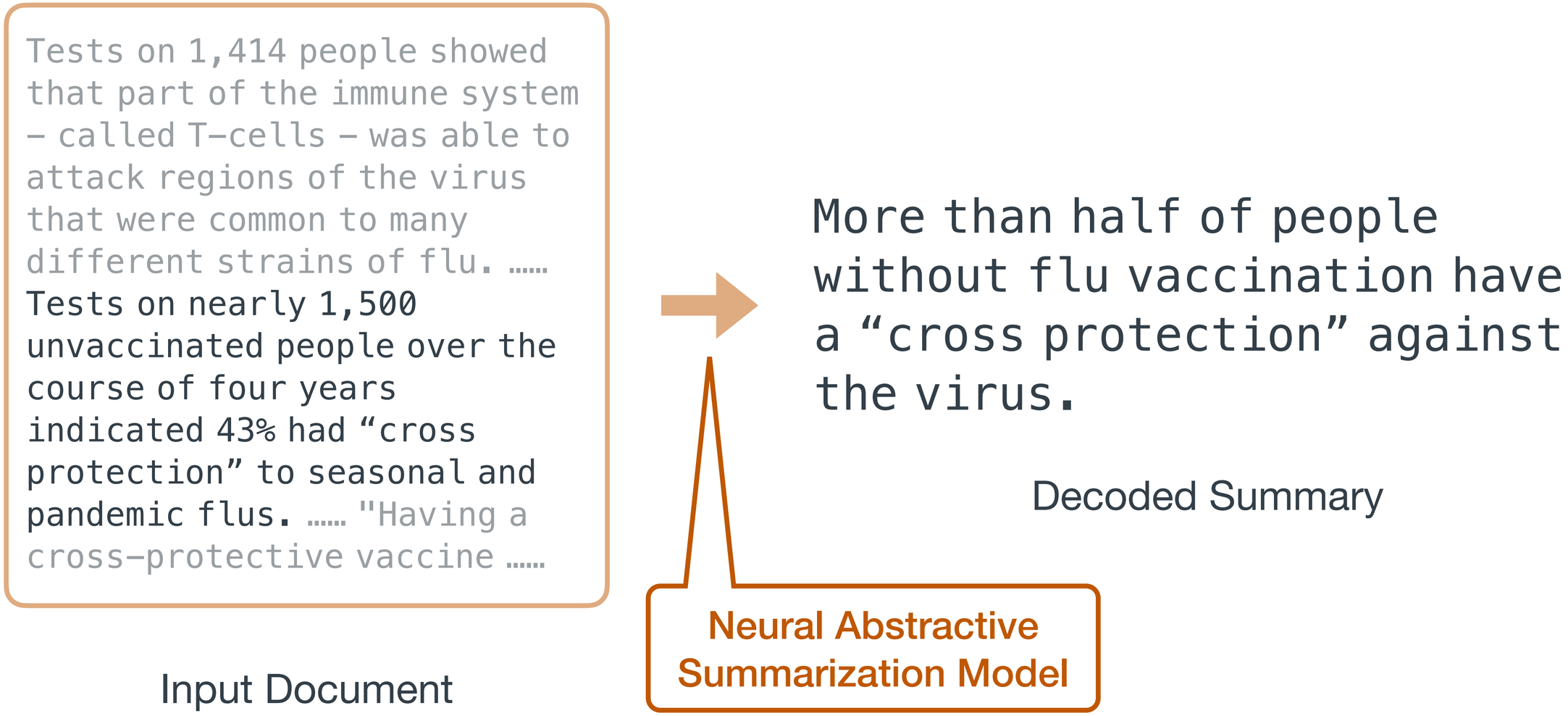
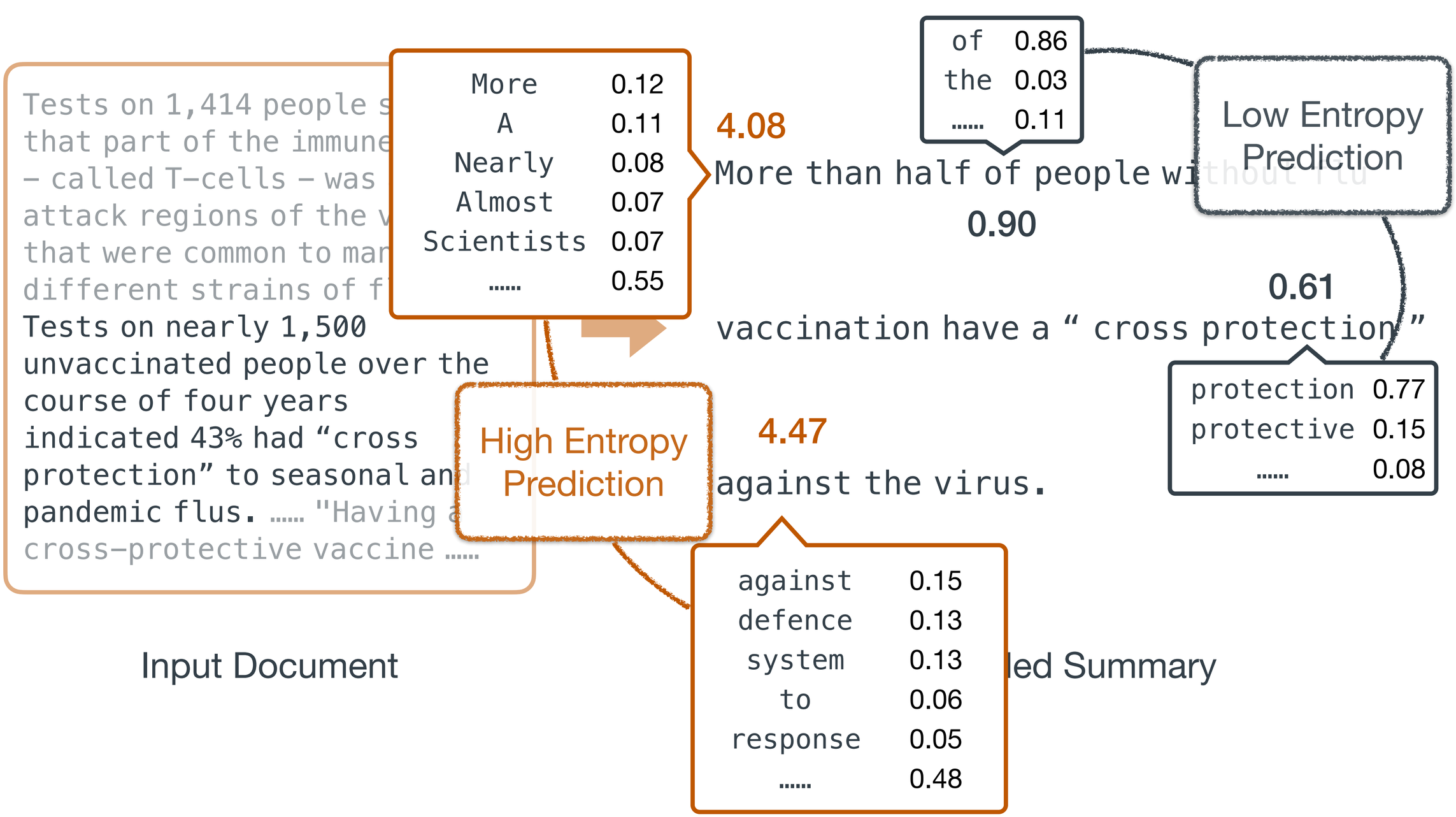
Understanding Neural Abstractive Summarization Models via Uncertainty

