input (query and documents) and
generate the blueprint and the response

What kind of a car is a Shelby



Q₁: Who built the Shelby Mustang from 1969 to 1970?

A₁: Ford

Q₂: During what years was the Shelby Mustang built by Shelby American?

A₂: 1965 to 1968

Q₃: In what year was the fifth generation of the Ford Mustang introduced?

A₃: 2005

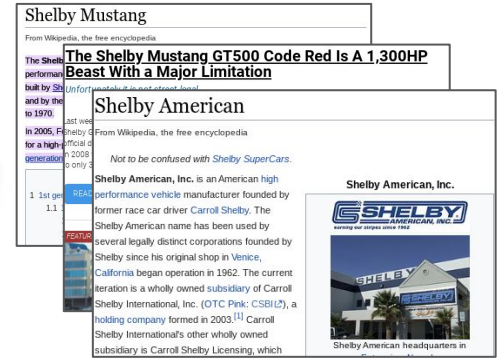
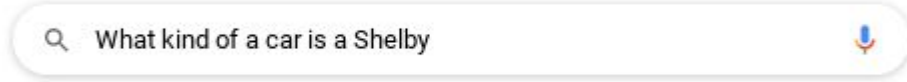
Q₄: What was the Shelby Mustang revived as?

A₄: 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.

Blueprint Models: Training and Inference

Inference



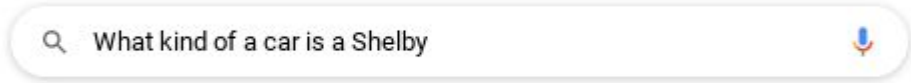
No extra annotation is required



Q ₁ : Who built the Shelby Mustang from 1969 to 1970?	A ₁ : Ford
Q ₂ : During what years was the Shelby Mustang built by Shelby American?	A ₂ : 1965 to 1968
Q ₃ : In what year was the fifth generation of the Ford Mustang introduced?	A ₃ : 2005
Q ₄ : What was the Shelby Mustang revived as?	A ₄ : 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.

Training Step 1: Question-Answering Blueprint Annotation



Q ₁ : Who built the Shelby Mustang from 1969 to 1970?	A ₁ : Ford
Q ₂ : During what years was the Shelby Mustang built by Shelby American?	A ₂ : 1965 to 1968
Q ₃ : In what year was the fifth generation of the Ford Mustang introduced?	A ₃ : 2005
Q ₄ : What was the Shelby Mustang revived as?	A ₄ : 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.

Blueprint Annotation

- **Overgenerate** QA pairs for all answer candidates
- Select **new information seeking questions** and **optimize coverage**

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 Step 1: Question-Answering Blueprint Annotation

What kind of a car is a Shelby



Question-answering Overgeneration

Identify all base **noun phrases and named entities** as answer candidates

Generate questions for each answer candidate using **SQuAD trained Question Generation model (T5-11B)**

Coverage

QA pairs should **cover all salient information** from the source in a **concise manner**

Rheme-based Selection

Answers should **provide new information** (theme-rheme structure)

Round-trip Consistency Check

Statement: The Shelby Mustang was built by Shelby American from 1965 to 1968, and from 1969 to 1970 by Ford.

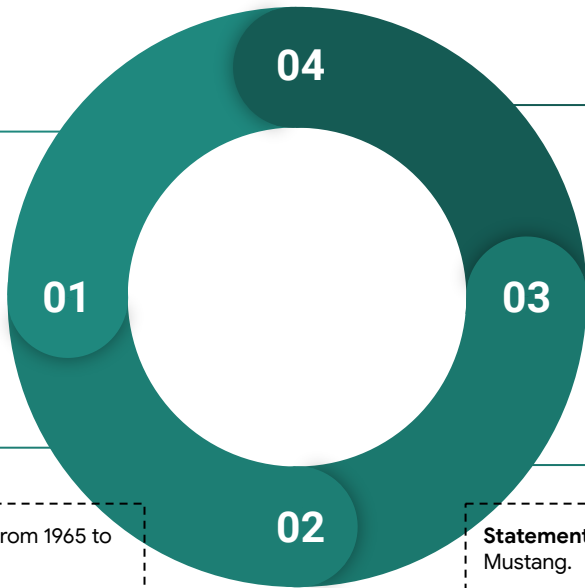
Q: Who built the Shelby Mustang from 1969 to 1970?
A: Ford

Q: What was the final year that Shelby American built the Mustang?
A: 1970

Statement: The Shelby Mustang is a high performance variant of the Ford Mustang.

Q: The Shelby Mustang is a high performance variant of what?
A: the Ford Mustang

Q: What is a high performance variant of the Ford Mustang?
A: The Shelby Mustang



Training Step 1: Question-Answering Blueprint Annotation

What kind of a car is a Shelby



Over Generated Question-Answer Pairs		Round Trip	Rheme	Coverage
Q1: What is a high performance variant of the Ford Mustang?	A1: The Shelby Mustang	✓	✗	
Q2: What is the high performance variant of the Ford Mustang called?	A2: Shelby	✓	✗	
Q3: What is a high performance variant of the Ford Mustang?	A3: Shelby Mustang	✓	✗	
Q4: What is a Shelby Mustang?	A4: a high performance variant	✓	✗	
Q5: The Shelby Mustang is a high performance variant of what?	A5: the Ford Mustang	✓	✓	✗
Q6: The Shelby Mustang is a high performance variant of what?	A6: Ford Mustang	✓	✗	
Q7: The Shelby Mustang is a high performance variant of what Ford model?	A7: Mustang	✓	✗	
Q8: Who built the Shelby Mustang from 1965 to 1968?	A8: Shelby American	✓	✓	✗
Q9: During what years was the Shelby Mustang built by Shelby American?	A9: 1965 to 1968	✓	✓	✓
Q10: In what year did Ford take over production of the Shelby Mustang?	A10: 1969	✓	✗	
Q11: What was the final year that Shelby American built the Mustang?	A11: 1970	✗		
Q12: Who built the Shelby Mustang from 1969 to 1970?	A12: Ford	✓	✓	✓
Q13: What event in 2005 led to the revival of the Shelby Mustang?	A13: the introduction	✗		
Q14: What generation of Mustang was introduced in 2005?	A14: the fifth generation	✓	✗	
Q15: What generation of Mustang was introduced in 2005?	A15: fifth	✓	✗	
Q16: In what year was the fifth generation of the Ford Mustang introduced?	A16: 2005	✓	✓	✓
Q17: What name was brought back for the 2005 Ford Mustang?	A17: the Shelby nameplate	✓	✗	
Q18: What was the Shelby Mustang revived as?	A18: a new high-performance model	✓	✓	✓
[The Shelby Mustang is a high performance variant of the Ford Mustang] _{P1} which [was built by Shelby American] _{P2} [from 1965 to 1968,] _{P3} and [from 1969 to 1970 by Ford.] _{P4} [Following the introduction of the fifth generation Ford Mustang in 2005,] _{P5} [the Shelby nameplate was revived as a new high-performance model, this time designed and built by Ford.] _{P6}				

