

Bringing Planning Back to Neural Generators

Shashi Narayan

Research Scientist, Google London

Google Research

NL4XAI NTP5, April, 2022

Bringing Planning Back to Neural Generators

Text Generation: Models and Issues

Planning based Generation for **faithful** and **diverse** text generation

Text Generation

- Given a source sequence \mathbf{x} , output a target \mathbf{y}

$$\mathbf{x} = (x_1, \dots, x_S)$$

$$\mathbf{y} = (y_1, \dots, y_T)$$

- Examples:
 - **Translation:** \mathbf{x} = English sentence, \mathbf{y} = French sentence
 - **Summarization:** \mathbf{x} = Document, \mathbf{y} = One paragraph summary
 - **Data to Text:** \mathbf{x} = Structured data, \mathbf{y} = Textual description.
 - **Story Generation:** \mathbf{x} = story prompt, \mathbf{y} = story continuation

Document Summarization



MailOnline

Article: Queen Victoria spent her holidays in Osborne House on the Isle of Wight. ... She would travel to Portsmouth by train and then by ferry to Ryde. From Ryde there was a railway line that passed not far from Osborne House but the nearest station was at Wootton, more than two miles from the property. So, in 1875, a station was built at Whippingham, the closest point on the line to Osborne House – just to serve the Royal residence. ... The building is now a five-bedroom family home, currently on the market for £625,000, while the track has become a cycle path. ...

Hermann et al.,
2015

Human-written Summary (Story highlights)

- Queen Victoria's holiday residence was Osborne House on the Isle of Wight
- But her journeys there involved train and ferry ride and then another train ride to a station more than two miles from the property
- In 1875, a station was built at Whippingham just to serve Royal residence
- Building is now a five-bedroom home, currently on the market for £625,000

Google Research

Generated text can be “Extractive”



MailOnline

Article: Queen Victoria spent her holidays in Osborne House on the Isle of Wight. ... She would travel to Portsmouth by train and then by ferry to Ryde. From Ryde there was a railway line that passed not far from Osborne House but the nearest station was at Wootton, more than two miles from the property. So, in 1875, a station was built at Whippingham, the closest point on the line to Osborne House – just to serve the Royal residence. ... The building is now a five-bedroom family home, currently on the market for £625,000, while the track has become a cycle path. ...

Hermann et al.,
2015

Human-written Summary (Story highlights)

- Queen Victoria's holiday residence was Osborne House on the Isle of Wight
- But her journeys there involved train and ferry ride and then another train ride to a station more than two miles from the property
- In 1875, a station was built at Whippingham just to serve Royal residence
- Building is now a five-bedroom home, currently on the market for £625,000

Google Research

“Abstractive” Text Generation

Extreme Summarization (XSum dataset, Narayan et al., 2018)



SUMMARY: *A man and a child have been killed after a light aircraft made an emergency landing on a beach in Portugal.*

DOCUMENT: Authorities said the incident took place on Sao Joao beach in Caparica, south-west of Lisbon.

The National Maritime Authority said a middle-aged man and a young girl died after they were unable to avoid the plane.

[6 sentences with 139 words are abbreviated from here.]

Other reports said the victims had been sunbathing when the plane made its emergency landing.

[Another 4 sentences with 67 words are abbreviated from here.]

Video footage from the scene carried by local broadcasters showed a small recreational plane parked on the sand, apparently intact and surrounded by beachgoers and emergency workers.

[Last 2 sentences with 19 words are abbreviated.]

Abstractive Data-to-text Generation

TEAM	WIN	LOSS	PTS	FG.PCT	RB	AS ...
Heat	11	12	103	49	47	27
Hawks	7	15	95	43	33	20

PLAYER	AS	RB	PT	FG	FGA	CITY ...
Tyler Johnson	5	2	27	8	16	Miami
Dwight Howard	4	17	23	9	11	Atlanta
Paul Millsap	2	9	21	8	12	Atlanta
Goran Dragic	4	2	21	8	17	Miami
Wayne Ellington	2	3	19	7	15	Miami
Dennis Schroder	7	4	17	8	15	Atlanta
Rodney McGruder	5	5	11	3	8	Miami
Thabo Sefolosha	5	5	10	5	11	Atlanta
Kyle Korver	5	3	9	3	9	Atlanta
...						

The Atlanta Hawks defeated the Miami Heat , 103 - 95 , at Philips Arena on Wednesday . Atlanta was in desperate need of a win and they were able to take care of a shorthanded Miami team here . Defense was key for the Hawks , as they held the Heat to 42 percent shooting and forced them to commit 16 turnovers . Atlanta also dominated in the paint , winning the rebounding battle , 47 - 34 , and outscoring them in the paint 58 - 26. The Hawks shot 49 percent from the field and assisted on 27 of their 43 made baskets . This was a near wire - to - wire win for the Hawks , as Miami held just one lead in the first five minutes . Miami (7 - 15) are as beat - up as anyone right now and it 's taking a toll on the heavily used starters . Hassan Whiteside really struggled in this game , as he amassed eight points , 12 rebounds and one blocks on 4 - of - 12 shooting ...

Figure 1: An example data-record and document pair from the ROTOWIRE dataset. We show a subset of the game's records (there are 628 in total), and a selection from the gold document. The document mentions only a select subset of the records, but may express them in a complicated manner. In addition to capturing the writing style, a generation system should select similar record content, express it clearly, and order it appropriately.

Neural Encoder Decoder Paradigm

$\mathbf{x} = (x_1, \dots, x_S)$ source text

$\mathbf{y} = (y_1, \dots, y_T)$ target text

$P(\mathbf{y}|\mathbf{x}) \approx P(\mathbf{y}|\mathbf{c})$ where $\mathbf{c} = F_{enc}(\mathbf{x})$

encoding function

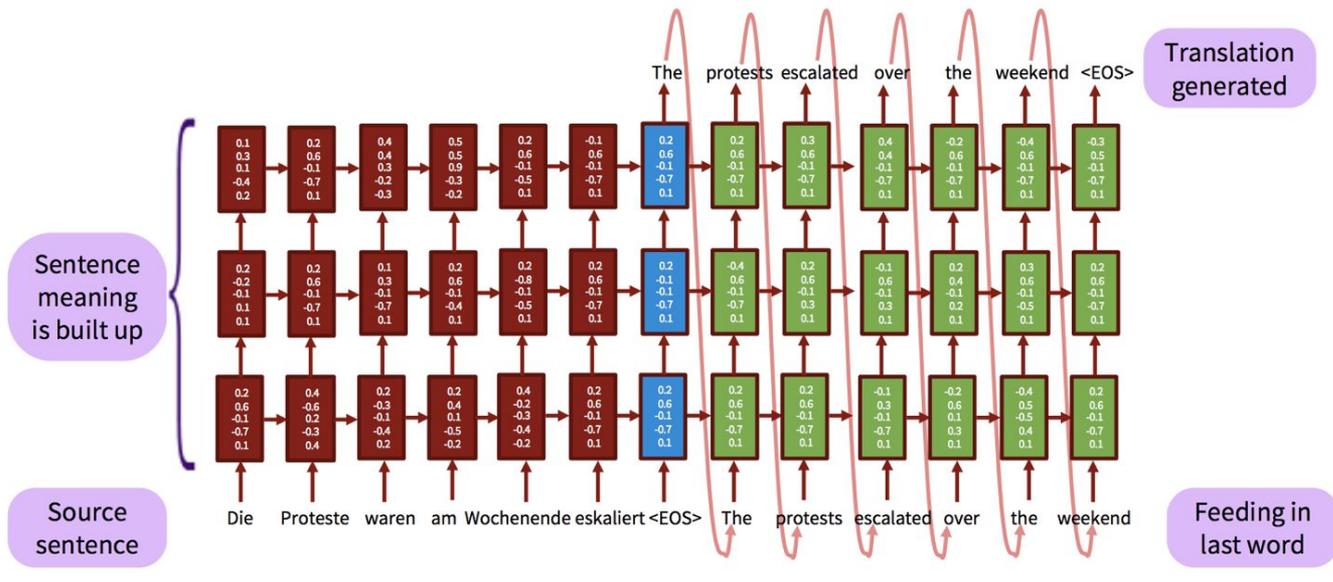
$$P(\mathbf{y}|\mathbf{c}) = \prod_{t=1}^T P(y_t | y_1, \dots, y_{t-1}, \mathbf{c})$$

language model conditioned on the source encoding

Encoder = some neural network

Decoder (conditional language model) = some neural network

Neural Encoder Decoder Paradigm



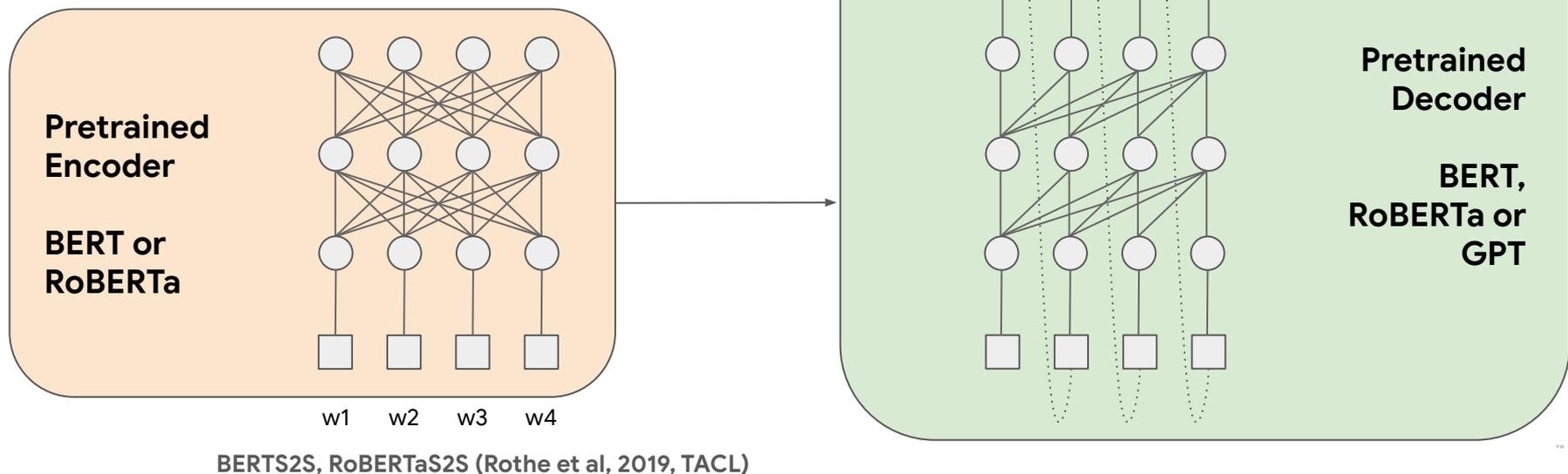
LSTMs, Attention [Sutskever et al. 2014, Bahdanau et al. 2014]

Transformers, Self- and multi-headed Attention [Vaswani et al, 2017]

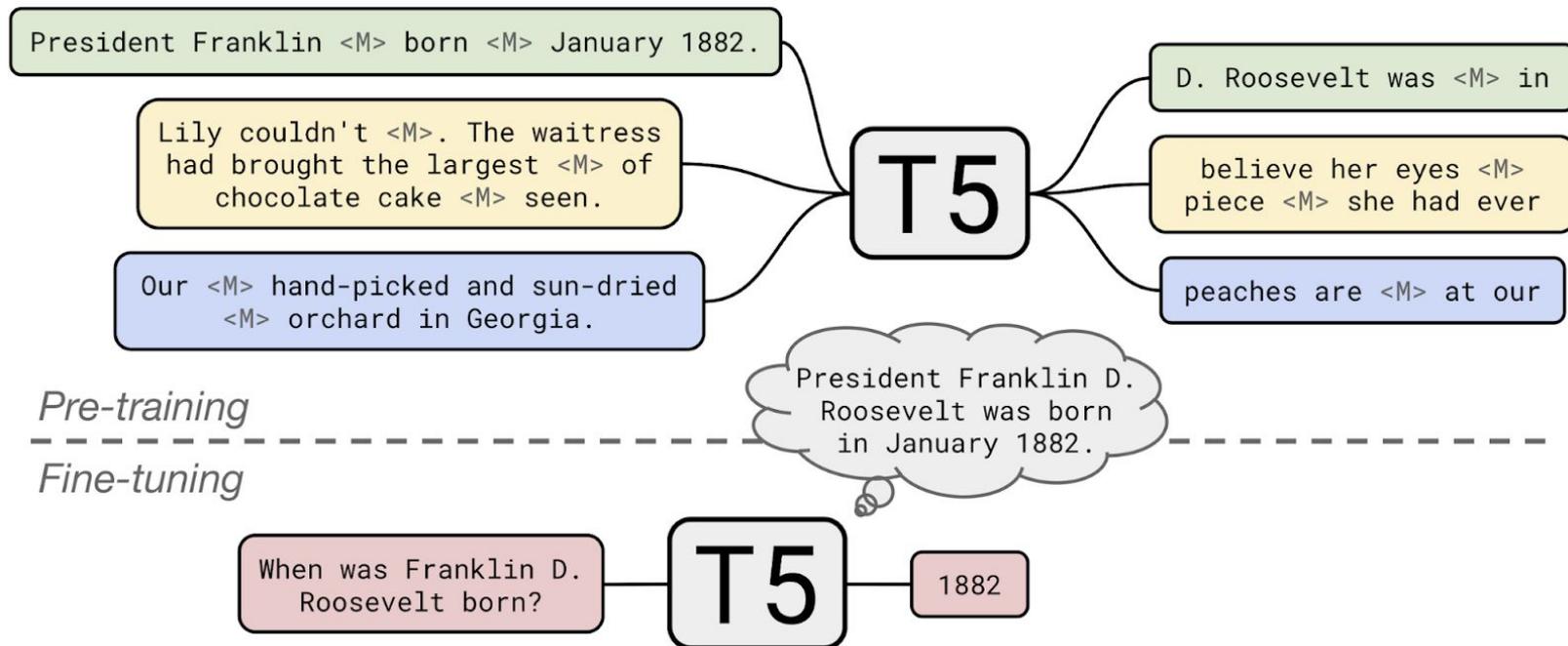
NLU Pretrained Models for Text Generation

Developed a transformer architectures compatible with public checkpoints of BERT, RoBERTa and GPT-2.