Jiacheng Xu Shrey Desai Greg Durrett
The University of Texas at Austin


Tests on 1,414 people showed that part of the immune system – called T-cells – was able to attack regions of the virus that were common to many different strains of flu. ……
Tests on nearly 1,500 unvaccinated people over the course of four years indicated 43% had “cross protection” to seasonal and pandemic flus. …… "Having a cross-protective vaccine ……
Input Document
Neural Abstractive Summarization Model
More than half of people without flu vaccination have a “cross protection” against the virus.
Decoded Summary


Tests on 1,414 people s
that part of the immune
– called T-cells – was
attack regions of the v
that were common to mar
different strains of f
Tests on nearly 1,500
unvaccinated people over the
course of four years
indicated 43% had “cross
protection” to seasonal and
pandemic flus. …… "Having a
cross-protective vaccine ……
More 0.12
A 0.11
Nearly 0.08
Almost 0.07
Scientists 0.07
…… 0.55
4.08
More than half of people wi
of 0.86
the 0.03
…… 0.11
Low Entropy Prediction
0.90
0.61
vaccination have a “ cross protection ”
protection 0.77
protective 0.15
…… 0.08
High Entropy Prediction
4.47
against the virus.
against 0.15
defence 0.13
system 0.13
to 0.06
response 0.05
…… 0.48
Input Document
led Summary