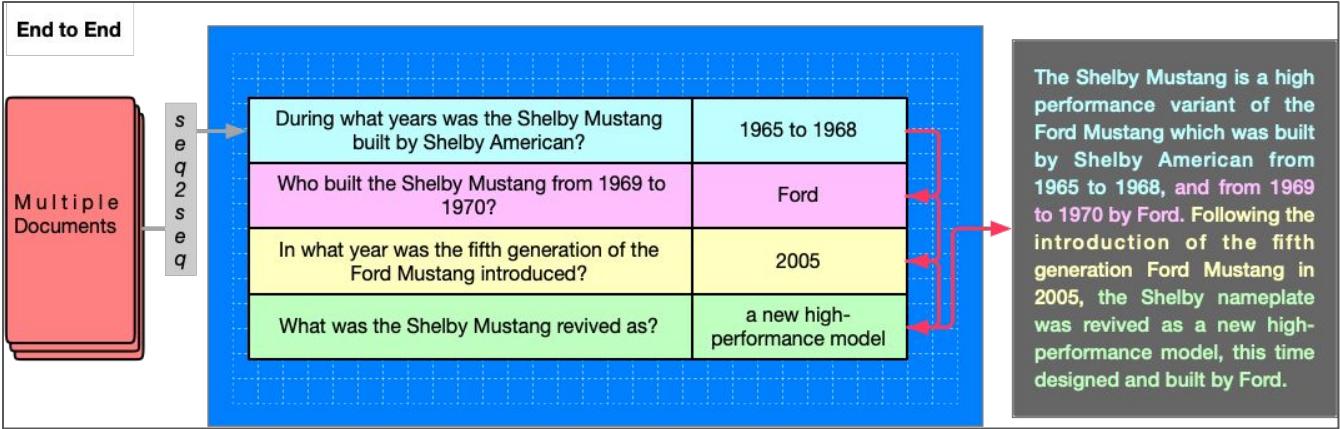
Step 2: Training Blueprint Models



- 1 End-to-End**
Generate the whole blueprint,
followed by the response
- 2 Multitask (Similar to Frost)**
Generate only the answer plans,
followed by the response
- 3 Iterative**
Iteratively plan and generate one
sentence at a time

Training Blueprint Models: End-to-End

1 **End-to-End**
Generate the whole blueprint, followed by the response



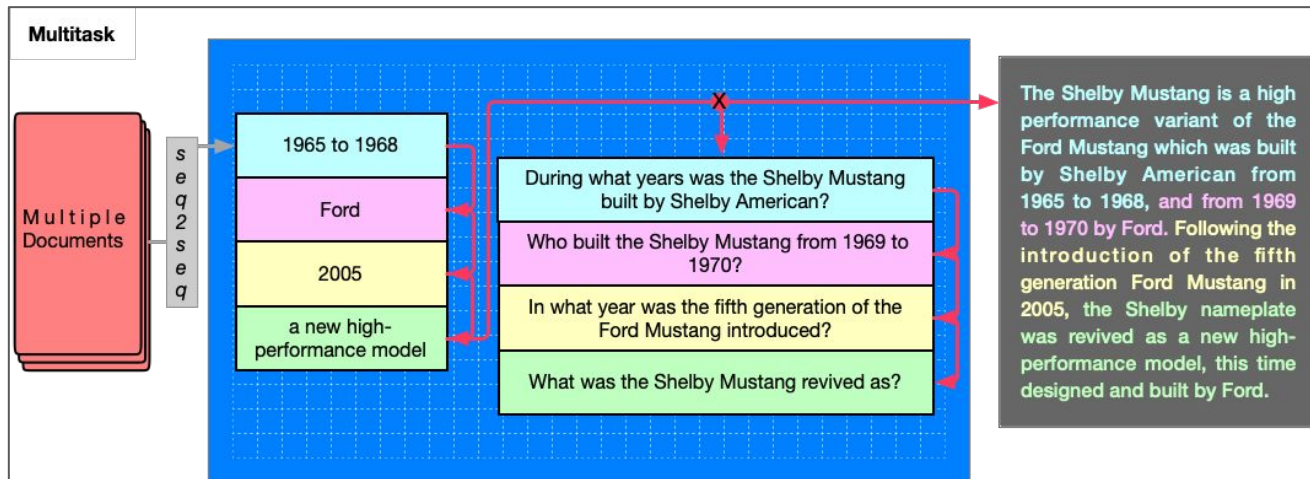
Training Blueprint Models: Multitask

2

Multitask (Similar to Frost)

