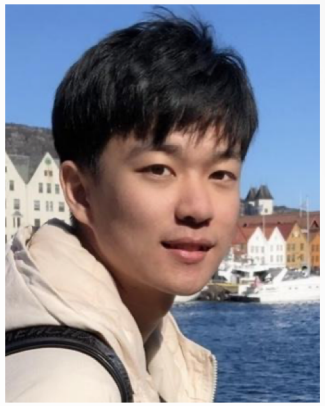
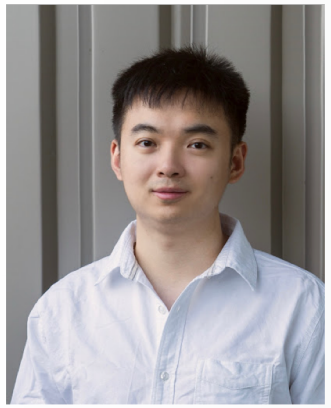
EMNLP 2021 Tutorial

Knowledge-Enriched Natural Language Generation






Wenhao Yu[1], Meng Jiang[1], Zhiting Hu[2], Qingyun Wang[3], Heng Ji[3,4], Nazneen Rajani[5]

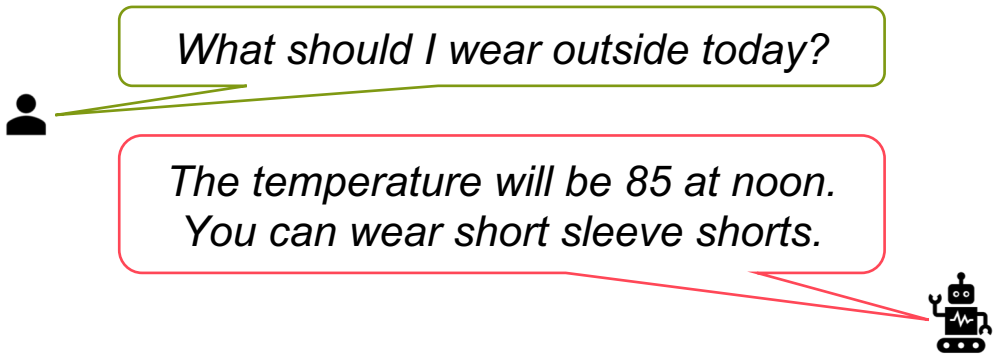
1 University of Notre Dame 2 University of California San Diego
3 University of Illinois at Urbana-Champaign 4 Amazon Scholar 5 Salesforce Research

Knowledge-Enriched NLG tutorial (Outline)

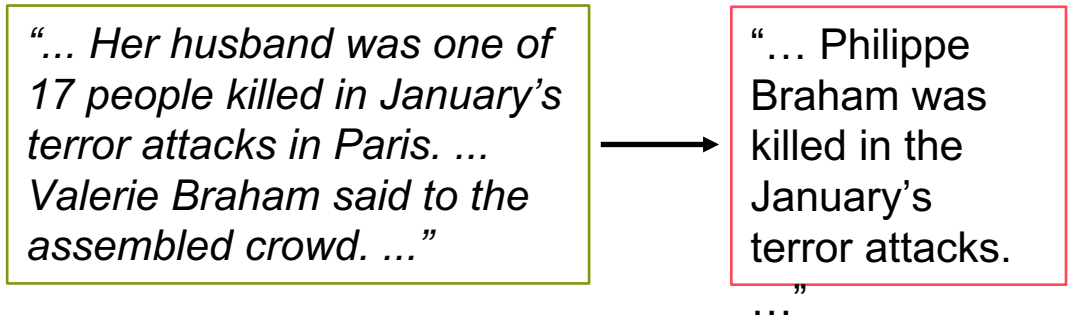
Time (AST / EST)	Content	Tutor
09:00 am – 09:10 am (08:00 am – 08:10 am)	Motivation and Key Questions	Meng Jiang
09:10 am – 09:50 am (08:10 am – 08:50 am)	General Methods of Integrating Knowledge into NLG	Zhiting Hu, Nazneen Rajani
09:50 am – 10:15 am (08:50 am – 09:15 am)	NLG Methods Enhanced by Various Knowledge Source	Wenhao Yu
10:15 am – 10:45 am (09:15 am – 09:45 am)	Coffee break	
10:45 am – 11:10 am (09:45 am – 10:10 am)	NLG Methods Enhanced by Various Knowledge Source	(cont.)
11:10 am – 12:10 pm (10:10 am – 11:10 am)	Applications, Benchmark Datasets, Code Library	Qingyun Wang
12:10 pm – 12:30 pm (11:10 am – 11:30 am)	Remaining Challenges and Future Directions	Heng Ji

Text generation is needed everywhere in our life!