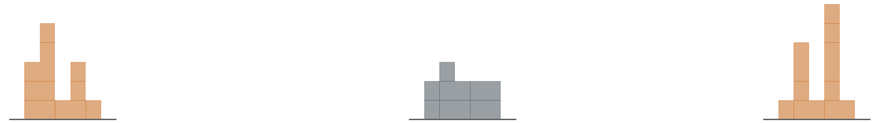
Experiment

- Datasets: CNN/DM (Hermann et al., 2015) and XSum (Narayan et al., 2018)
- PEGASUS (Zhang et al., 2020)
  - Configuration: `pagesus-large-xsum` & `pagesus-large-cnndm` from Huggingface
- BART (Lewis et al., 2020)
  - Results including comparison are not covered here; check our paper

Tests on 1,414 people showed that part of the immune system – called T-cells – was able to attack regions of the virus that were common to many different strains of flu. ……
Tests on nearly 1,500 unvaccinated people over the course of four years indicated 43% had “cross protection” to seasonal and pandemic flus. …… "Having a cross-protective vaccine ……

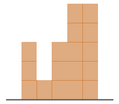
Input Document

More than half of people without flu

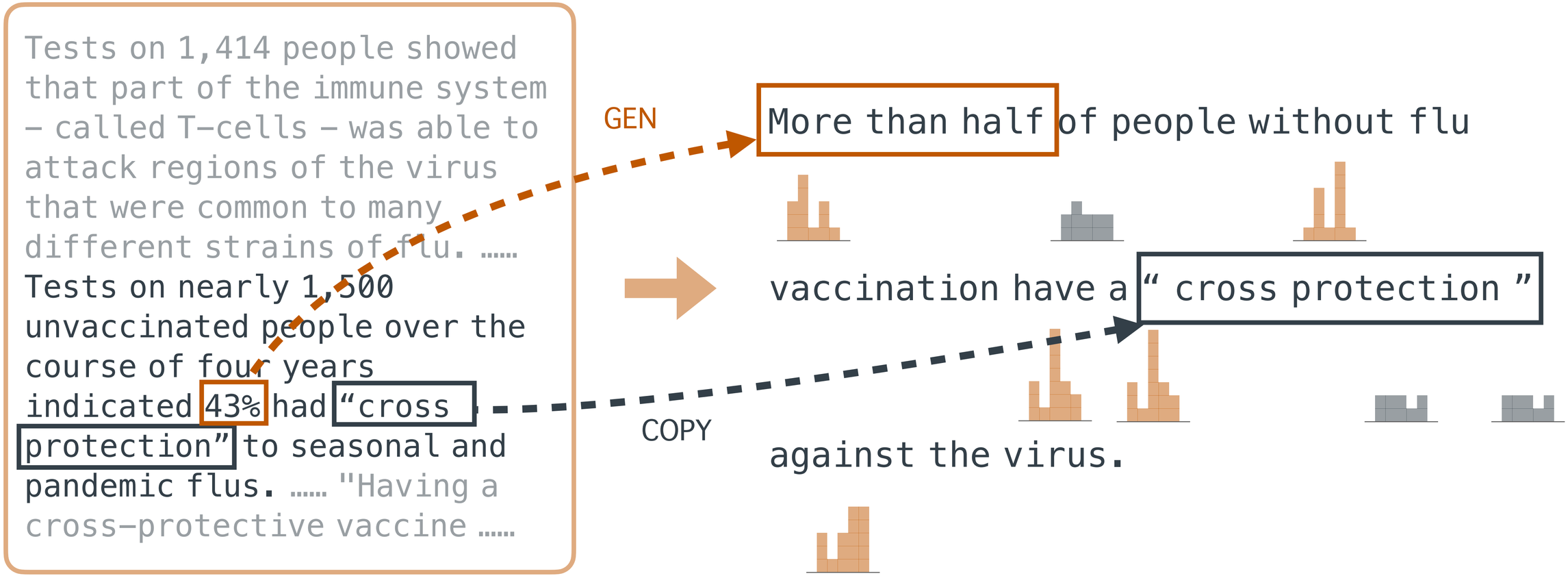


vaccination have a “ cross protection ”

against the virus.