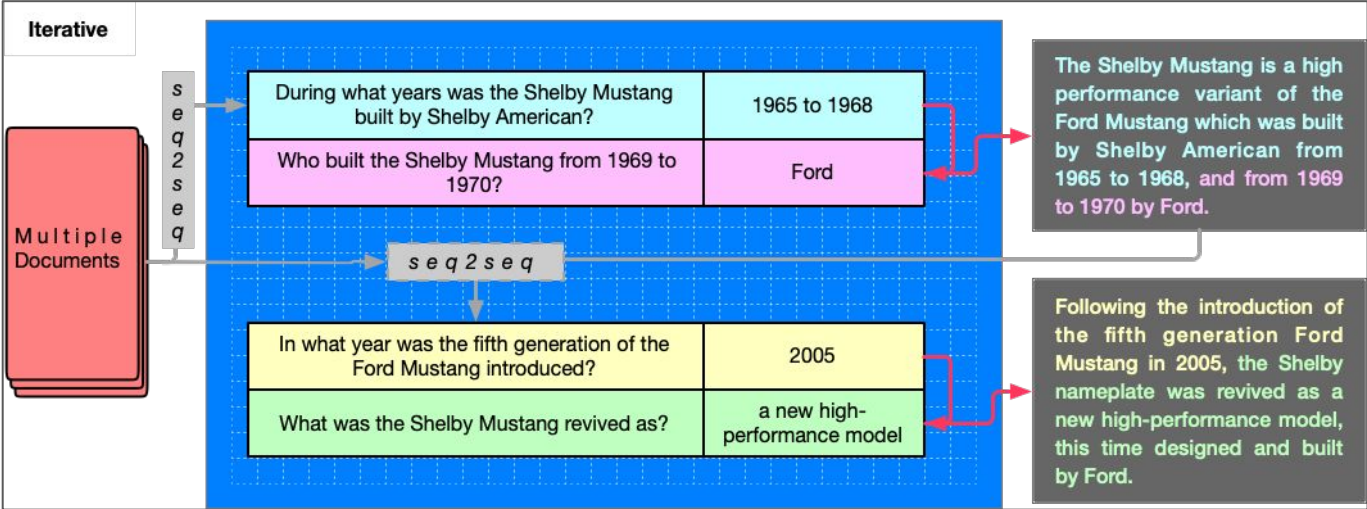
Generate only the answer plans followed by the response

What kind of a car is a Shelby

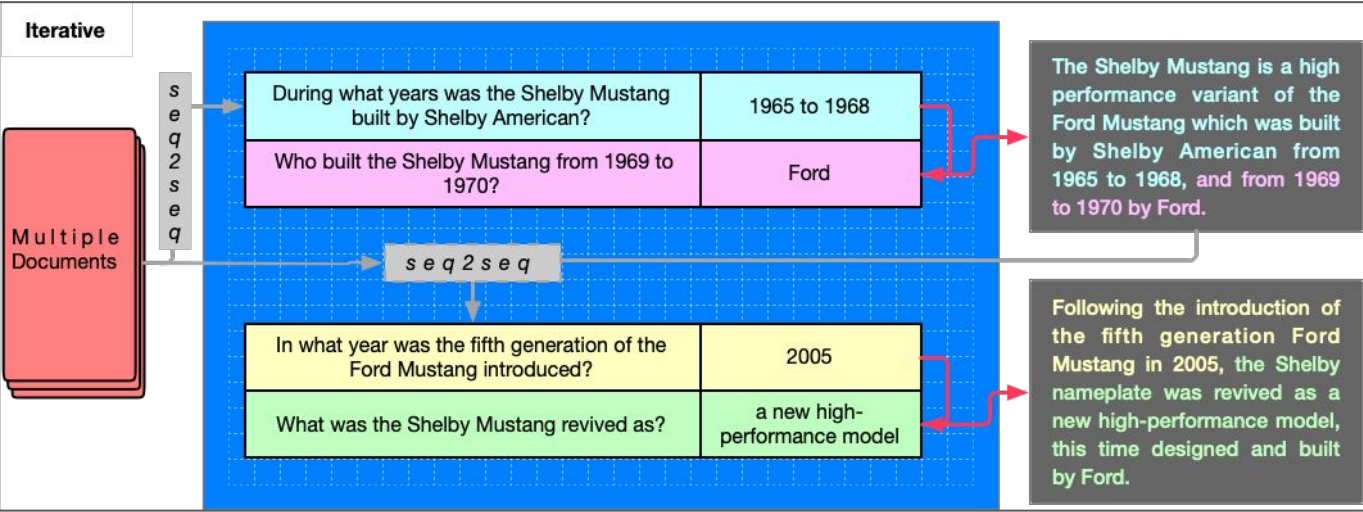


Training Blueprint Models: Iterative

3 Iterative
Iteratively plan and generate one sentence at a time



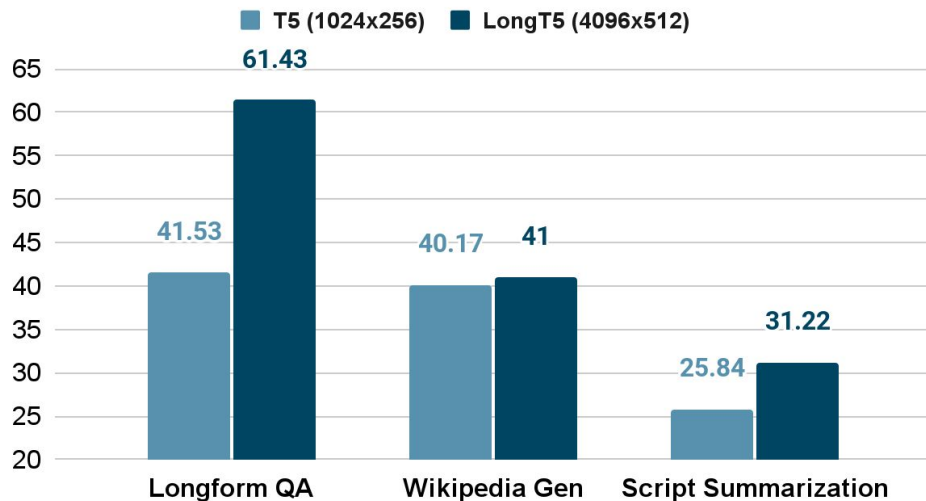
Training Blueprint Models: How Blueprints are Generated Iteratively?