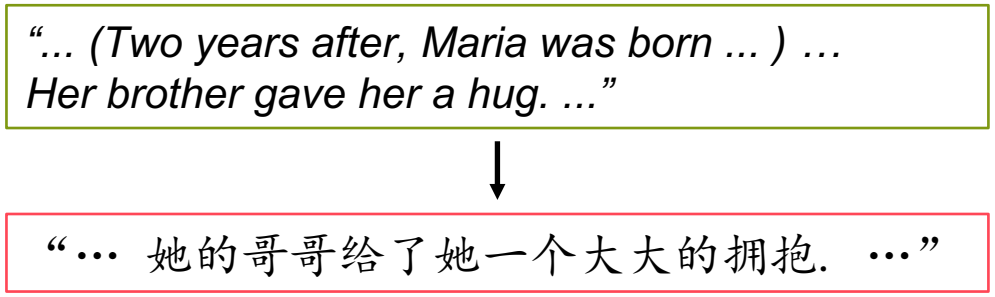
- Leveraging machine intelligence to make many things EASY and FAST



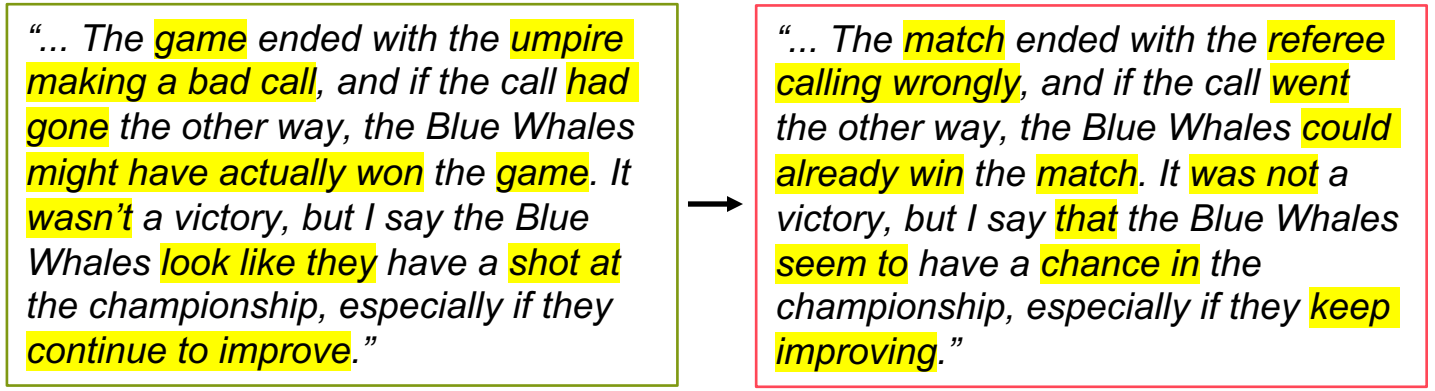
Dialog systems



Summarization



Machine Translation



Paraphrasing

Text generation is needed everywhere in our life!

Image Captioning:
Image-to-Text Generation

	85	Final	95	

ROCKETS LAKERS STATS NEWS

Player	Min	Reb	Ast	Pts
6 LeBron James · F	35	7	8	15
9 Kent Bazemore · F	27	2	1	9
3 Anthony Davis · C	33	13	2	16
0 Russell Westbrook · G	35	8	9	20
20 Avery Bradley · G	30	3	0	2
7 Carmelo Anthony	25	3	0	23
15 Austin Reaves	20	2	0	2
11 Malik Monk	19	2	0	0
10 DeAndre Jordan	17	3	2	8

HOU 1-5 **85** Final **95** LAL 4-3

LOS ANGELES -- — Carmelo Anthony scored 23 points in a reserve role and Russell Westbrook added 20 points in the Los Angeles Lakers' 95-85 victory over the Houston Rockets on Sunday night.

Anthony Davis had 16 points and 13 rebounds, and LeBron James scored 15 points in the Lakers' fourth win in five games after opening the season with two losses. Westbrook added eight rebounds and nine assists.