Decoded Summary

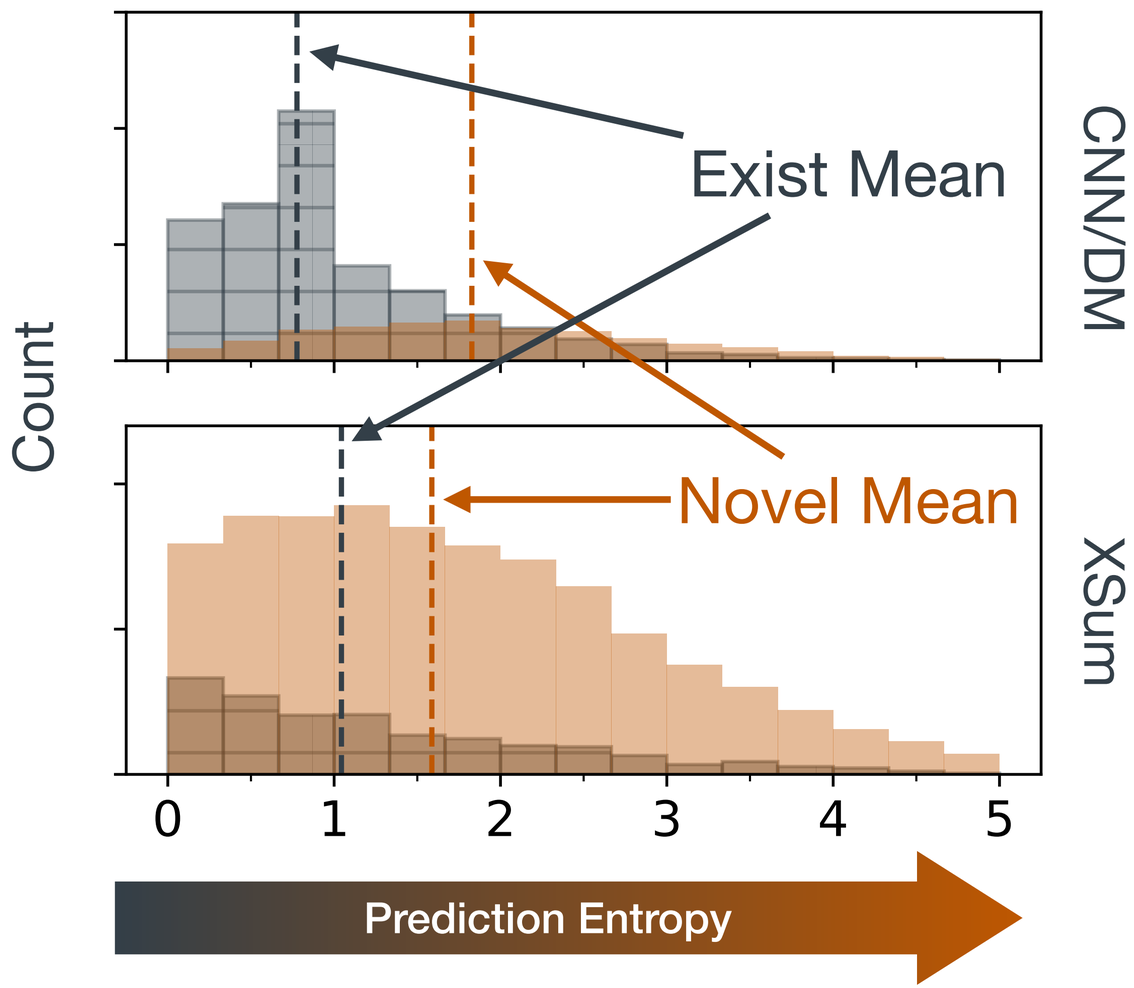


Can entropy tell us where ***copies*** / ***generation*** are happening?