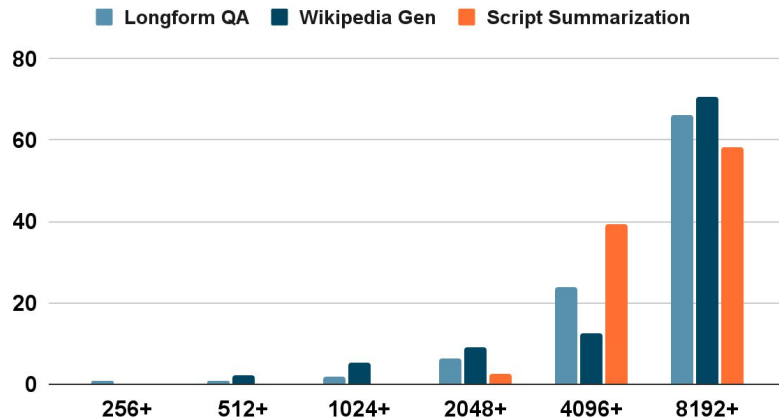
LongT5 to Train Our Blueprint Models

LongT5 are better equipped with modeling long and complex inputs

Rouge (Rouge-L, n-gram overlap with the reference)



Sentencepieces: Input distributions

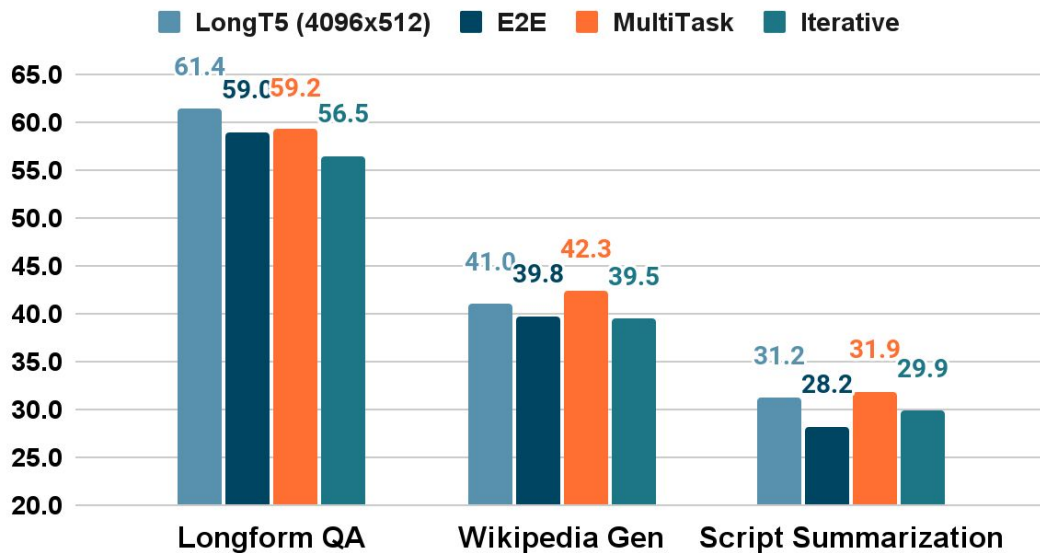


Still a lot that needs to be done for aggregating information and reasoning over multiple- or complex sources

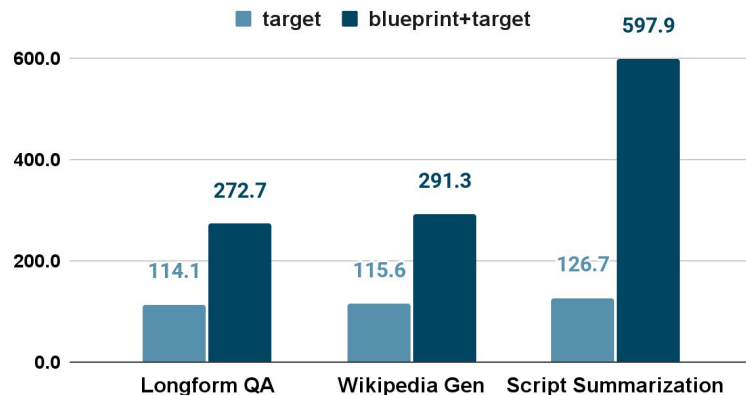
How Blueprint Models Fare in Terms of Rouge

All blueprint variants underperform compared to LongT5, with few exceptions

Rouge (Rouge-L, n-gram overlap with the reference)

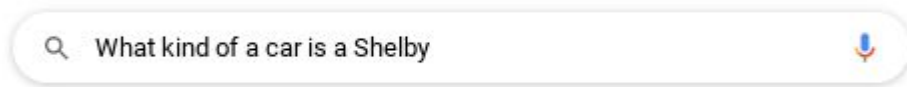


Number of words in target



Blueprint-conditioned Generation Is More Grounded

QA-F1 (Comparing predicted Blueprint questions and Response)



Q₁: Who built the Shelby Mustang from 1969 to 1970?

A₁: Ford

Q₂: During what years was the Shelby Mustang built by Shelby American?

A₂: 1965 to 1968

Q₃: In what year was the fifth generation of the Ford Mustang introduced?

A₃: 2005

Q₄: What was the Shelby Mustang revived as?

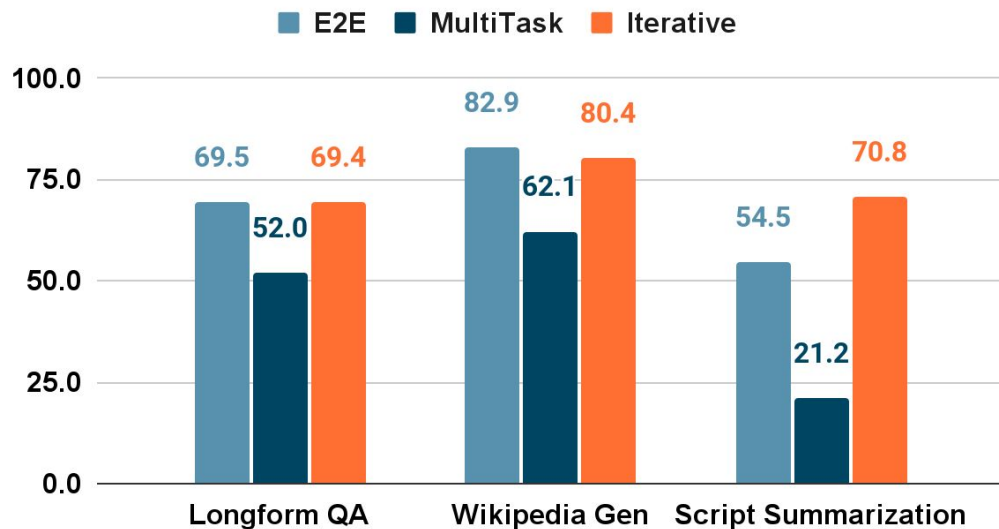
A₄: a new high-performance model

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.