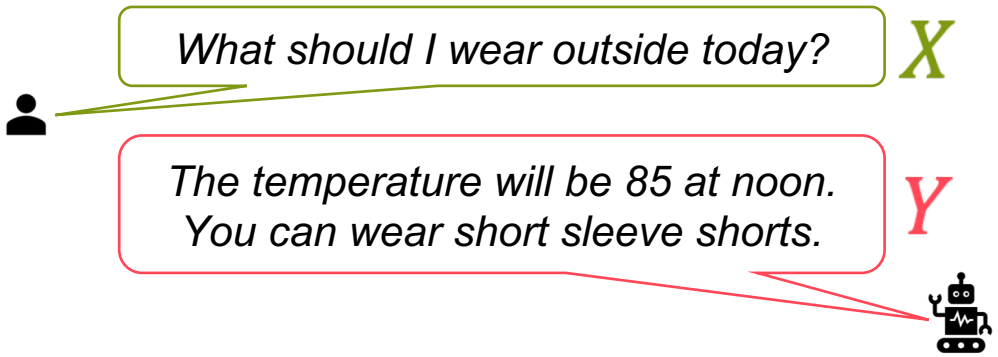
Eric Gordon scored 17 points and Christian Wood had 16 for the Rockets, who lost their fourth straight at the start of a five-game road trip.

Sports News Generation:
Table-to-Text Generation

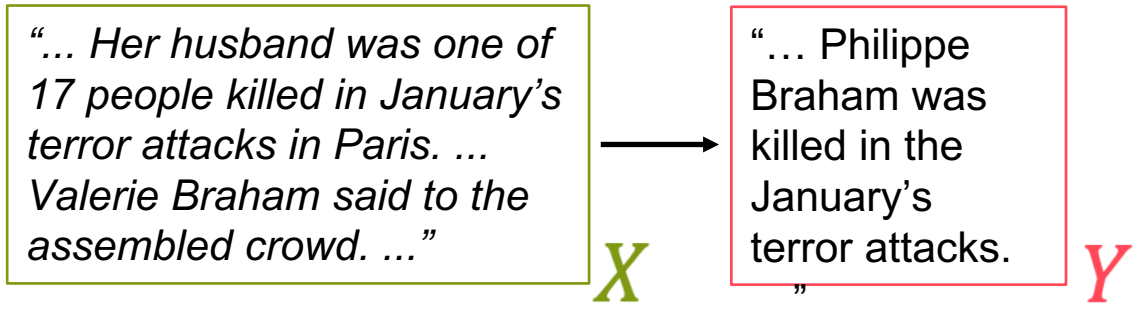
Not the main focus of this tutorial.

Text-to-text generation is needed everywhere!

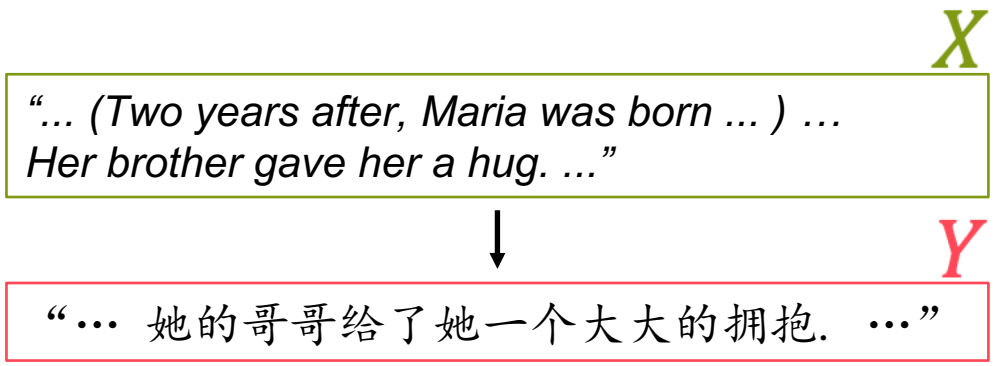
- $P(Y|X) = P(y_1, \cdots, y_m | x_1, \cdots, x_n) = \prod_{t=1}^{m} p(y_t | X, y_1, \cdots, y_{t-1})$, when Y is text and X is text.