Entropy of {Existing, Novel} Bigram

Summary	…… people without flu vaccination ……	
Extract Bigram	flu_vaccination	without_flu
Found in Input Document?	✓ addition to the flu vaccination for adults	✗
Novel or Existing	Existing	Novel
Indication of Behavior	COPY	GEN

Entropy of {Existing, Novel} Bigram



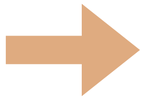
- Dotted lines: Mean of {Existing, Novel} Bigram entropy value
- Different datasets ➡ different behavior
  - Different prediction entropy distribution
  - More COPY (Existing Bigram) on CNN/DM
  - More GEN (Novel Bigram) on XSum
- Behavior Pattern

Entropy of {Existing, Novel} Bigram

Summary	…… people without flu vaccination ……	
Extract Bigram	flu_vaccination	without_flu
Found in Input Document?	✓ addition to the flu vaccination for adults	✗
Novel or Existing	Existing	Novel
Indication of Behavior	COPY	GEN
Avg Entropy	Low	High

Tests on 1,414 people showed that part of the immune system – called T-cells – was able to attack regions of the virus that were common to many different strains of flu. …… Tests on nearly 1,500 unvaccinated people over the course of four years indicated 43% had “cross protection” to seasonal and pandemic flus. …… "Having a cross-protective vaccine ……

Input Document



More than half of people without flu vaccination

Higher Entropy

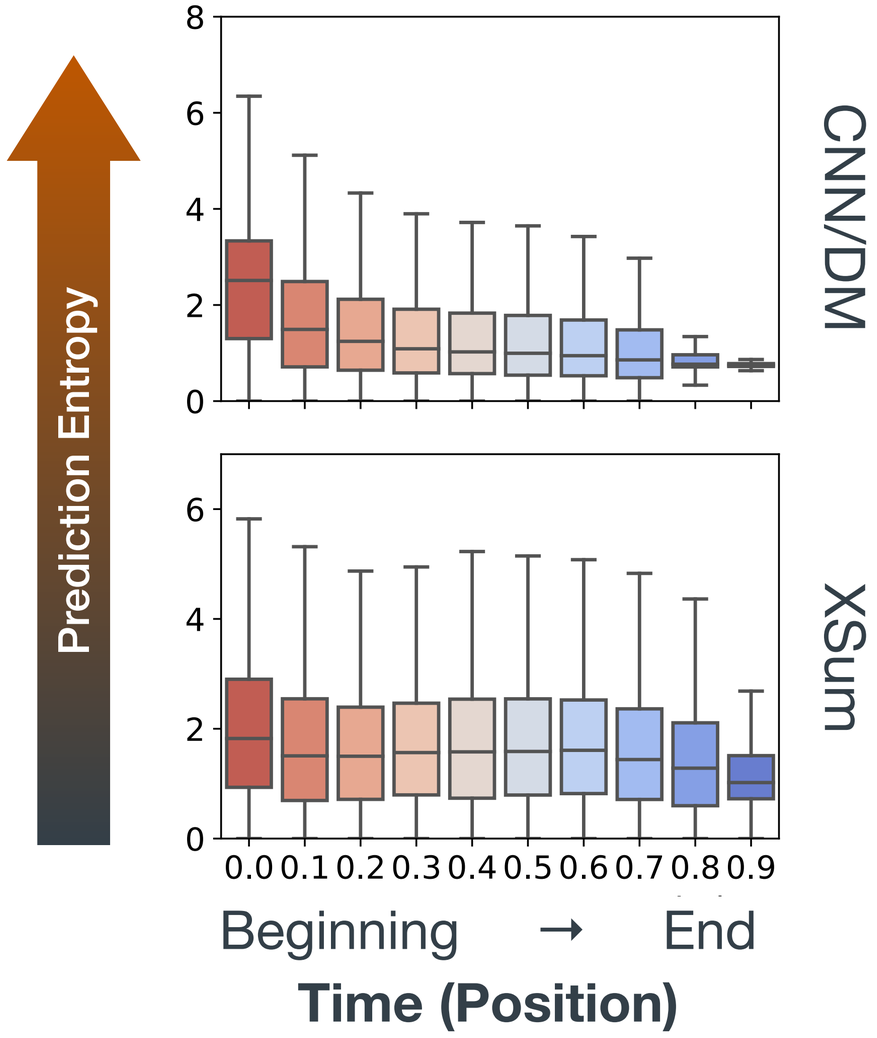
have a “ cross protection ” against the virus.