Blueprint-conditioned Generation Is More Grounded

Multitask is inferior to E2E and Iterative

QA-F1 (Comparing Blueprint questions and Response)

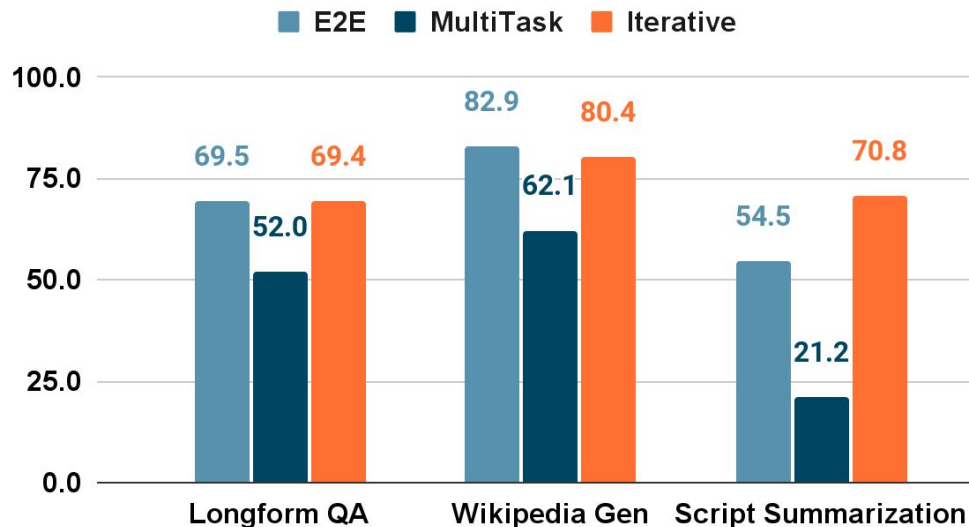


Blueprint-conditioned Generation Is More Grounded

Multitask is inferior to E2E and Iterative

It is important to condition on the whole blueprint

QA-F1 (Comparing Blueprint questions and Response)



Multitask Predictions

Partial Blueprint: Ford | 1965 to 1968 ||| 2005 | a new high-performance model

Response: 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.

E2E Prediction

Blueprint

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

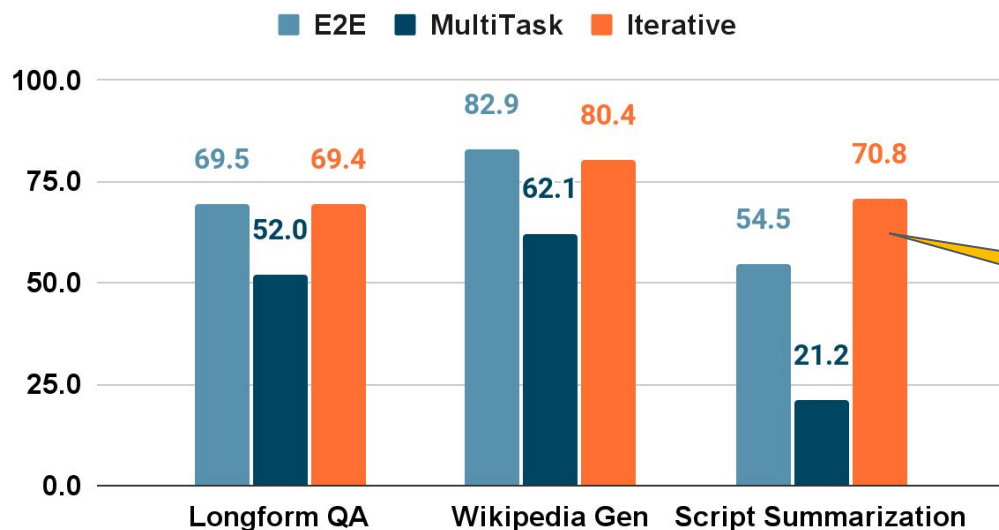
Response: The Shelby Mustang is a high performance variant of the Ford Mustang which was built by Shelby American from 1965 to 1968, and from 1969 to 1970 by Ford. Following the introduction of the fifth generation Ford Mustang in 2005, the Shelby nameplate was revived as a new high-performance model,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

Blueprint-conditioned Generation Is More Grounded

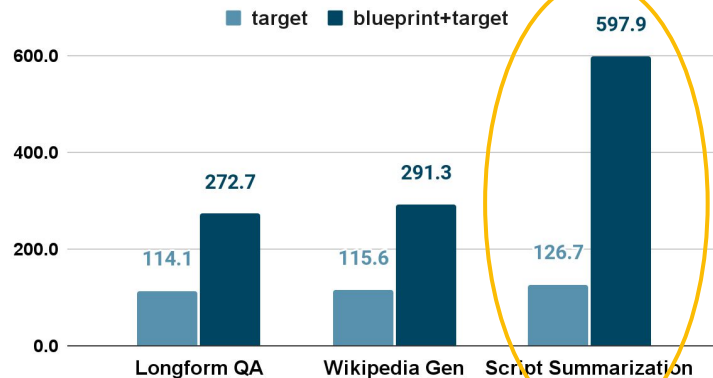
Multitask is inferior to E2E and Iterative

It is important to condition on the whole blueprint

QA-F1 (Comparing Blueprint questions and Response)



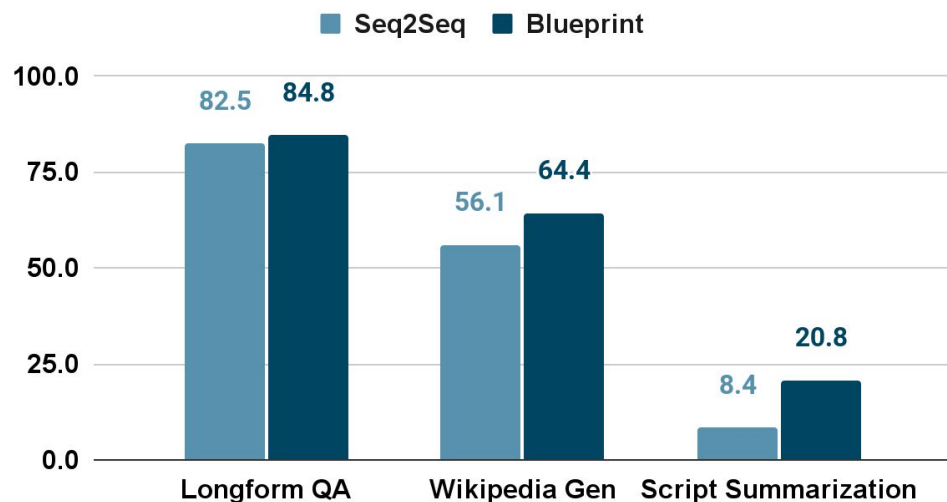
Number of words in target



For script summarization, iterative seems to be the only way to do grounded generation