Publicly available pre-trained checkpoints for warm-starting sequence generation models.

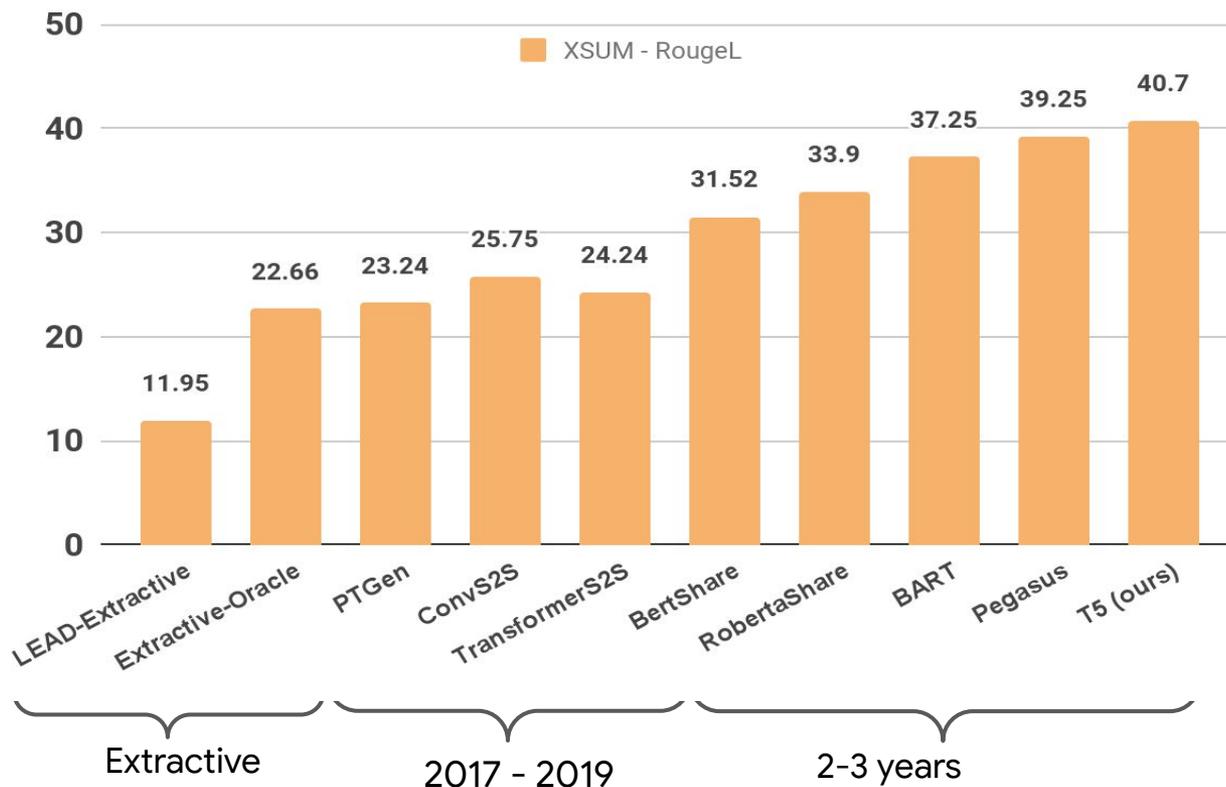


End-to-End Pretraining of Generation Models



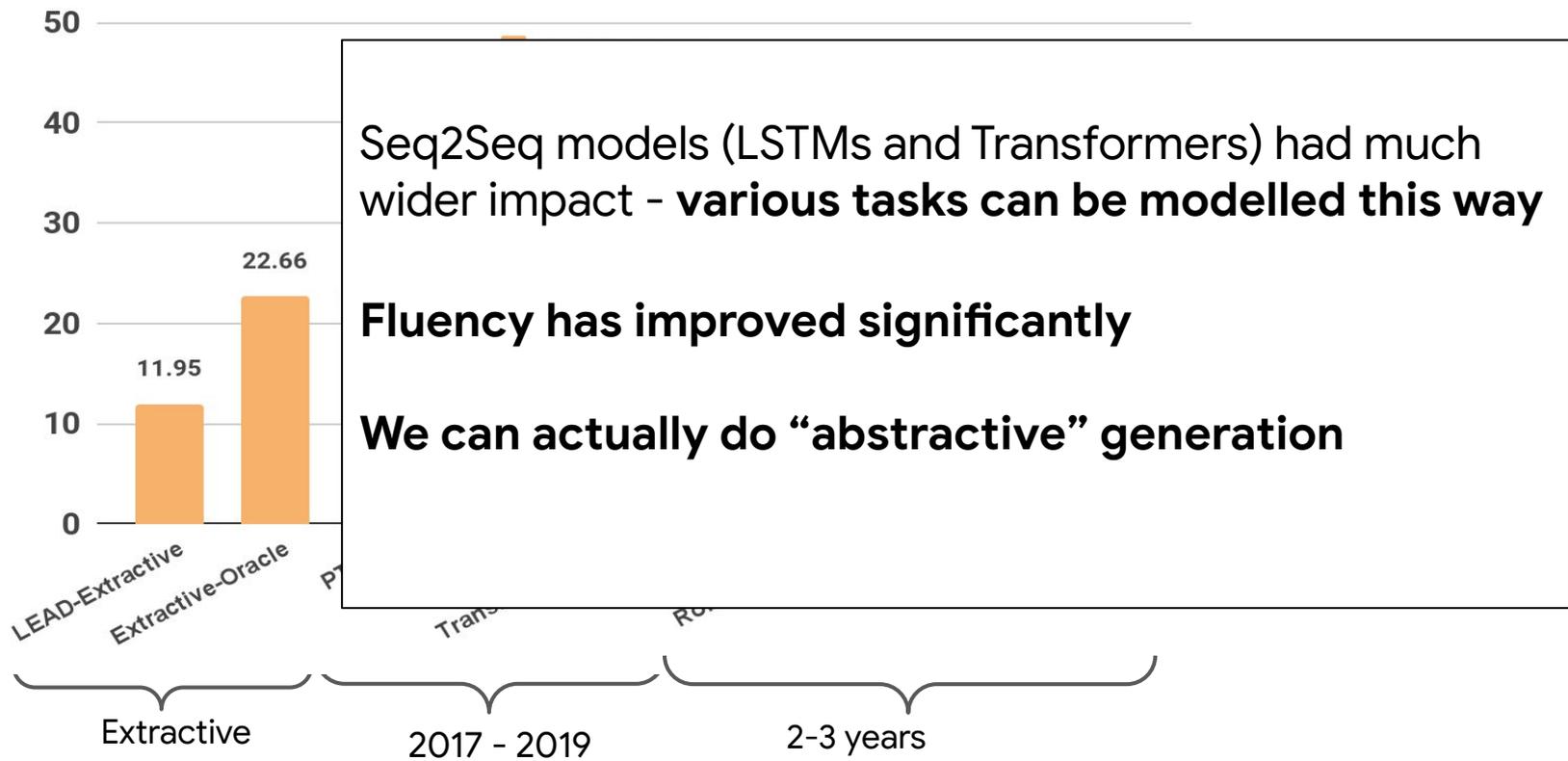
See others: GPT, MASS, BART, UniLM, Pegasus, LamDa etc.

Pretrained Model Performance on Abstractive Summarization



Extreme Summarization (XSum dataset, Narayan et al., 2018)

Pretrained Model Performance on Abstractive Summarization



Challenges for Trustworthy Neural Generation

- Hallucinations are very common, but hard to control or explain

Reference

School lunches can tempt fussy eaters to try new foods, a survey for the school food trust has suggested.

Generated

Parents who eat lunches at school are more likely to eat a meal than their children, a survey suggests.

Challenges for Trustworthy Neural Generation

- Hallucinations are very common, but hard to control or explain

XSum Input: Leahy, 24, scored 12 goals in 158 appearances with Falkirk, having joined the club from Rugby Town in 2012. The left-back made 38 appearances last season, helping the club finish second in the Scottish second tier before they lost to Dundee United in the play-offs. He joins Walsall on a free transfer after his contract expired and is the League One club's first summer signing. Find all the latest football transfers on our dedicated page.

Reference: Walsall have signed defender Luke Leahy on a two-year contract from Scottish Championship side Falkirk.

Output: Walsall have signed defender Paddy Leahy from Scottish Championship side Falkirk on a three-year deal.

Challenges for Trustworthy Neural Generation

**Evaluation and
Benchmarking**

**Controllable
Models**

Challenges for Trustworthy Neural Generation

**Evaluation and
Benchmarking**

**Understand and measure types of
hallucinations**

**Controllable
Models**

Planning-based Controllable and Faithful
Generation

Understand and measure types of hallucinations

- ❖ What types of hallucinations occur?
- ❖ How frequently do systems hallucinate?
- ❖ Are there automatic means of measuring these hallucinations?

We investigated these
questions via a large scale
human study

Major Questions in Study

- ❖ **What types of hallucinations occur?**
- ❖ How frequently do systems hallucinate?
- ❖ Are there automatic means of measuring these hallucinations?

Intrinsic Hallucinations

Source Text: Germany emerge victorious in 2-0 win against Argentina on Saturday

Summary: Germany beat Argentina 2-0

- ❖ Model misrepresents information present in the input document

Intrinsic Hallucinations

Argentina beat Germany 2-0

Argentina emerged on Saturday

Extrinsic Hallucinations

Source Text: Germany emerge victorious in 2-0 win against Argentina on Saturday

Summary: Germany beat Argentina 2-0

- ❖ Model adds information not inferable from the input document

Extrinsic Hallucinations

Germany beat Argentina at **home**

Germany beat Argentina **on the moon**

Argentina lost **the Copa America**

Factuality in Hallucinated Content

Source Text: Germany emerge victorious in 2-0 win against Argentina on Saturday

Summary: Germany beat Argentina 2-0

Hallucinations can be **factual**

- ❖ Model might memorize the information
- ❖ Information might be world knowledge

Germany beat Argentina at home

Factual

Research Questions

- ❖ What types of hallucinations occur?
- ❖ **How frequently do systems hallucinate?**
- ❖ Are there automatic means of measuring these hallucinations?

Large Scale Annotation

10 min per rating in average

- ❖ Raters are given the news article and the system summary
- ❖ They are tasked with selecting spans that aren't supported by the input

The screenshot shows the BBC News website interface. The main headline is "Zac Goldsmith chosen as Conservative London mayoral candidate" dated 2 October 2015. The article text includes: "Zac Goldsmith will contest the 2016 London mayoral election for the Conservatives, it has been announced." "The Richmond Park and North Kingston MP said he was 'honoured' after winning 70% of the 9,227 votes cast using an online primary system." "He beat London Assembly Member Andrew Boff, MEP Syed Kamall and London's deputy mayor for crime and policing Stephen Greenhalgh." "Mr Goldsmith's main rival is likely to be Labour's Sadiq Khan." "Mr Khan beat former cabinet minister Tessa Jowell and a number of other current MPs to win the Labour nomination last month." "Caroline Pidgeon is the Lib Dem candidate, Sian Berry will contest the election for the Greens and UKIP has chosen its culture spokesman Peter Whittle. Former Respect MP George Galloway is also running." "Mr Goldsmith, who was the favourite for the Tory nomination, balloted his constituents earlier this year to seek permission to stand."