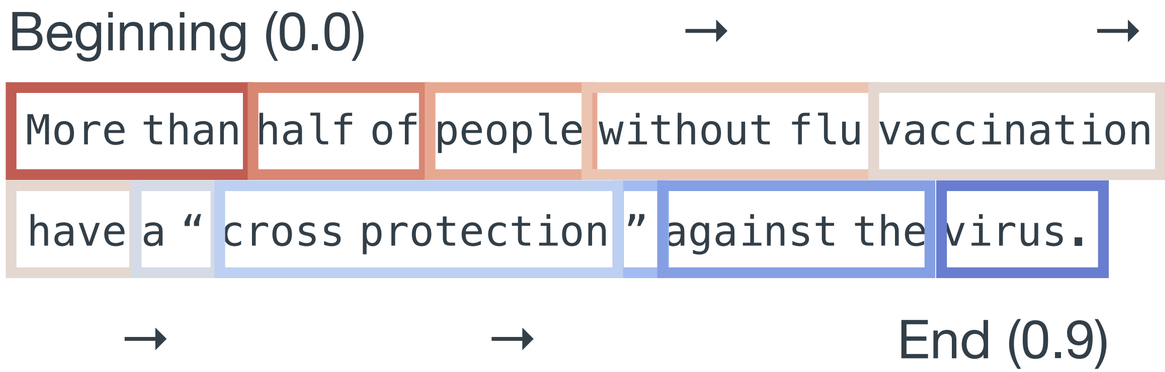
Lower Entropy

Decoded Summary

How does entropy vary over *time*?

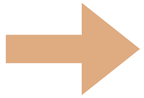
Entropy varies over time

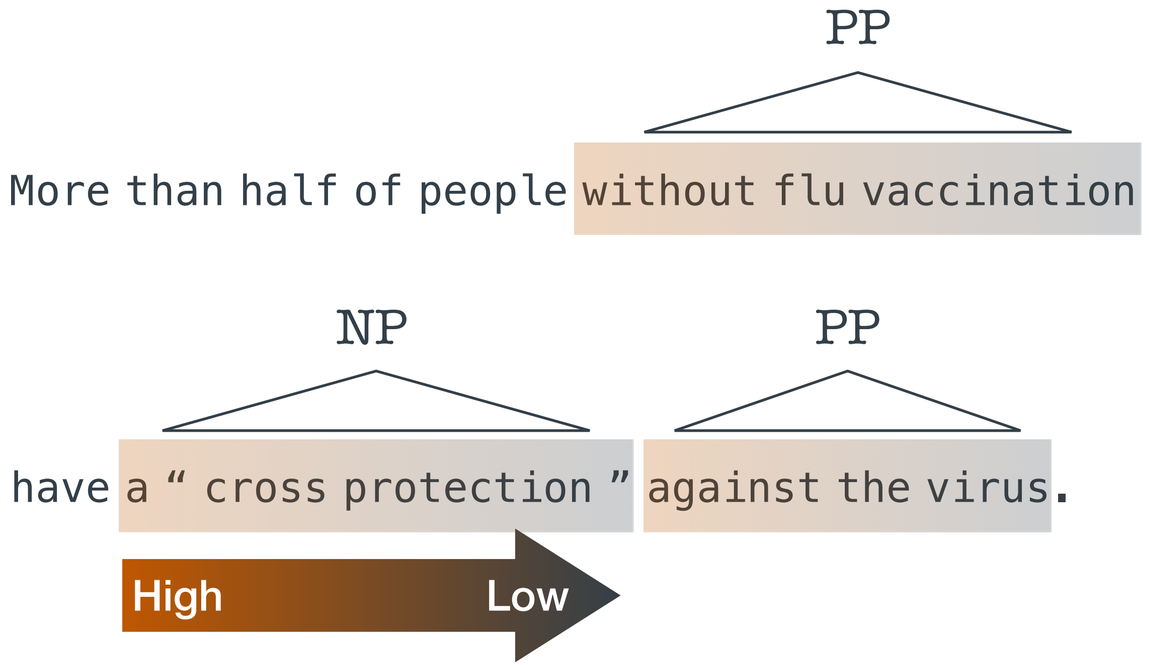




- Content planing happens at the beginning of the sentence ➡ relative higher entropy
- Entropy decreases during finishing a sentence, especially on CNN/DM

Tests on 1,414 people showed that part of the immune system – called T-cells – was able to attack regions of the virus that were common to many different strains of flu. ……
Tests on nearly 1,500 unvaccinated people over the course of four years indicated 43% had “cross protection” to seasonal and pandemic flus. …… "Having a cross-protective vaccine ……





Input Document

Decoded Summary

Does entropy connect to *syntactic* environment?