Grounded Generation Is Most Faithful

% of responses entailed by the input



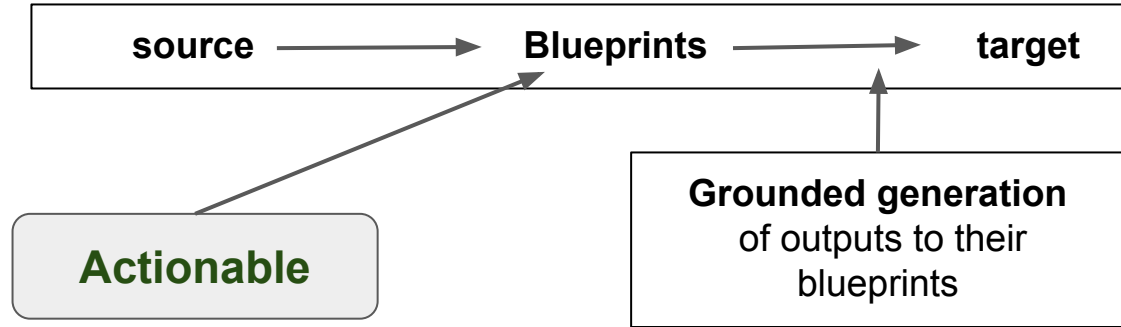
LongT5 prediction

Grissom and Catherine investigate when a man is found dead in a dumpster. They soon discover a lot more went on in the kitchen than cooking. Meanwhile Nick and Sara are called to the scene of a double homicide. The victims are a husband and his wife who were both in the process of selling off their rare records. Suspicion quickly falls on the wife's ex-boyfriend, but the evidence increasingly points to the husband.

Blueprint Iterative prediction

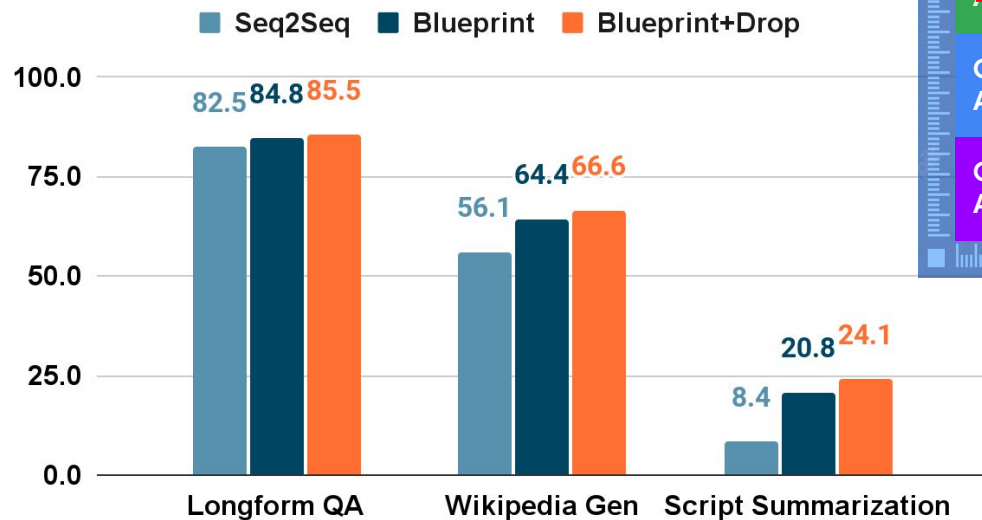
Grissom, Catherine and David investigate when a man is found dead in a dumpster. The man had been eating at a restaurant called Aunt Jackpot's Pretzels. They discover that he ate himself to death. Meanwhile Nick and Sara look into the disappearance of a husband and wife who are found dead in their home. Also missing is a record collection that the husband had been collecting. They discover that the wife's neck was slashed in the attack. CSIs track down Missy Halter, a woman who helped them find the records.

Blueprint Models are Controllable



Grounded and Controllable Generation to Improve Faithfulness

% of responses entailed by the input



A vertical stack of question-answer pairs for the RMS Titanic, illustrating the concept of faithfulness. The questions and answers are as follows:

- Q: What kind of ship was the RMS Titanic?
A: A British passenger liner
- Q: What line operated the RMS Titanic?
A: the White Star Line
- Q: ~~How many ships did the white star line make?~~
A: ~~three ships~~
- Q: In what ocean did the RMS Titanic sink?
A: ~~Pacific Ocean~~
- Q: During what voyage did the RMS Titanic hit an iceberg?
A: its maiden voyage from Southampton, UK, to New York City

RMS Titanic was a British passenger liner, operated by the White Star Line, which sank in the North Atlantic Ocean after striking an iceberg during its maiden voyage from Southampton, UK, to New York City.

Controlling Content with Blueprint Planning

ITERATIVE	
Q ₁ : In what country was Abraham Verghese born? A ₁ : Ethiopia Q ₂ : From what country did Verghese's parents come? A ₂ : India	Abraham Verghese was born in Ethiopia, while his parents, teachers from Kerala, India, lived there.
Q ₃ : Cutting for Stone is set in Ethiopia and what other country? A ₃ : the United States Q ₄ : What are the main characters in Cutting for Stone? A ₄ : two physicians	The main characters in this gripping family saga, set in Ethiopia and the United States, are two physicians from India.
Q ₅ : In what year did Cutting for Stone win the Man Booker Prize? A ₅ : 2007	Cutting for Stone won the Man Booker Prize in 2007.
Q ₆ : Who is the author of Cutting for Stone? A ₆ : Ethiopian-born Verghese	The author is Ethiopian-born Verghese.
ITERATIVE+Q1	
Q ₁ : In what country was Abraham Verghese born? A ₁ : Ethiopia	Abraham Verghese was born in Ethiopia.
Q ₂ : From what country did Verghese's parents come? A ₂ : India	His parents were teachers from Kerala, India.
Q ₃ : Cutting for Stone is set in Ethiopia and what other country? A ₃ : the United States	Cutting for Stone is a gripping family saga, set in Ethiopia.
Q ₄ : What are the main characters in Cutting for Stone? A ₄ : two physicians	The main characters are two physicians from India.