Conservative MP Zac Goldwin's bid to become Labour's candidate in the 2016 London mayoral election.
Hallucination x Conservative MP Zac Goldwin's bid to become Labour's candidate in the 2016 London mayoral election.
The span cannot be inferred from the document because: <ul style="list-style-type: none"><input type="radio"/> It introduces new concepts that are not presented in the document.<input checked="" type="radio"/> It wrongly presents concepts from the document.
Hallucination x Conservative MP Zac Goldwin's bid to become Labour's candidate in the 2016 London mayoral election.
The span cannot be inferred from the document because: <ul style="list-style-type: none"><input checked="" type="radio"/> It introduces new concepts that are not presented in the document.<input type="radio"/> It wrongly presents concepts from the document.
Hallucination x Conservative MP Zac Goldwin's bid to become Labour's candidate in the 2016 London mayoral election.
The span cannot be inferred from the document because: <ul style="list-style-type: none"><input checked="" type="radio"/> It introduces new concepts that are not presented in the document.<input type="radio"/> It wrongly presents concepts from the document.

Results

Model	Intrinsic ↓	Extrinsic ↓	Faith. ↑	+Fact. ↑
PTGen	19.9	63.3	24.7	27.3
TConvS2S	17.7	71.5	21.5	26.9
Transformer	19.1	68.1	20.7	25.3
BertS2S	16.9	64.1	26.9	34.7
Gold	7.4	73.1	23.1	-

Pretraining improves factuality and faithfulness

Best system is faithful **only** 27% of the time

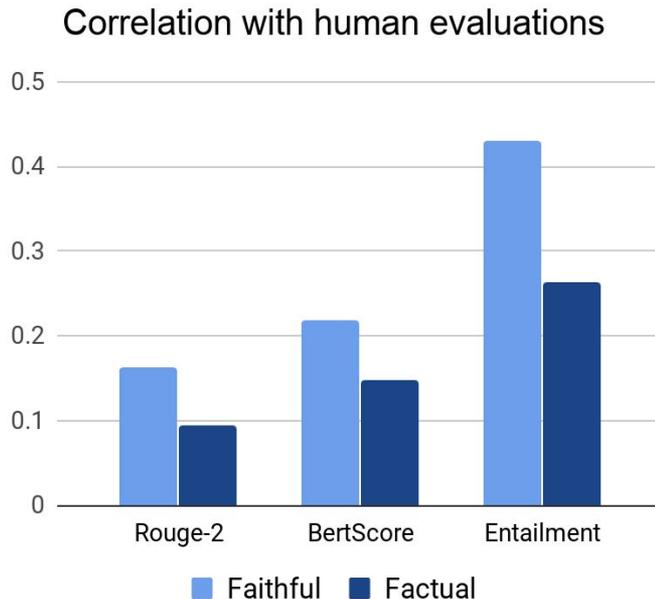
Best system is faithful or factual 35% of the time

Research Questions

- ❖ What types of hallucinations occur?
- ❖ How frequently do systems hallucinate?
- ❖ **Are there automatic means of measuring these hallucinations?**

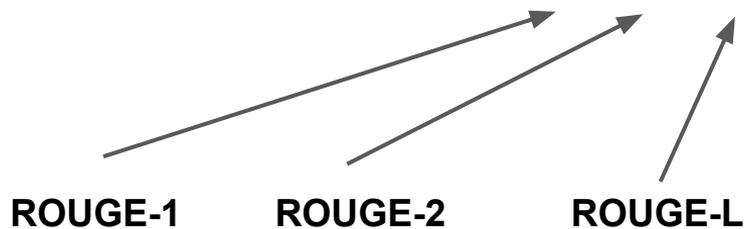
Metrics Correlation

- ❖ Ngram overlap metrics (e.g. Rouge) don't correlate well
- ❖ Quality learned metrics (e.g. BertScore) don't correlate well
- ❖ Entailment scores are better correlated with faithfulness



ROUGE prefers coverage/fluency over faithfulness

GOLD	Zac Goldsmith will contest the 2016 London mayoral election for the Conservatives, it has been announced.	
PTGEN	UKIP leader Nigel Goldsmith has been elected as the new mayor of London to elect a new Conservative MP.	[45.7, 6.1, 28.6]
TCONVS2S	Former London mayoral candidate Zac Goldsmith has been chosen to stand in the London mayoral election.	[50.0, 26.7, 37.5]
TRANS2S	Former London mayor Sadiq Khan has been chosen as the candidate to be the next mayor of London.	[35.3, 12.5, 23.5]
GPT-TUNED	Conservative MP Zac Goldwin's bid to become Labour's candidate in the 2016 London mayoral election.	[42.4, 25.8, 36.4]



Challenges for Trustworthy Neural Generation

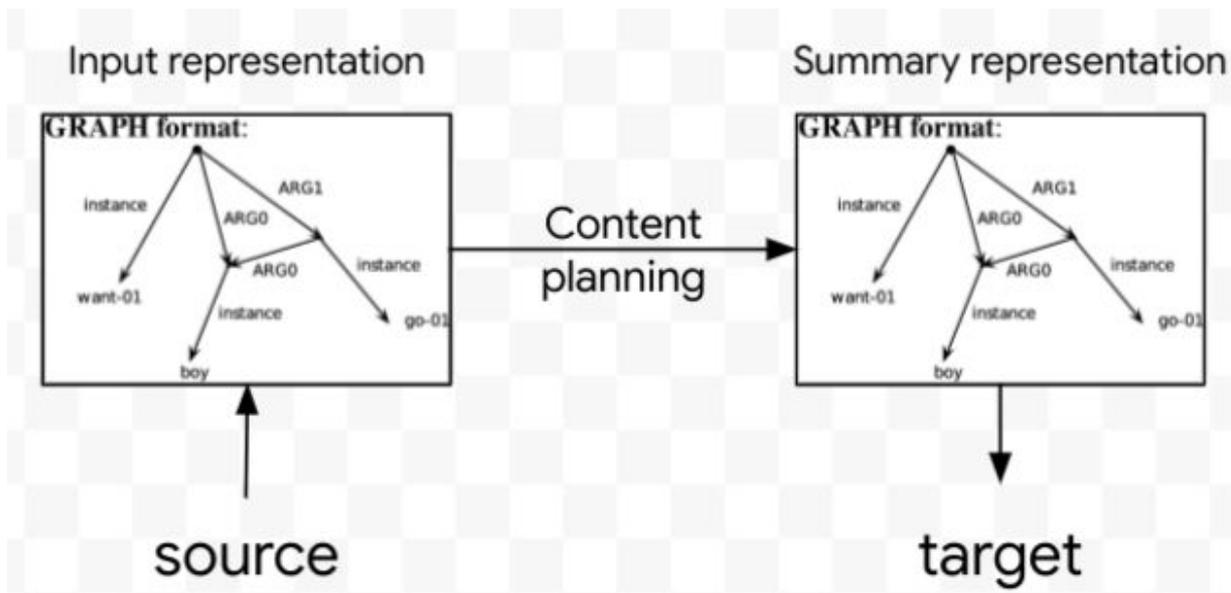
**Evaluation and
Benchmarking**

Understand and measure types of
hallucinations

**Controllable
Models**

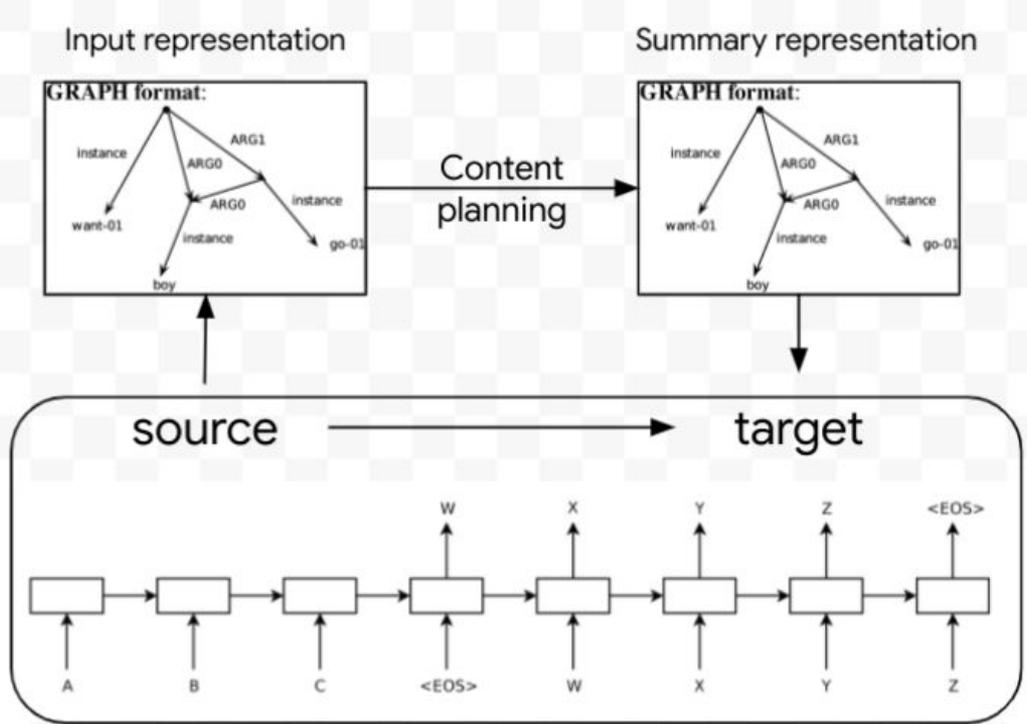
**Planning-based Controllable and
Faithful Generation**

Traditional NLG Systems were more Controlled and Grounded

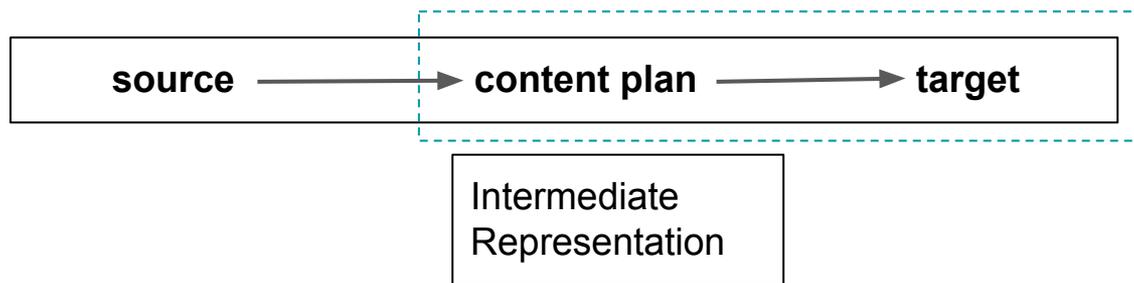
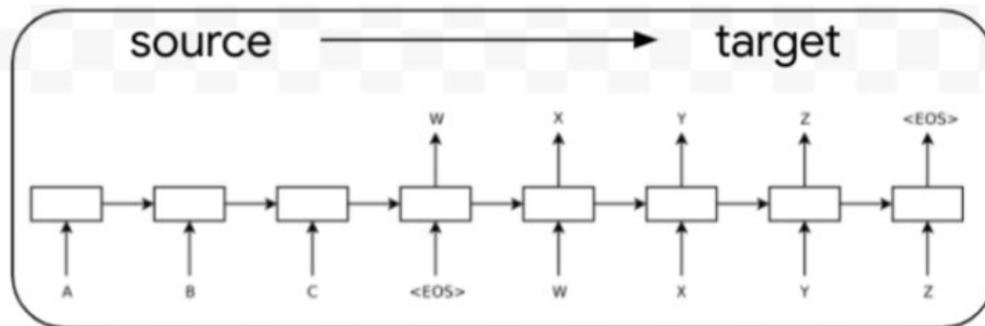


See Jones (1993) and Reiter and Dale (2000).