Entropy connects to Syntactic Distance

Summary: More than half of people without flu vaccination have a “ cross protection ” against the virus.

Constituency Parser

Linearized Tree: ((((More than half)) (of (people (without (flu vaccination))))) (have (a “ (cross protection) ” (against (the virus)))) .)

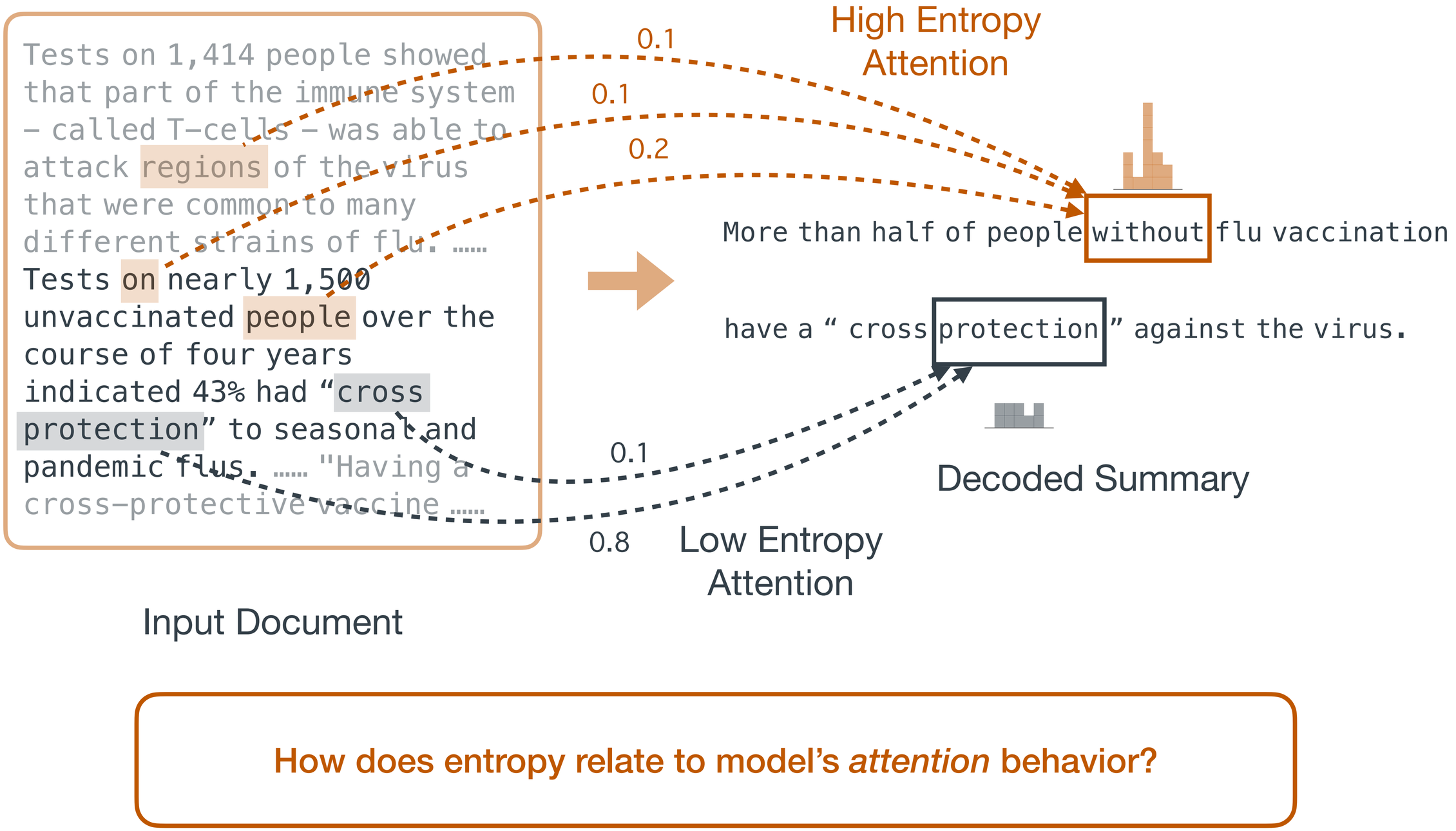
Syntactic Distance:
D(vaccination, have) = len(“)))))(”) = 6
D(cross, protection) = len(“”) = 0