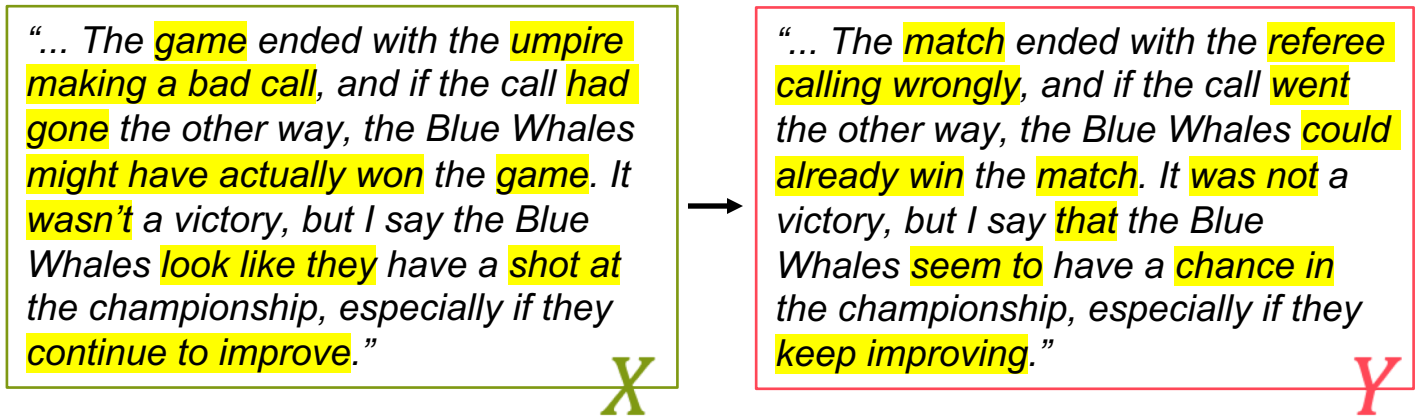
Dialog systems



Summarization



Machine Translation



Paraphrasing

Knowledge is needed everywhere in text generation!

- Mistakes may occur due to lack of knowledge.

What should I wear outside today?

The temperature will be ~~65~~ 85 at noon. You can wear short sleeve shorts.

Dialog systems

"... Her husband was one of 17 people killed in January's terror attacks in Paris. ... Valerie Braham said to the assembled crowd. ..."

→ "... ~~Valerie Braham~~ Philippe Braham was killed in the January's terror attacks. …"

Summarization

"... (Two years after, Maria was born ...) … Her brother gave her a hug. ..."

↓

"… 她的~~弟弟~~哥哥给了她一个大大的拥抱. …"

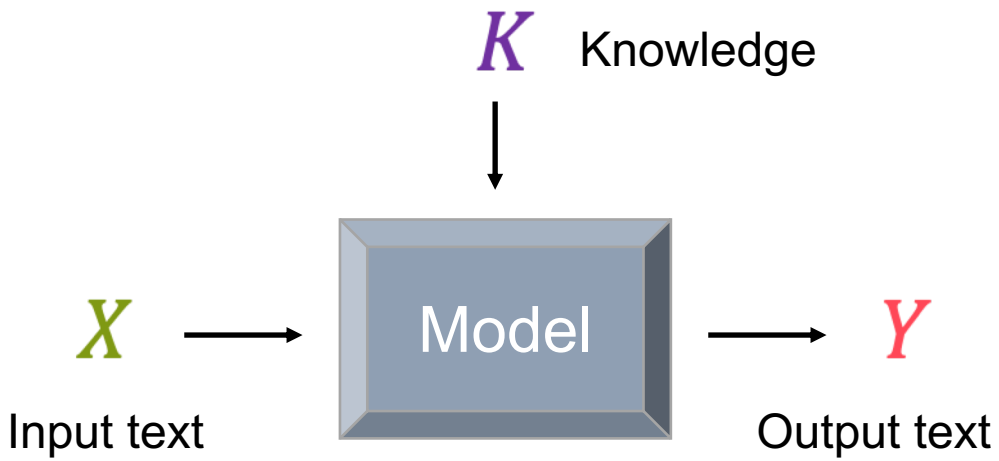
Machine Translation

"... The game ended with the umpire making a bad call, and if the call had gone the other way, the Blue Whales might have actually won the game. It wasn't a victory, but I say the Blue Whales look like they have a shot at the championship, especially if they continue to improve."

→ *"... The match ended with the referee calling wrongly, and if the call went the other way, the Blue Whales could ~~never~~ already win the match. It was not a victory, but I say that the Blue Whales seem to have a chance in the championship, especially if they keep improving."*

Paraphrasing

Knowledge-Enriched Natural Language Generation (KE-NLG)



Key questions: (Outline of this tutorial)

- Where does the Knowledge come from? In other words, what are the Knowledge sources?
- What are the data/representations of Knowledge?
- What are the methods for integrating Knowledge into NLG models? How to utilize different types of Knowledge?
- What are the concrete NLG models/techniques that have been enhanced by Knowledge? What are their advantages and disadvantages?
- Where can we find benchmark datasets, code library, and hands-on tutorial for doing research on this topic?
- What are the remaining challenges and future directions?

Material for this tutorial

- Website: https://kenlg-tutorial.github.io/
  - Schedule
  - Slides
  - Tutor information
- A pre-print survey on the topic of Knowledge-enhanced NLG:
  - https://arxiv.org/abs/2010.04389
- A reading list and a collection of open-source code:
  - https://github.com/wyu97/KENLG-Reading