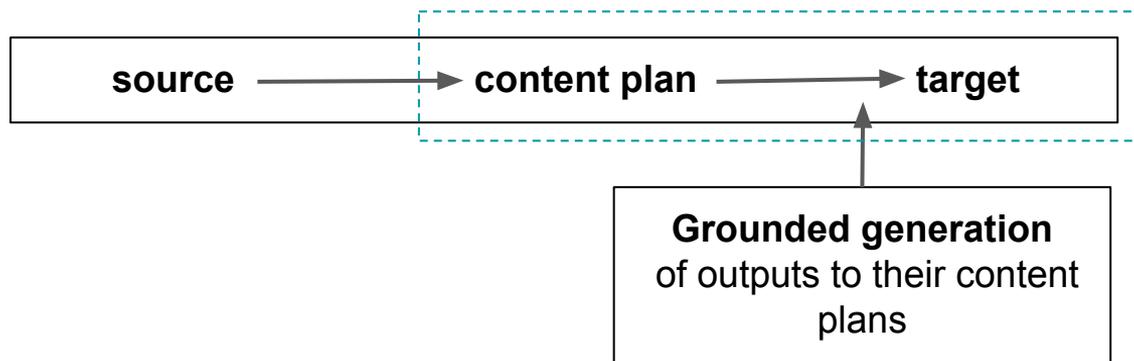
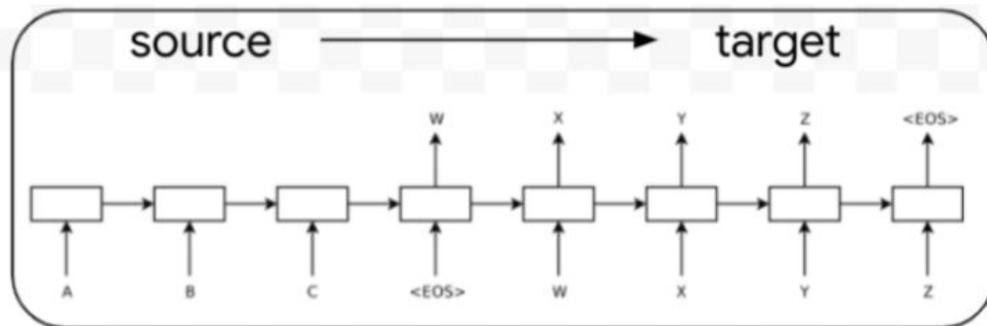
But Neural Generation is not Ground in the Same Way



Planning for Trustworthy Generation



Planning for Trustworthy Generation

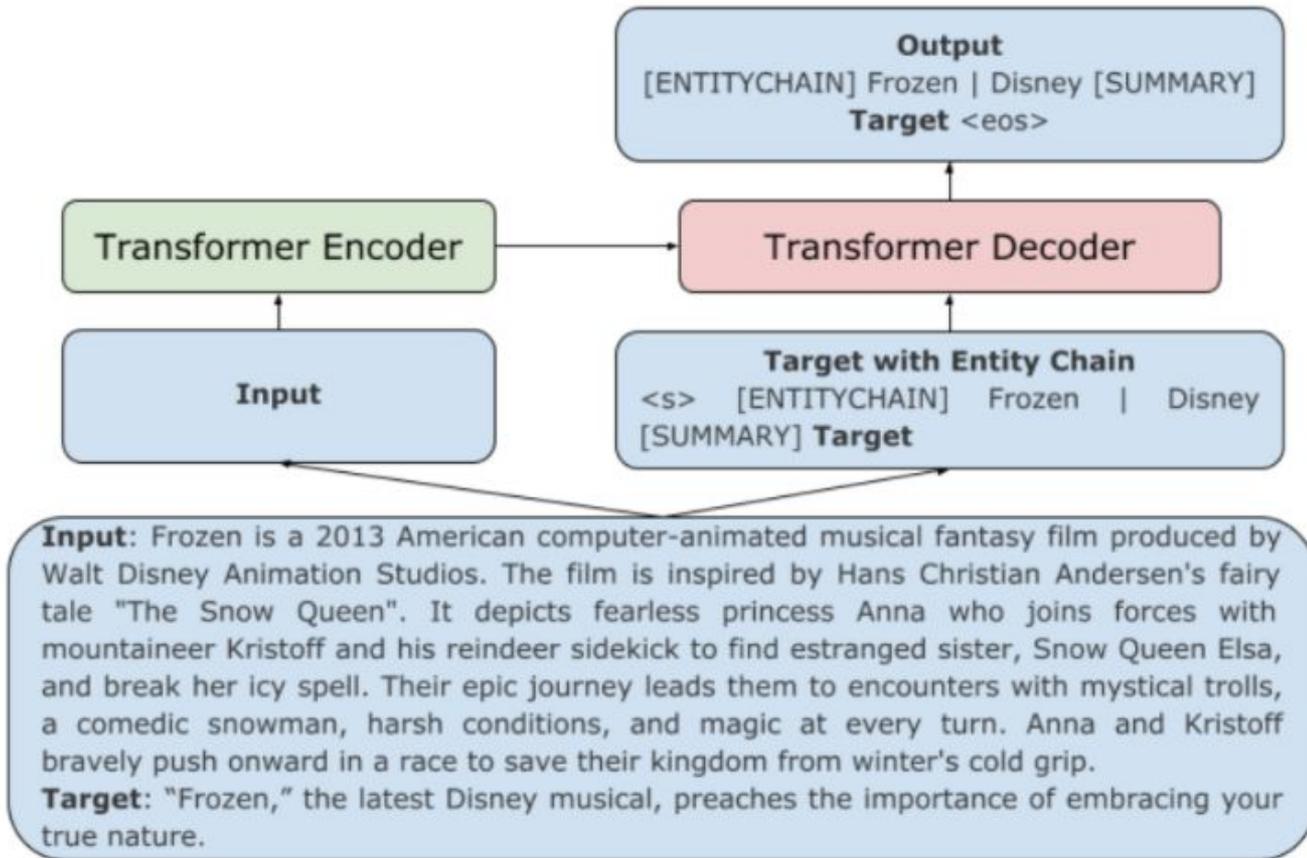


Controllable Entity-Based Planning and Grounding for Generation

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



Supervised Training with Entity Chains



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.

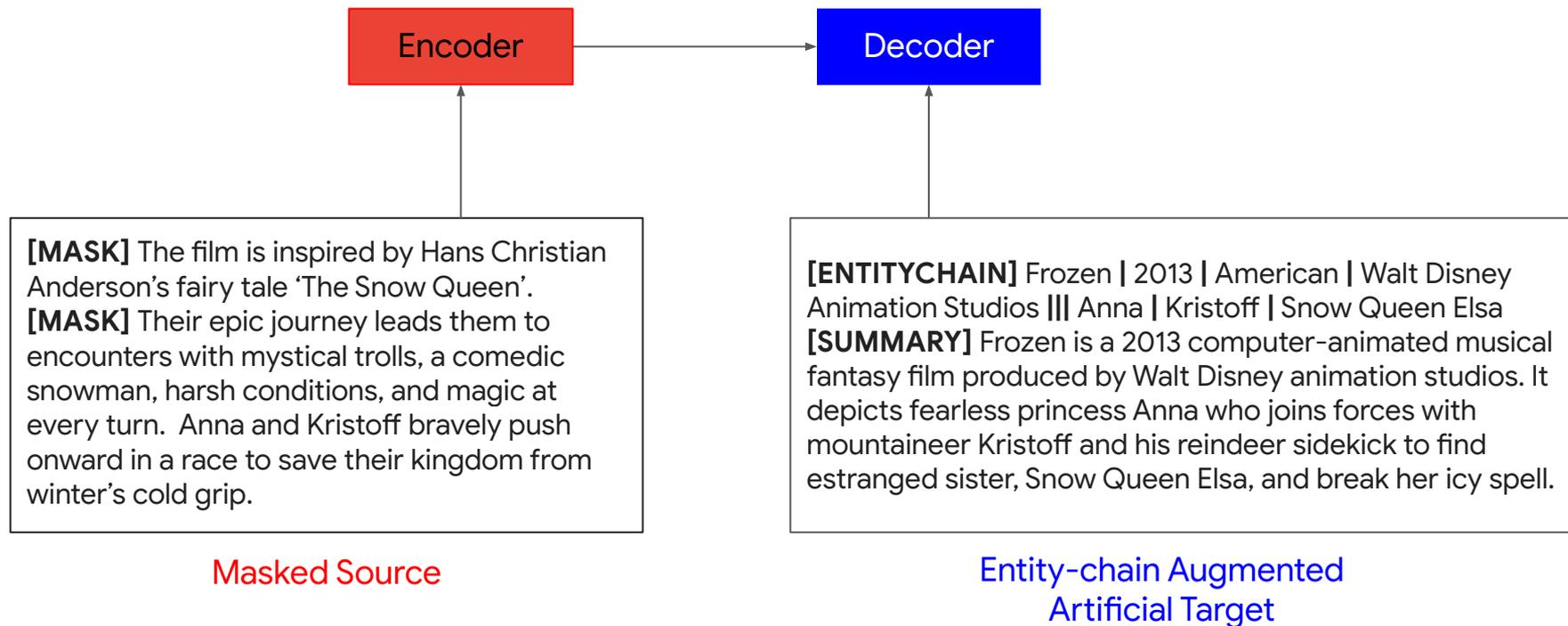
Masked Source

[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.

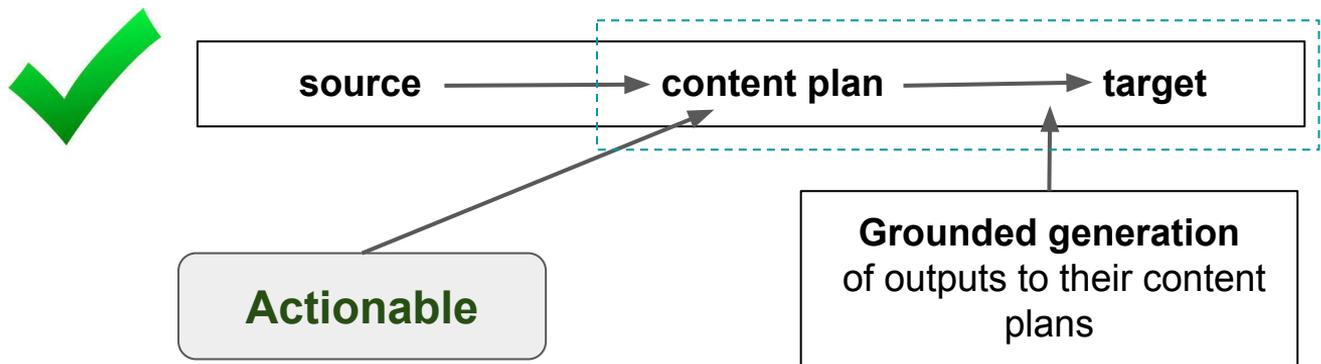
Entity-chain Augmented
Artificial Target

Pretraining with Entity Chain Planning

Can be used in pre-training:



Planning for Trustworthy Generation

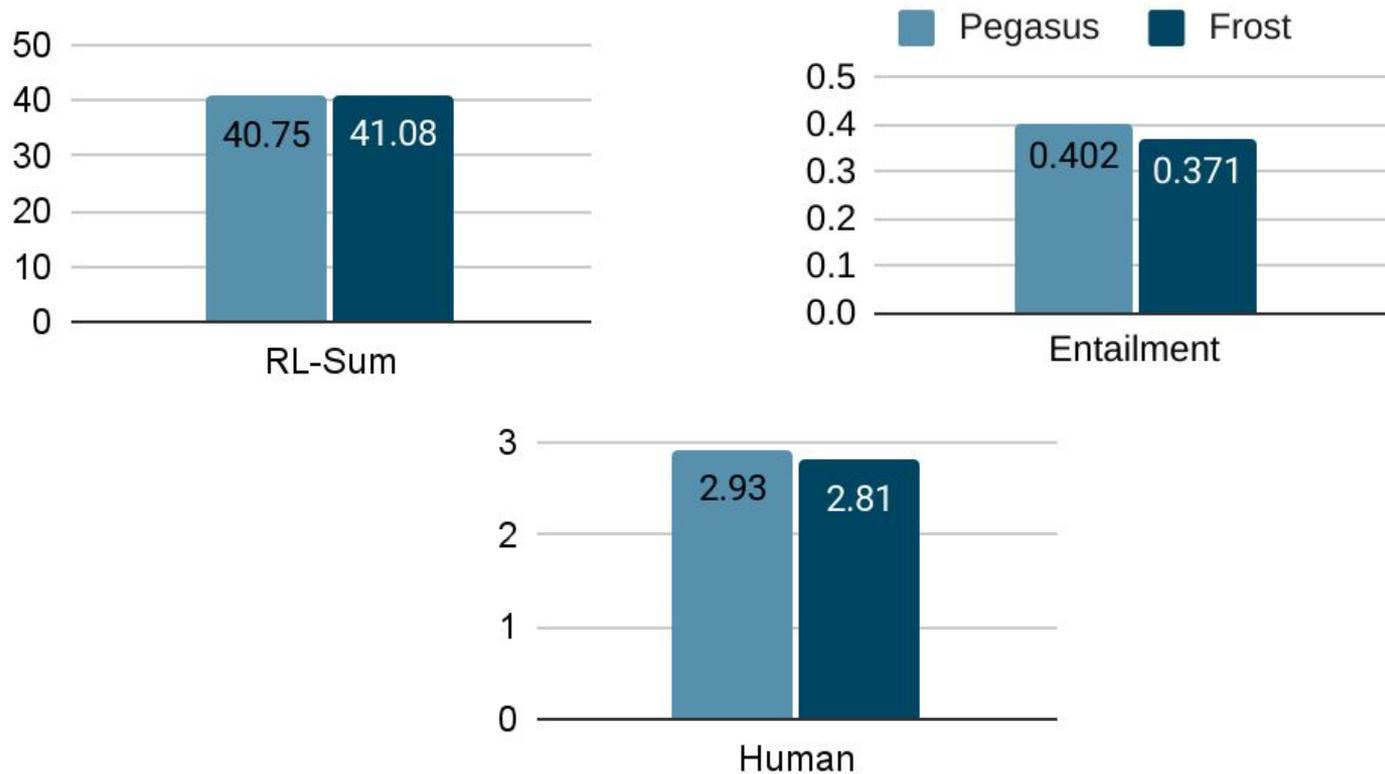


Improved explainability and controllability

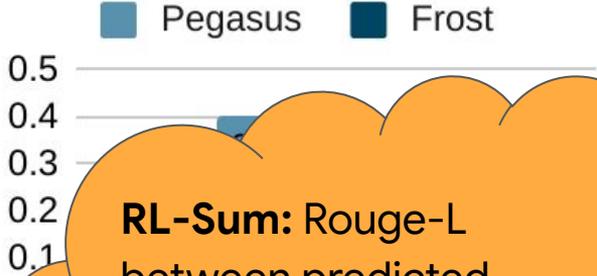
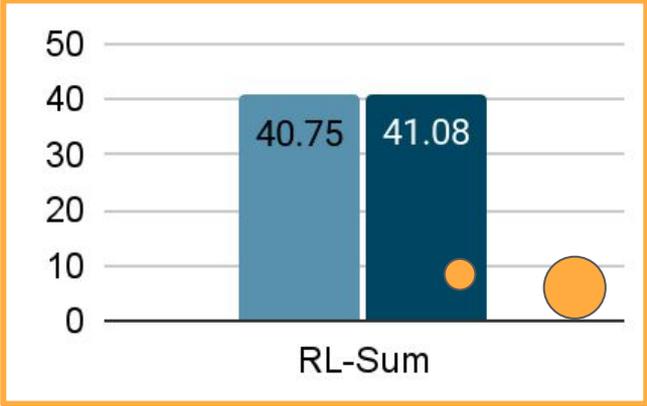
Many advantages:

1. **No change in architecture**, no additional annotations required during inference. Plans can be fully automatically generated by decoder.
2. **Controllable and Faithful Generation**: Can be semi-automatically generated by correcting hallucinated entities. Or, can be manually controlled.
3. Can be used to generate diverse summaries.

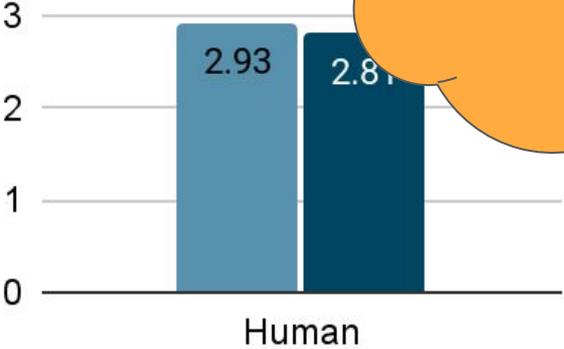
Rouge and Faithfulness Results



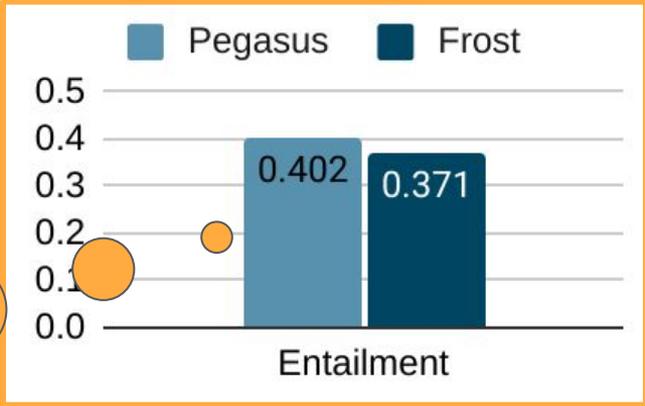
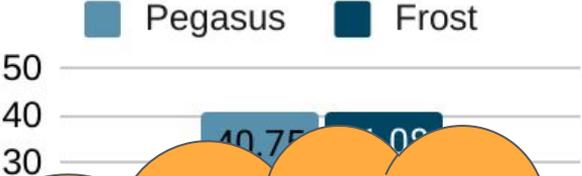
Rouge and Faithfulness Results



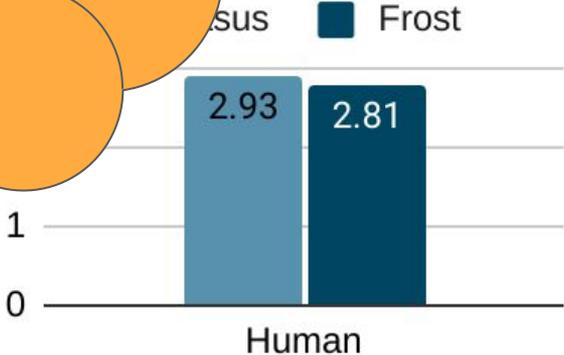
RL-Sum: Rouge-L between predicted summary and reference summary



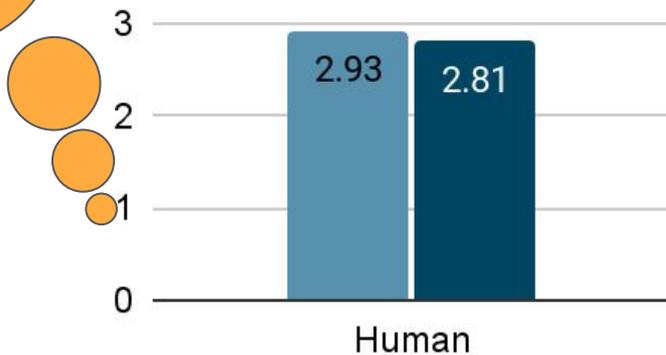
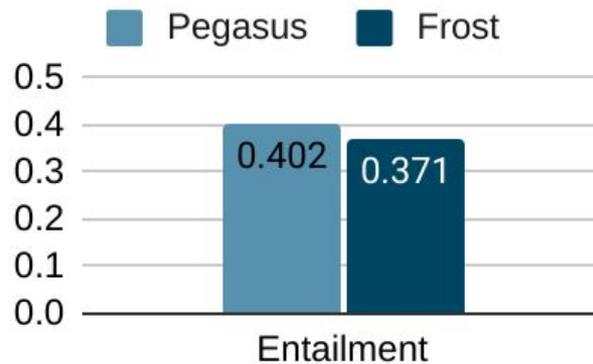
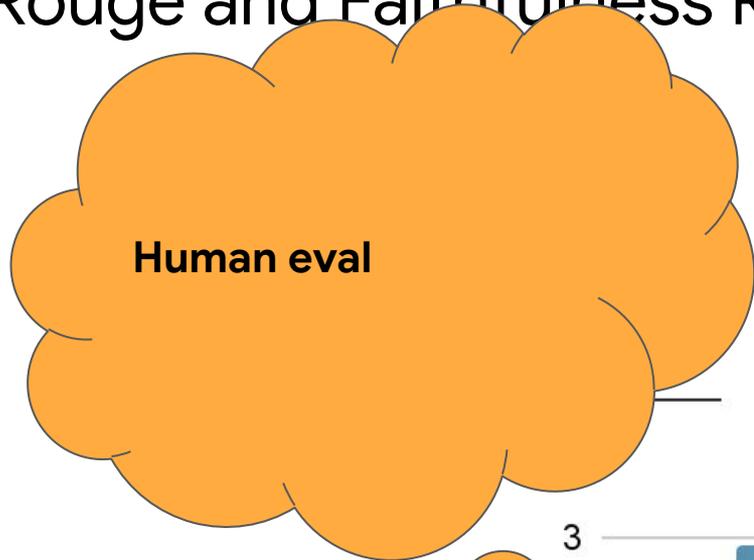
Rouge and Faithfulness Results



Entailment:
 $p(\text{document entails summary})$



Rouge and Faithfulness Results



Faithfulness Results: Human Eval Template

Instructions: Read a news article and rate summaries on a scale of 1 (fully unfaithful) to 5 (fully faithful). A summary is "fully faithful" if all of its content is fully supported or can be inferred from the input article.

- Fully unfaithful: None of the content is supported by the input document.
- Somewhat unfaithful: Except some parts, most of the content in the summary is not supported by the input article. If chosen, **please explain what was faithful in the summary.**
- 50-50: Approximately half of the content is supported by the input article. If chosen, **please explain what was faithful and what was unfaithful in the summary.**
- Somewhat faithful: Except some parts, most of the content in the summary is supported by the input article. If chosen, **please explain what was unfaithful in the summary.**
- Fully faithful: All contents are supported by the input article.

Please refer this [guidelines](#) document for detailed instructions and examples

Article

Highlight

Unhighlight

David Vallenilla's 22-year-old son, also called David, was shot by military police on Thursday. Mr Vallenilla says he once worked with the president on Caracas' transport system, and called on his "former colleague" to clear his son's name. The president has said firearms must not be used on protesters. "I want to speak to my former colleague of the Caracas Metro," Mr Vallenilla told a group of reporters outside the morgue. "Nicol ?? s Maduro, you know that we worked together, I'm Supervisor Vallenilla." He said he used to be Mr Maduro's boss, although the past working relationship between them has not yet been confirmed. President Maduro is a former bus driver, who got involved in trade unions and then national politics, becoming president in 2013 after his predecessor Hugo Chavez died. "Please, Nicol ?? s, my friend," Mr Vallenilla said with tears streaming down his face, before calling on the leader to make it clear that his only son "was not a criminal, he was a graduating student". "You met him when he was little," he added. Earlier in the week, Fabi ?? n Urbina, 17, was also shot and killed as security forces and demonstrators clashed in Caracas. There have been almost daily anti-government protests in Venezuela for over two months as the country's economic and political crisis has worsened. More than 70 people have been killed in protest-related violence since 1 April, according to figures released by the chief prosecutor's office. On Thursday, Interior Minister Néstor Reverol tweeted to confirm the death of a protester and said a police sergeant had fired an "unauthorised weapon".

Summary 1:

The father of a Venezuelan protester shot dead by security forces has made an emotional plea to President Nicol ?? s Maduro outside a morgue in Caracas.

Please rate the summary on a scale 1-5:

Fully unfaithful Somewhat unfaithful 50-50 Somewhat faithful Fully faithful

Please give 1-2 sentence explanation of your choice:

Previous

Next

Frost

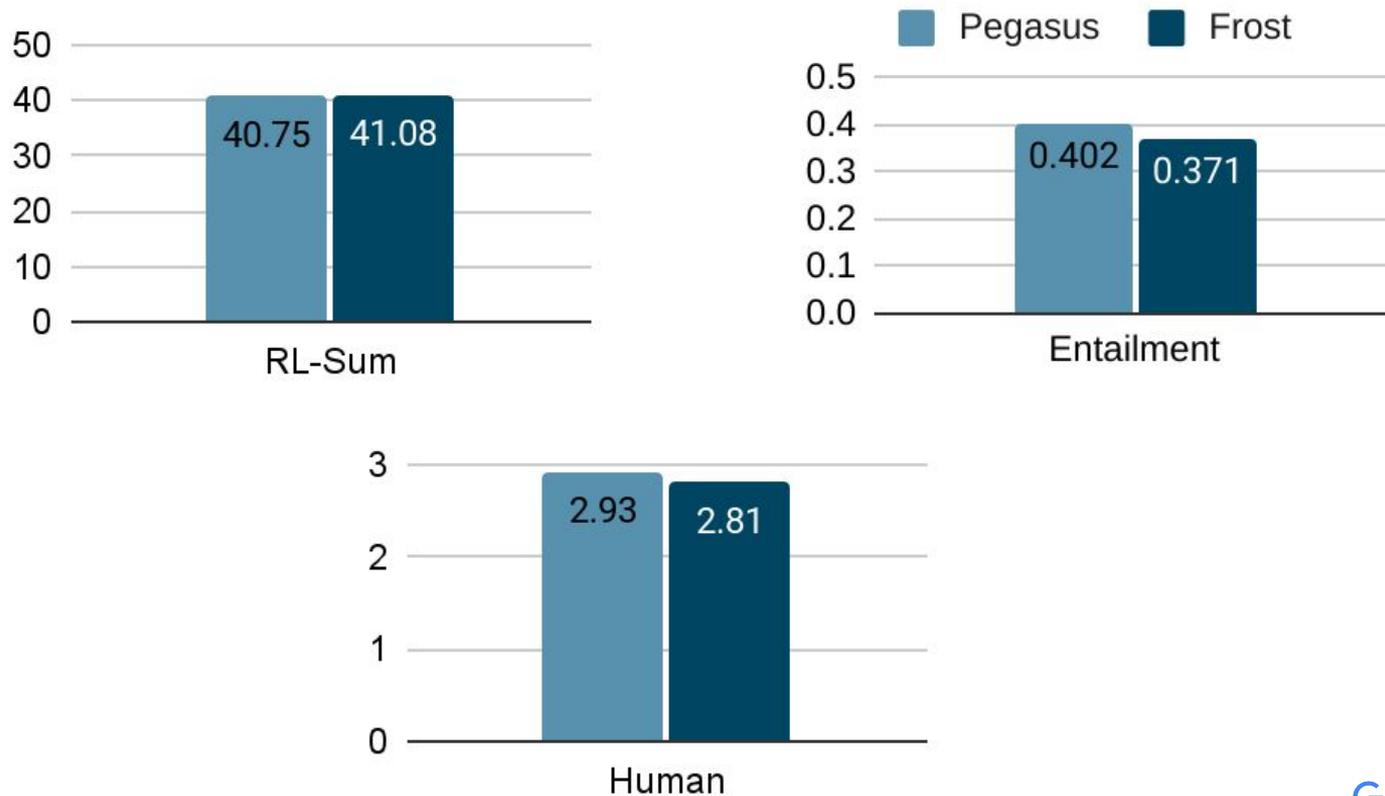
0.371

ent

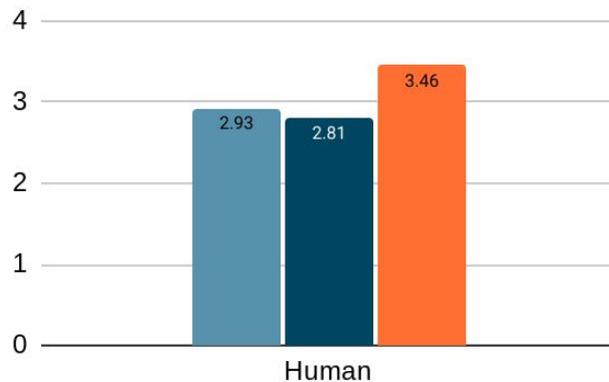
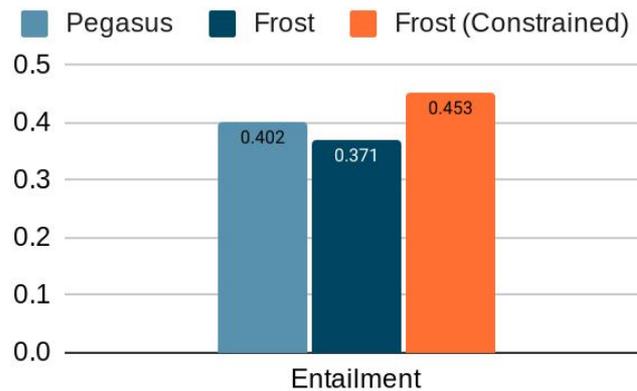
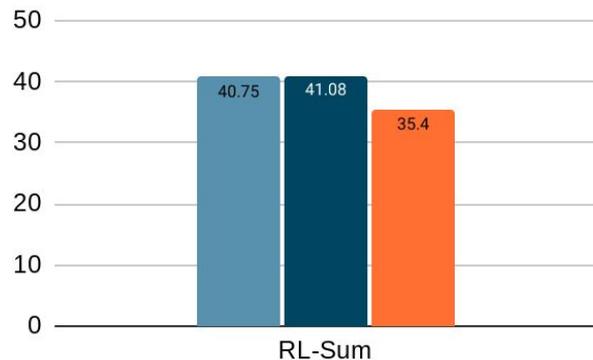
2.81

an

Planning Itself Does not Help with Faithfulness



But Planning allows Controllable and Faithful Generation



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 .**

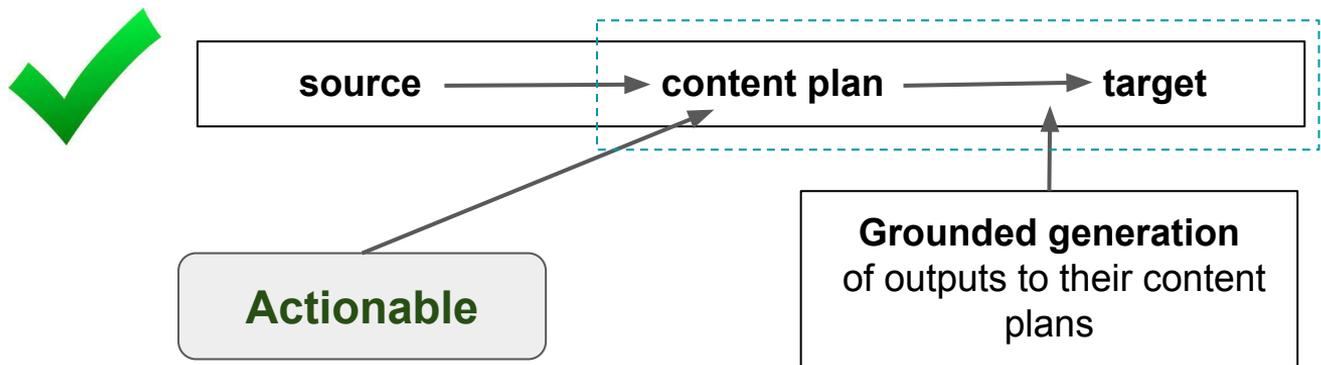


Average lengths: target (60.7), Pegasus (68.6) and Frost (65.0)

Frost models Generate Summaries with Specific Details

Pegasus (d -> t)	Frost	Target
<p>BBC One's new drama about a family dealing with a child's diagnosis of autism has been met with a mixture of shock and relief.</p>	<p>A new BBC One drama, <i>The A Word</i>, tells the story of a family dealing with a child's diagnosis of autism.</p>	<p>New TV drama <i>The A Word</i> tells the story of a family discovering that their quirky five-year-old son is in fact autistic.</p>
<p>A storm-hit Jersey beach is to be inspected before it is reopened to the public, the Department of Infrastructure has said.</p>	<p>A Guernsey beach is to be inspected to see if it is safe to reopen, the Department of Environment has said.</p>	<p>Steps at a bay in Guernsey could be opened by Easter, according to the environment department.</p>

FROST: Conclusion



Improved explainability and controllability

Many advantages:

1. **No change in architecture**, no additional annotations required during inference. Plans can be fully automatically generated by decoder.
2. **Controllable and Faithful Generation**: Can be semi-automatically generated by correcting hallucinated entities. Or, can be manually controlled.

Using Frost to Generate Diverse Summaries with Different Styles and Themes

GOLD: [ENTITYCHAIN] BP | Lord Browne | North Sea [Summary] Former BP chief executive Lord Browne has warned North Sea oil operators' costs must fall in order for them to compete globally.

FROST (ECPP): [ENTITYCHAIN] Browne | BP [Summary] Lord Browne, the former boss of BP, has warned that some oil companies could go out of business if costs continue to rise.

FROST (ECPP), c_{mod} , Style diversity
[ENTITYCHAIN] BP | Browne [Summary] The former boss of BP, Lord Browne, has warned that some oil companies could go out of business if costs continue to rise.
[ENTITYCHAIN] BBC | BP | Browne [Summary] In an interview with the BBC, former BP chief executive Lord Browne has warned the oil and gas industry is in danger of "going to the wall".

FROST (ECPP), c_{mod} , Topical diversity
[ENTITYCHAIN] Brent Crude [Summary] The former boss of the world's biggest oil company has warned that the price of Brent Crude will not recover until next year.
[ENTITYCHAIN] Brent Crude | Unions [Summary] Oil companies will have to cut costs dramatically if they are to survive the fall in the price of Brent Crude, according to the former chief executive of the Unions.

NLG Tasks often Exhibit One-to-Many Relationships

Summarization

Chelsea's Eden Hazard and Arsenal's Santi Cazorla are set to reach a Premier League milestone this weekend when they each make their 100th appearance. Both players have been hugely influential since they moved to London in the summer of 2012, but who has been the most exciting import to watch? Here, Sportsmail's reporters choose the player they most enjoy seeing in action. Eden Hazard (L) and Santi Cazorla are both set to make their 100th Premier League appearance this weekend....

⇒ Eden Hazard and Santi Cazorla are both set to make their 100th Premier League appearances this weekend. Both players have been hugely influential since they moved to London. Here, Sportsmail's reporters choose the player they most enjoy seeing in action.

⇒ Eden Hazard and Santi Cazorla are set to make their 100th appearances for Chelsea and Arsenal respectively in the Premier League this weekend. Both players have been hugely influential since they moved to London in the summer of 2012. But who has been the most exciting import to watch?

⇒ Chelsea's Eden Hazard and Arsenal's Santi Cazorla will both make 100th appearances this weekend. Sportsmail's reporters pick the player they most enjoy seeing in action. Both players have been hugely influential since moving to London.

Question Generation

On Tesla's 75th birthday in 1931, Time magazine put him on its cover. The cover caption "All the world's his power house" noted his contribution to electrical power generation. He received congratulatory letters from more than 70 pioneers in science and engineering, including Albert Einstein.

⇒ Who appeared on Time magazine's cover on his 75th birthday?

⇒ Which famous scientist was in the cover of Time Magazine in 1931?

⇒ Which mad scientist received more than a 70 people congratulating him on his birthday?

⇒ What famous scientist was also 75?

Diversity: It is important to be able to generate multiple diverse outputs from a model.

Diversity should not Come at the expense of Correctness

Summarization

Chelsea's Eden Hazard and Arsenal's Santi Cazorla are set to reach a Premier League milestone this weekend when they each make their 100th appearance. Both players have been hugely influential since they moved to London in the summer of 2012, but who has been the most exciting import to watch? Here, Sportsmail's reporters choose the player they most enjoy seeing in action. Eden Hazard (L) and Santi Cazorla are both set to make their 100th Premier League appearance this weekend...

⇒ Eden Hazard and Santi Cazorla will each make their 100th Premier League appearance this weekend. [nightstandapplication.com](https://www.nightstandapplication.com). Sportsmail's hovercraft reporters choose their man of the match countermeasures.

⇒ By making their 100th Premier League appearances this weekend, Eden Hazard and Santi Cazorla will set new records. Here, Anna Coren and Dominic King select their favourites.

Alternate summaries are not valuable if they are unfaithful to the input document(s)

(Maynez et al., 2020; Kryscinski et al., 2020)

Standard Sampling Methods Degenerate Text, Unfaithful or Inconsistent to the Input

Anna Petrovna of Russia was wife of King William II of the Netherlands

Sampling Methods (e.g., top-p or Nucleus Sampling)

Anna Pavlovna of Russia was a queen consort of the Netherlands by marriage to king William II.

Beam Search



This Work: Composition Sampling for Diverse Conditional Generation

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

Nucleus Sampling

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

Beam Search



Anna Petrovna of Russia was wife of King William II of the Netherlands

Sampling Methods (e.g., top-p or Nucleus Sampling)



Anna Pavlovna of Russia was a queen consort of the Netherlands by marriage to king William II.

Beam Search



This Work: Composition Sampling for Diverse Conditional 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**



*“A Well-Composed Painting is Half Done”:
French painter Pierre Bonnard*

Compositional sampling

- combines planning with stochastic sampling
- samples a composition (or plan) in the form of an entity chain and then uses beam search to generate the best possible sequence grounded to the sampled entity chain
- avoids degeneration by instilling diversity in composition, rather than directly on the surface form

Composition Sampling: Hybrid Decoding with Planned Generation



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

Nucleus Sampling

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

Beam Search



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

Nucleus Sampling

[SUMMARY] Anna Petrovna of Russia was wife of King William II of the Netherlands

Nucleus Sampling



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

Beam Search

[SUMMARY] Anna Pavlovna of Russia was a queen consort of the Netherlands by marriage to king William II.

Beam Search

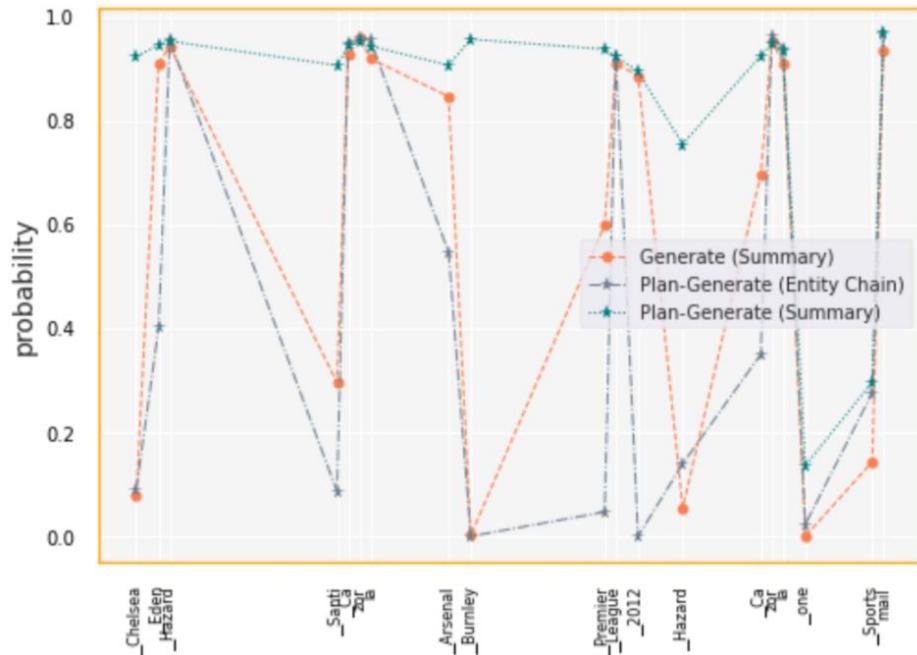


Why is Planning Useful for Diverse Generation?