Entropy:
H(have) = 2.3
H(protection) = 0.6

- Low and high entropy decisions can be localized to constituent span boundaries
  - Finishing the current span ➡ Lower entropy
  - Starting a new span ➡ Higher entropy

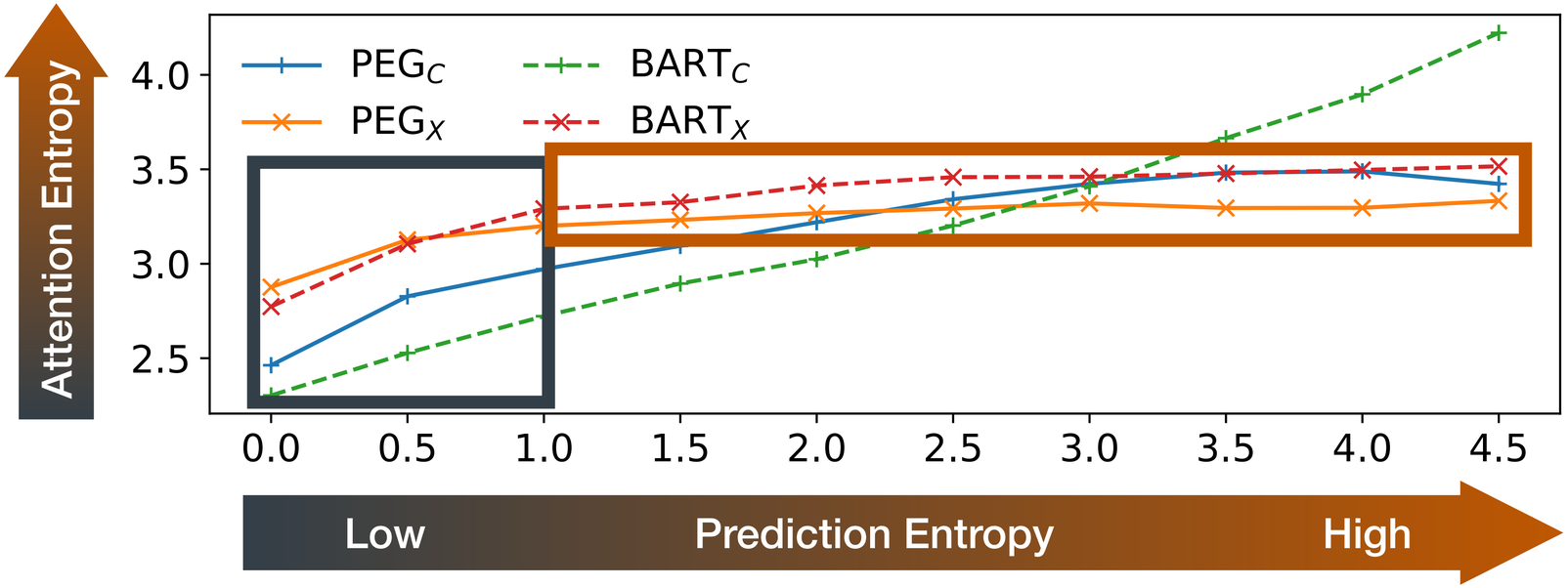
Entropies of Syntactic Productions

Check our paper!



How does entropy relate to model’s *attention* behavior?

$H_{\text{attn}} = f(H_{\text{pred}})$ when H_{pred} is low



$$H_{\text{attn}} = \begin{cases} f(H_{\text{pred}}), & \text{when } H_{\text{pred}} < 1 \\ \text{no correlation, otherwise} \end{cases}$$

Conclusion

- Uncertainty metrics like entropy is easy to access
  - Providing a lens into the inner workings of the model
- Uncertainty can help us understand
  - Model decision like COPY vs. GEN
  - Model behavior in different syntactic environments
  - Coarse properties of attention

Thank you!

Code: https://github.com/jiacheng-xu/text-sum-uncertainty
TAUR Lab: http://taur.cs.utexas.edu/
Supported by: TACC, NSF, Salesforce