Controlling Content with Blueprint Planning

	now extinct?
Q ₁ :	What breed existed but is no longer extinct?
A ₁ :	Old English Bulldogs
Q ₂ :	Along with the Old English Bulldog and Toy Bulldog, what breed was considered extinct at the end of the 19th century? What was the Old English Bulldog bred for?
A ₂ :	Bullenbeisser Fighting in public arenas

Old English Bulldogs refers to a breed of dog that once existed but is no extinct. At the end of the 19th century, three breeds – the Old English Bulldog, Toy Bulldog, and Bullenbeisser – were considered extinct.

Old English Bulldogs refers to a breed of dog that existed but is now extinct. It was bred for fighting in public arenas.

Blueprint Surfaces Interesting Facets of the Query

Who directed Scarface?



Scarface (1983 film)
From Wikipedia, the free encyclopedia

Scarface is a 1983 American crime drama film directed by Brian De Palma and written by Oliver Stone. Loosely based on the 1929 novel of the same name and serving as a loose remake of the 1932 film, it tells the story of Cuban refugee Tony Montana (Al Pacino), who arrives penniless in Miami during the Miami South and becomes a powerful and extremely homicidal drug lord. The film co-stars Steven Bauer, Michelle Pfeiffer, Mary Elizabeth Mastrantonio and Robert Loggia. De Palma dedicated this version of Scarface to the writers of the original film, Howard Hawks and Ben Hecht. Pacino became interested in a remake of the 1932 version after seeing it, and

Directed by Brian De Palma

Scarface (novel)
From Wikipedia, the free encyclopedia

This article **may** **primary place undue weight** on the subject as a **whole**, **allocating** any relevant in detail that may be against (January 2020)

Scarface is a novel written by Armitage Trail in 1929 and published in 1930. The 1932 film Scarface was based on it. The twenty-eight year-old author died suddenly of a heart attack in 1930.^[1]

Scarface (1932 film)
From Wikipedia, the free encyclopedia

Scarface (also known as **Scarface: The Shame of a Nation** and **The Shame of a Nation**) is a 1932 American pre-Code gangster film directed by Howard Hawks and produced by Hawks and Howard Hughes. The screenplay, by Ben Hecht, is based loosely on the 1929 novel by Armitage Trail which was inspired by Al Capone. The film stars Paul Muni as Italian immigrant gangster Antonio "Tony" Camonte, a gangster who violently rises through the Chicago gangland, with a supporting cast that includes George Raft and Boris Karloff. Camonte's rise to power dovetails with his relentless pursuit of his boss's mistress while his own inner passions

Directed by Howard Hawks

What is 'Scarface'?

a 1983 American crime film

Who directed 'Scarface'?

Brian De Palma

Who wrote 'Scarface'?

Oliver Stone

Along with Al Pacino, Steven Bauer, Michelle Pfeiffer and Mary Elizabeth Mastrantonio, who starred in Scarface?

Robert Loggia

What is the name of the 1929 novel that Scarface is based on?

the same name

Who directed the 1932 film of the same name?

Howard Hawks

Along with Howard Hawks, who wrote the original Scarface?

Ben Hecht

What premiered in New York City on December 1, 1983?

The film

On what date was Scarface released?

December 9, 1983

What studio released Scarface?

Universal Pictures

"Scarface is a 1983 American crime film directed by Brian De Palma and written by Oliver Stone.

Loosely based on the 1929 novel of the same name and serving as a loose remake of the 1932 film, it tells the story of Cuban refugee Tony Montana (Al Pacino), who arrives penniless in Miami during the Mariel boatlift and becomes a powerful and extremely homicidal drug lord.

The film co-stars Steven Bauer, Michelle Pfeiffer, Mary Elizabeth Mastrantonio and Robert Loggia.

De Palma dedicated this version of Scarface to the writers of the original film, Howard Hawks and Ben Hecht.

The film premiered in New York City on December 1, 1983, and was released on December 9, 1983, by Universal Pictures.

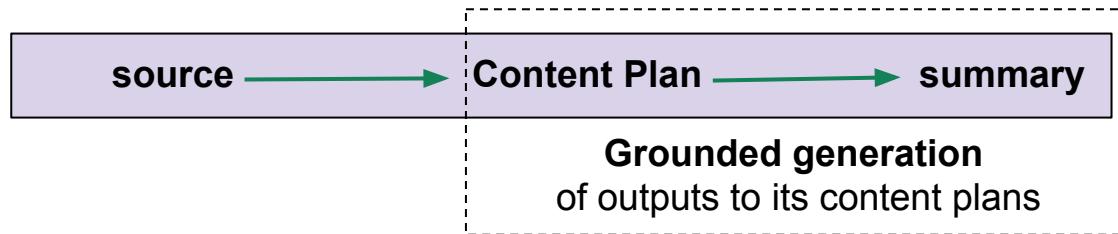
The film grossed \$45 million at the domestic box office and \$66 million worldwide."

Human Evaluation results in Script Summarization

- Iterative summaries are selected best similarly often to LongT5
- Plans are perceived to be coherent (relevant and ordered) with varying levels of redundancy
- Iterative summaries are used*, grounded+ and have new information

Models	summary	blueprint		grounding		
	best 0-to-1	quality 1-to-3	redund. %	abs. %	contr. %	new-info 3-to-1
R2T-BART	0.10	—	—	—	—	—
LONGT5	0.42 ^{◇†‡}	—	—	—	—	—
E2E	0.26	2.43	49.3	26.6	5.3 [◇]	1.76 [◇]
MULTITASK	0.39 ^{◇*}	2.28 [◇]	33.0	63.0	8.7	2.60
ITERATIVE	0.44[†]	2.09	81.7	12.0[◇]	2.8[†]	1.53
+drop	0.39 ^{‡*}	2.28 [◇]	63.7	14.2 [◇]	3.6 ^{◇†}	1.89 [◇]