Grounded generation: once the hard part of planning the summary composition is done, the model is much less perplexed during the generation of the summary

Pegasus	$p(y x)$:	perplexity:	1.478
Frost	$p(y h;x)$:	perplexity:	1.15
Frost	$p(h;y x)$:	perplexity:	1.16

*“Chelsea star **Eden Hazard** is set to make his 100th top-flight appearance. **Santi Cazorla** should hit the same milestone when **Arsenal** meet **Burnley**. Both players have impressed since moving to the **Premier League** in 2012. **Hazard** has more goals this season but **Cazorla** has **one** more assist. **Sportsmail**’s reporters choose the player who has excited them the most.”*



Why is Planning Useful for Diverse Generation?

Entity chains are good at separating content from surface form

- Learning $p(h | x)$ is much easier than learning $p(y | x)$ - so we can sample from h
- We can then sample from $p(h | x)$, and compute $\operatorname{argmax}_y p(y | x, h)$

Pegasus $p(y|x)$: neg_log_likelihood: 121.18

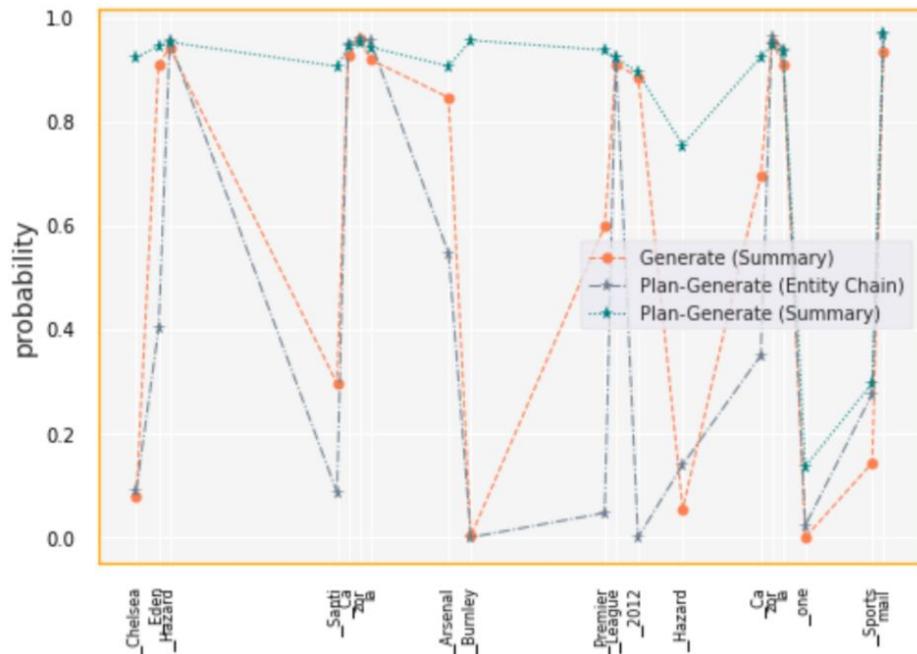
Frost $p(h|x)$: neg_log_likelihood: 46.95

Frost $p(y|h;x)$: neg_log_likelihood: 95.08

In top 15 token for Pegasus vs FROST

Avg Nucleus Prob: 0.2830 vs 0.4332

*“Chelsea star **Eden Hazard** is set to make his 100th top-flight appearance. **Santi Cazorla** should hit the same milestone when **Arsenal** meet **Burnley**. Both players have impressed since moving to the **Premier League** in 2012. **Hazard** has more goals this season but **Cazorla** has **one** more assist. **Sportsmail**’s reporters choose the player who has excited them the most.”*



Nucleus Sampled CNN/DailyMail Summaries

Nucleus Sampling: PEGASUS

s_1 → Nick Clegg reveals **balance902** of power lies with his wife Miriam Gonzalez Durantez **irish**. Couple performed joint interview **to promote new series** on ITV News Tonight. Cleggs spoke of pressure of moving from their Putney home into a taxpayer-funded mansion.

s_3 → Lib Dem leader admitted wife Miriam has the final say over family life. Couple chose not to move their three **Laundry** to Whitehall home **earlier this May**.

s_4 → Nick Clegg and his wife Miriam Gonzalez Durantez open up in TV interview. Lib Dem leader revealed **she Bloomberg-style 'discussions'** in their home. Couple revealed they **opted not to stay with their sons** in their £1.5m house.

s_5 → Liberal Democrats leader revealed balance of power lies **30-plus metres away**. He brought cameras into family home due to Cameron and Miliband controversies. Lib Dem leader joked that wife Miriam vetoed their move to Whitehall.

Nucleus Sampling: FROST

$c_1; s_1$ → [CONTENT] Liberal Democrats | Nick Clegg | Miriam Gonzalez Durantez ||| Putney | London ||| Cleggs ||| ITV ||| Couple [SUMMARY] Liberal Democrats leader Nick Clegg reveals balance of power with wife Miriam Gonzalez Durantez in joint interview. They invited cameras into kitchen of £1.5million family home in Putney, south west London. Cleggs are seen trying white wine as they discuss family life and **girlfriends**. They were **Furness on ITV programme** and said they chose home to protect family. Couple say choosing home stopped them veering off from wider society 'in a bubble'

Composition Sampled CNN/DailyMail Summaries

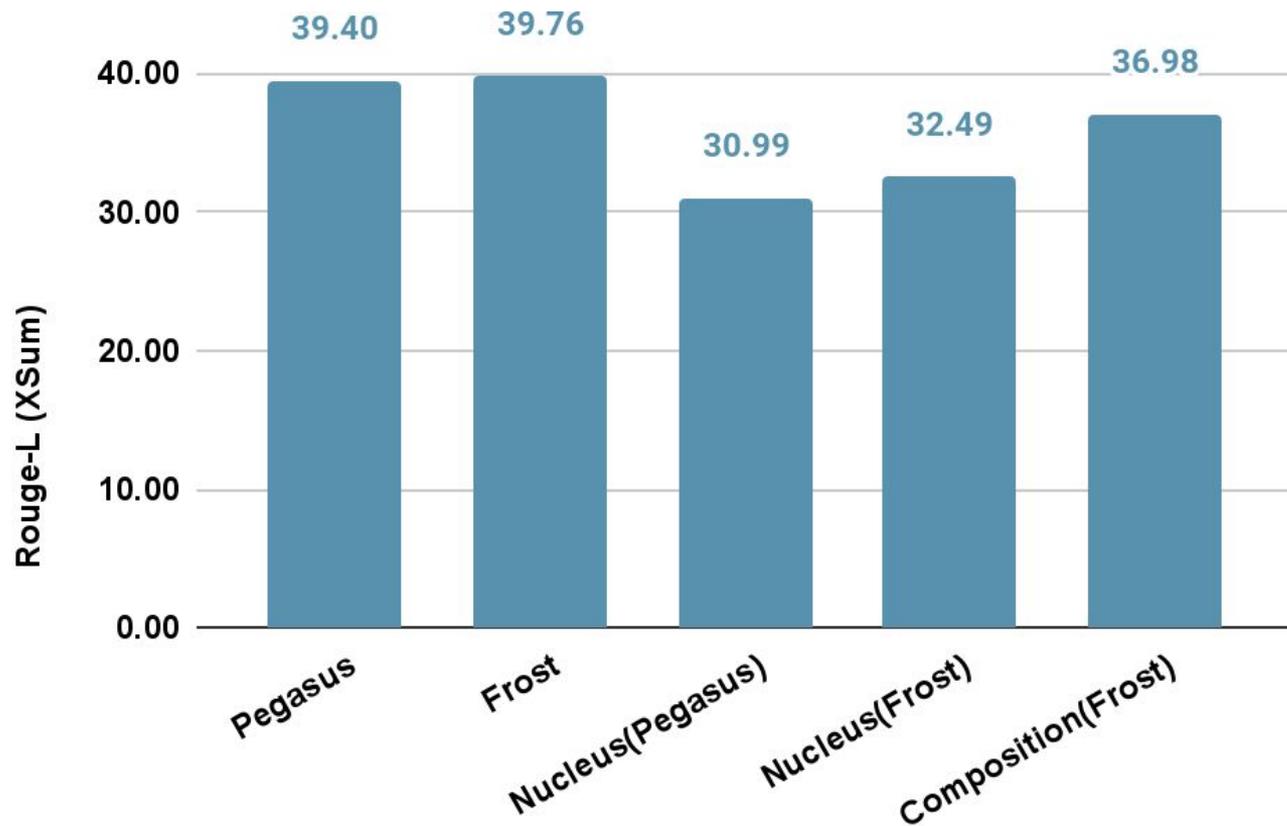
Composition Sampling: FROST

$c_1; s_1 \rightarrow$ [CONTENT] Liberal Democrats | Nick Clegg | Miriam Gonzalez Durantez ||| Putney | London ||| Cleggs ||| ITV ||| Couple [SUMMARY] Liberal Democrats leader Nick Clegg reveals balance of power lies with wife Miriam Gonzalez Durantez. The couple invited the cameras into the kitchen of their £1.5million family home in Putney, south west London. Cleggs were seen drinking white wine and cooking in the kitchen as they chatted about their family life. Focus groups conducted by the ITV programme found he was characterised as the 'chihuahua in a handbag' of the government. Couple believe that choosing to remain in their family home has stopped them from becoming cut off from the rest of the world.

$c_2; s_2 \rightarrow$ [CONTENT] Lib Dem | ITV | Tonight | Miriam Gonzalez Durantez ||| ||| Couple | Putney | London [SUMMARY] Lib Dem leader appeared on ITV's Tonight programme with wife Miriam Gonzalez Durantez. He was given the option of moving his family into a grace-and-favour government mansion - but the move was vetoed by his wife. Couple invited the cameras into the kitchen of their £1.5million family home in Putney, south west London.

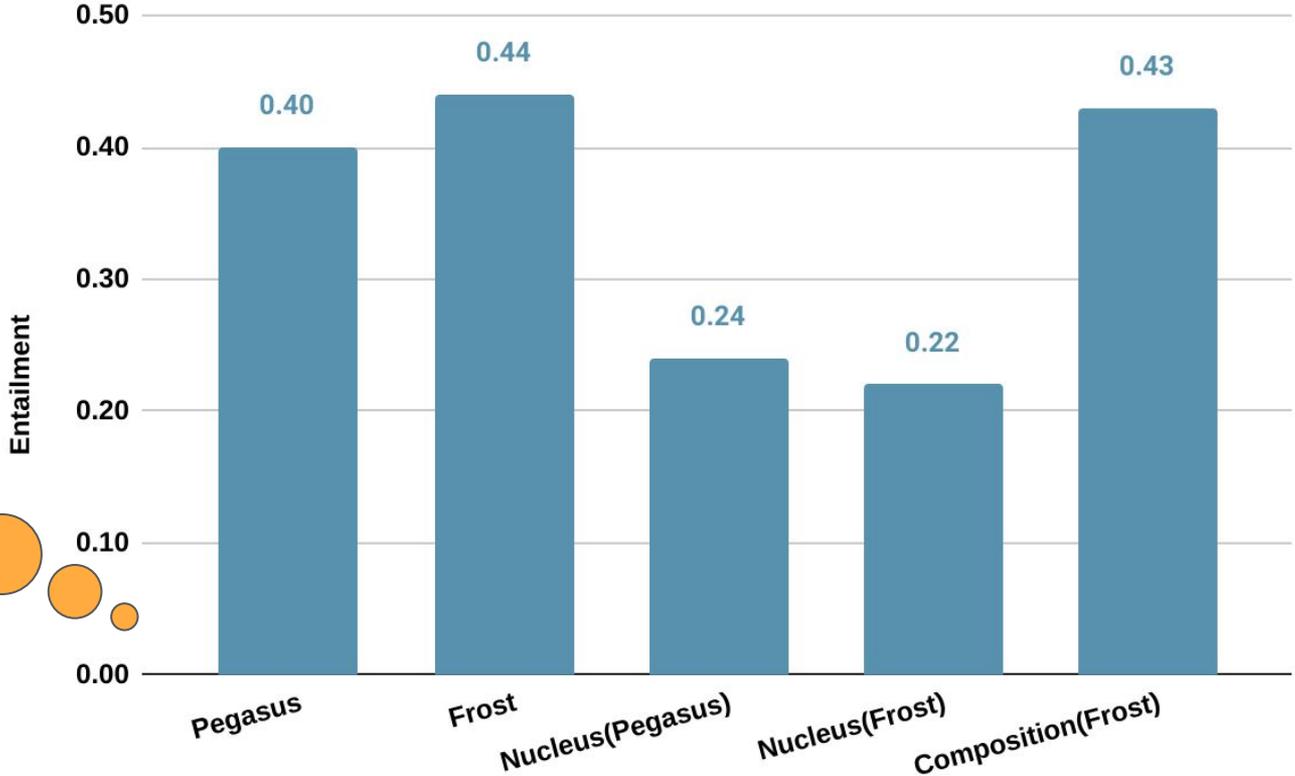
$c_3; s_3 \rightarrow$ [CONTENT] Lib Dems | Miriam ||| Couple | ITV | Tonight ||| Putney | London ||| bestseller | Miliband [SUMMARY] Lib Dems leader reveals balance of power lies with wife Miriam in joint interview. Couple invited the cameras into their kitchen for ITV's Tonight programme. They were seen drinking wine and cooking in their £1.5million home in Putney, south west London. Interview comes after bestseller's row over Miliband's small kitchen.