Planning-based Models are Faithful, Controllable, & Attributable



Planning-based Generation

- **Frost** and **Composition Sampling**: Planning with Entity Chains
- **Blueprint**: Planning with Question Answer Blueprints

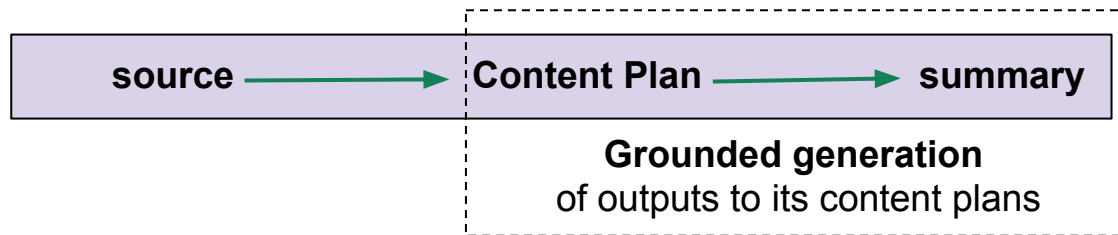
Planning-based models

- can generate response in a **controllable and explainable fashion**

Blueprint models:

- Can generate **long and multifaceted response**, summarizing a **diverse set of documents**
- Can **organize information** using the QA blueprints

Planning-based Models are Faithful, Controllable, & Attributable



- Paper: <https://arxiv.org/pdf/2207.00397.pdf> (Will appear at TACL)
- Demo at EACL 2023
- Experiments with LLMs
 - Query Refinement Prompts for Closed-book Long-form QA
 - <https://arxiv.org/pdf/2210.17525.pdf>

Questions?

What's next?

Retrieval-based Generation with Blueprints

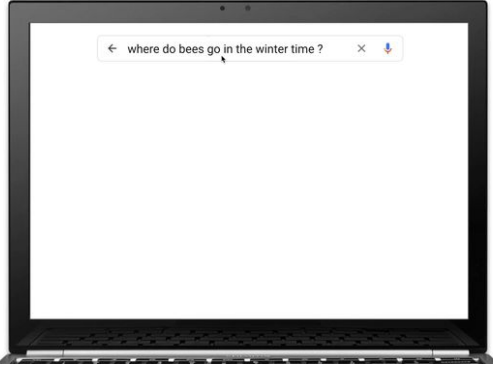
Attribution with Blueprints

Text-Blueprint

Surfacing personalized query and response

Handling Multi-sources and complex inputs

Better content selection, planning and attribution



Q&A Bank

- Q: What is the product?
A: iPhone case
- Q: How much does it cost?
A: 25 euros
- Q: How is it made?
A: 3D printed
- Q: What is it made of?
A: High-gloss Polycarbonate
- Q: Is it scratch-resistant?
A: Yes
- Q: Does shipping cost?
A: Free for all orders over 69
- .
- .

text-blueprint^{durability}

- Q: Is it scratch-resistant?
A: Yes
- Q: What is the product?
A: iPhone case
- Q: What is it made of?
A: High-gloss Polycarbonate

Output

<gCase> is a scratch-resistant iPhone case. This is due to being made by high-gloss polycarbonate.

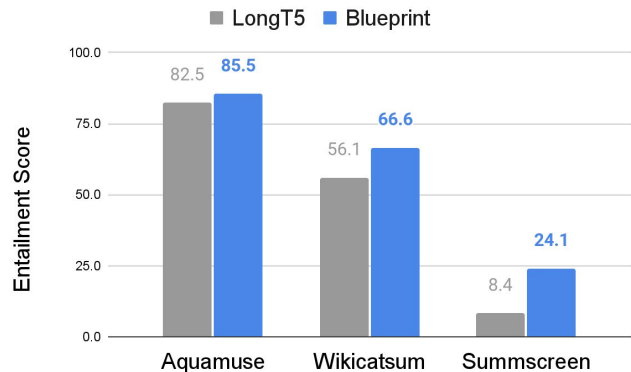
text-blueprint^{price}

- Q: How much does it cost?
A: 25 euros
- Q: Does shipping cost?
A: Free for all orders over 69

Output

<gCase> costs 25 euros and is shipped for free if order exceeds 69 euros.

Blueprint Models are Faithful, Controllable, & Attributable

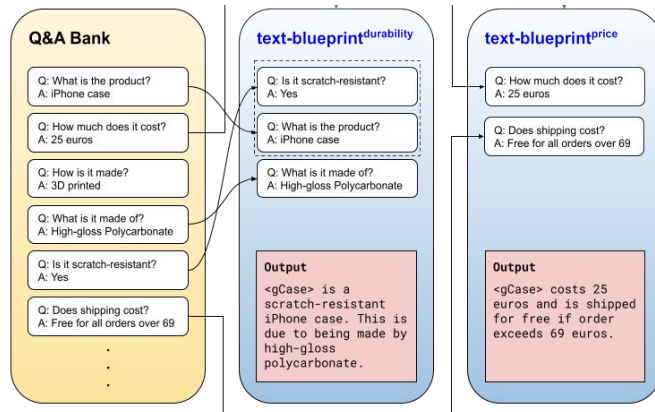
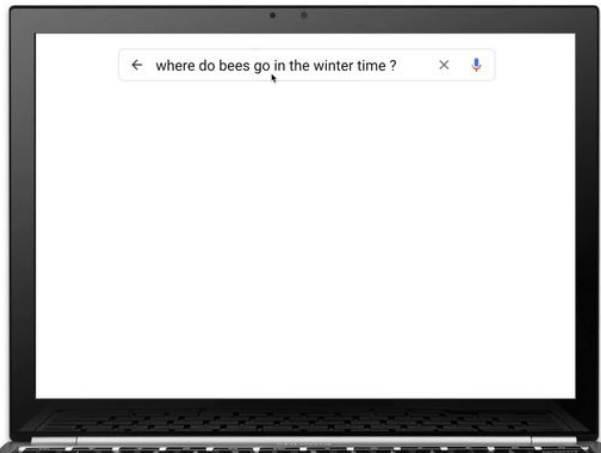


Faithful

Results on long-form summarization datasets show **significant increase in entailment scores**

Controllable

Summaries are **highly grounded** on the Blueprint (> 70 QA-F1)



Attributable

We can **create a Q&A database** with Blueprint, where each QA pair is attributable to the source