ROUGE: Composition Sampling is not as Performance Diminishing as Nucleus Sampling

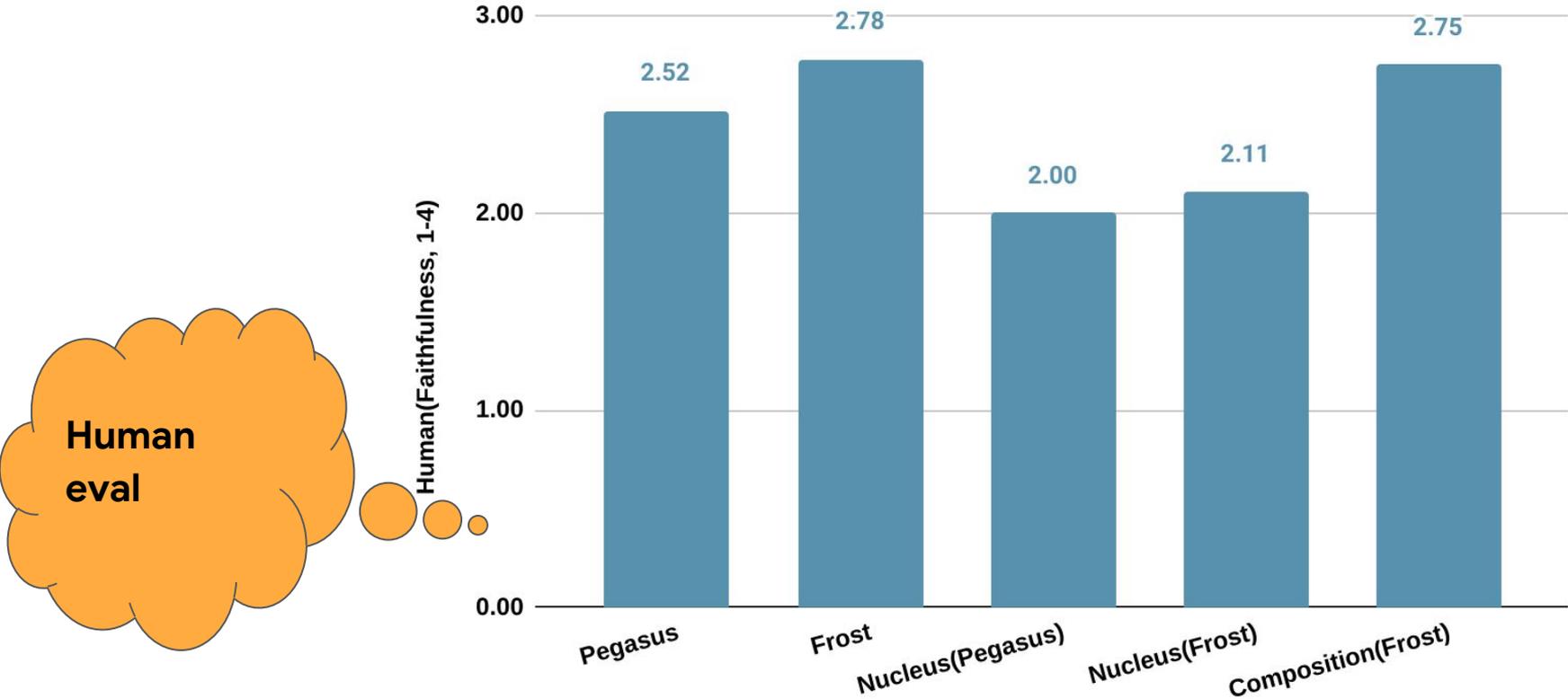


Faithfulness: Constrained Composition is Most Effective in Generating Meaningful Diverse Summaries

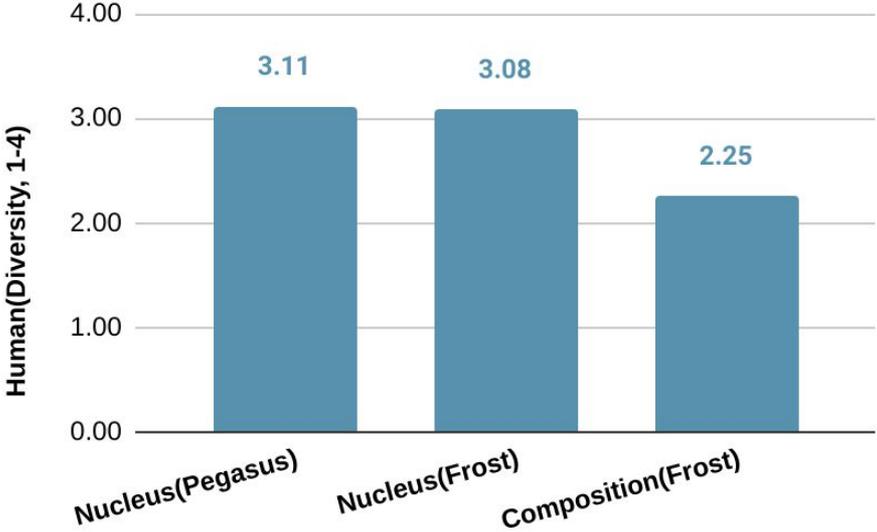
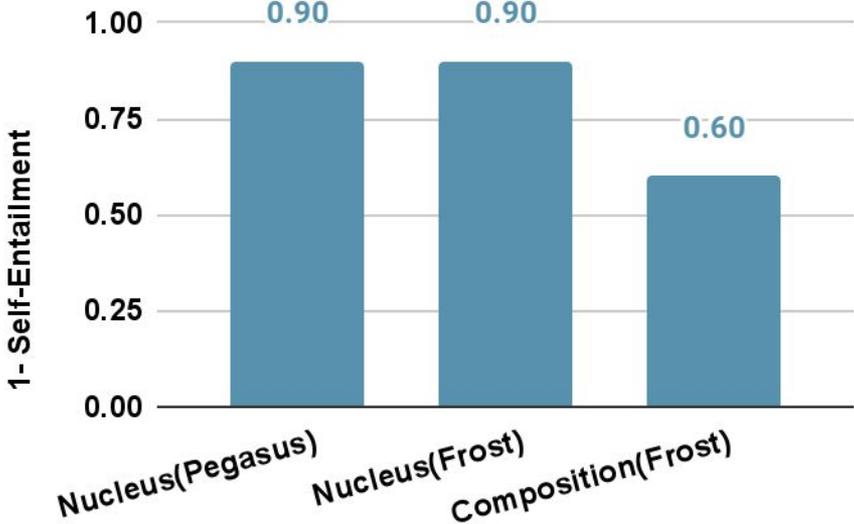
Entailment:
 $p(\text{document entails summary})$
(Maynez et al, 2020)



Human Eval for Faithfulness: Constrained Composition is Most Effective in Generating Meaningful Diverse Summaries



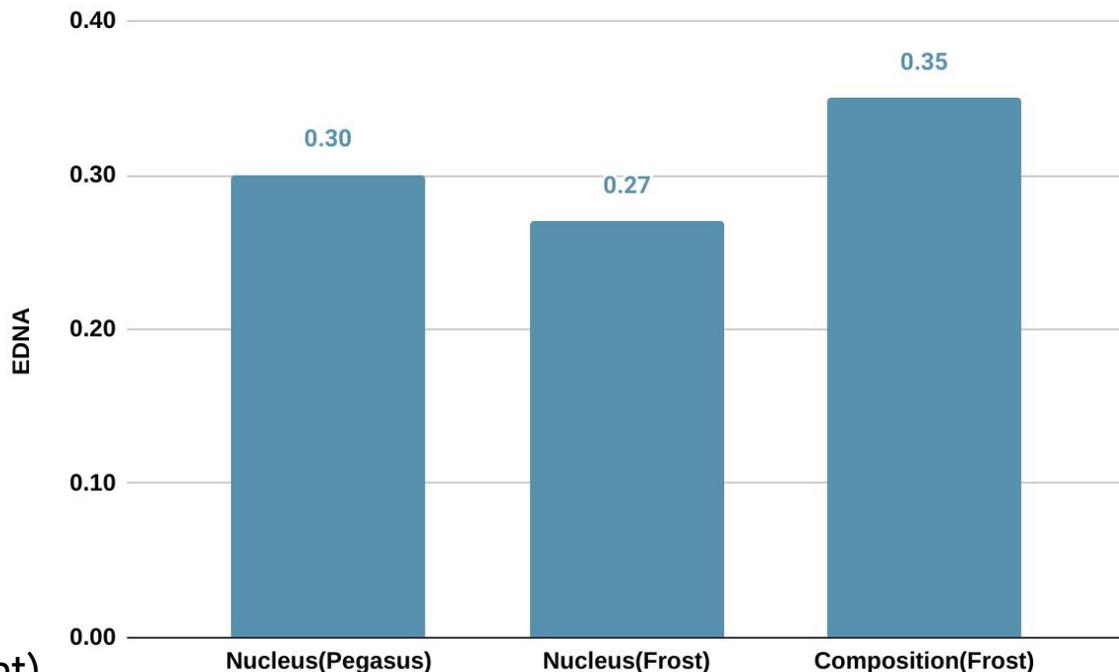
Novel Semantic Diversity Evaluations: Pairwise Entailment and BERTScore



EDNA: Novel Faithfulness and Diversity Metric

EDNA

- “Evaluating Diversity and Faithfulness” in summaries
- Harmonic mean between Entailment and (1-Self-Entailment)



EDNA Best Correlates with Human Judgements for Faithfulness and Diversity

Metric	Faithfulness	Diversity
ROUGE-L	0.197	-0.164
BERTScore	0.209	-0.195
Entailment	0.588	-0.067
1 - Self-BLEU	-0.208	0.880
1 - Self-Entailment	-0.187	0.771
1 - Self-BERTScore	-0.198	0.873
EDNA	0.482	0.174

Different automatic metrics and their correlation against human assessments using Spearman's rank coefficient

Composition Sampling: Conclusion

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

**Sampled text
Composition**

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

**Generated output grounded to
its composition**



- Currently the best available decoding strategy for generating diverse and meaningful output.
 - avoids degeneration by instilling diversity in composition, rather than directly on the surface form
- Composition Sampling is straightforward to implement and does not require any external system to augment the input during inference

Summary

Evaluation and Benchmarking

Understand and measure types of hallucinations

- ❖ Hallucinations are still common
- ❖ **Common metrics** (Rouge, BLEU, BertScore) are **not correlated** with faithfulness
- ❖ Learned metric-based **candidate selection improves faithfulness**

Planning-based Controllable Models

Planning-based Controllable and Faithful Generation

- ❖ **Provides improved explainability and controllability**
- ❖ Control Hallucinations and generate faithful diverse texts
- ❖ **No change in architecture**

Thank You

Shashi Narayan

On Faithfulness and Factuality in Abstractive Summarization, Joshua Maynez, Shashi Narayan, Bernd Bohnet and Ryan McDonald, ACL 2020.

FROST: Planning with Learned Entity Prompts for Abstractive Summarization, Shashi Narayan, Yao Zhao, Joshua Maynez, Gonçalo Simões, Vitaly Nikolaev, Ryan McDonald, Transactions of the Association for Computational Linguistics (TACL), Volume 9.

A Well-Composed Text is Half Done! Composition Sampling for Diverse Conditional Generation, Shashi Narayan, Gonçalo Simões, Yao Zhao, Joshua Maynez, Dipanjan Das, Michael Collins, Mirella Lapata. ACL 2022.

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