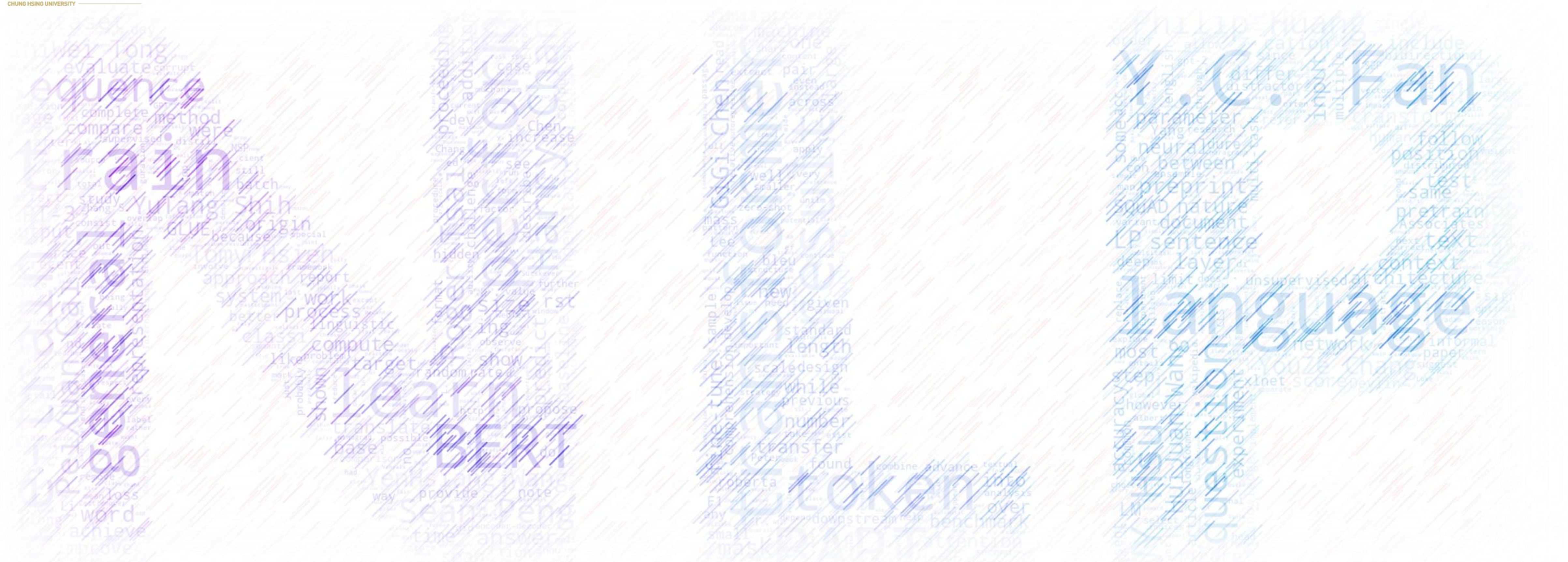
國立中興大學
NATIONAL CHUNG HSING UNIVERSITY




# Keyword Provision Question Generation for Facilitating Educational Reading Comprehension Preparation

Ying-Hong Chan, Ho-Lam Chung, Yao-Chung Fan
National Chung Hsing University, Taichung, Taiwan

INLG 2022

# Outline

- Introduction
- Educational QG based on Pre-trained LM
- Keyword Provision Question Generation
- Conclusion

# Introduction

# Motivation

- Question generators have great potential in the education market
- Current data-driven QG models are trained with factoid QA datasets (e.g., SQuAD or NewsQA), which are too simple for advanced reading practice assessment
- QG model suffers from the model controllability concern. If the question generated might not be a user expected result, which lowers user experiences in practical educational preparation scenarios

Context

At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars. After he was awarded the Ph.D. at Princeton University, he went back to California institute of technology.

Candidate Questions


1. Who once worked on the project to conquer planet mars?
2. Who was the youngest American to have ever won the gold medal in the International Physics Olympiad?

AI

# Our Goal

- Build an educational QG model using the pre-trained language model
- Propose KPQG (Keyword Provision Question Generation) model that allows users to provide keywords for guiding QG direction

**Input**

Context

At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars. After he was awarded the Ph.D. at Princeton University, he went back to California institute of technology.

Keywords

project, mars

**Output**

Question

Who once worked on the project to conquer planet mars?

# Educational QG based on Pre-trained LM

# Masked-LM Generation Architecture

- We follow QG model architecture (Chan, 2019) on Encoder Only Pre-trained LM model (e.g., BERT, RoBERTa, and DeBERTa)

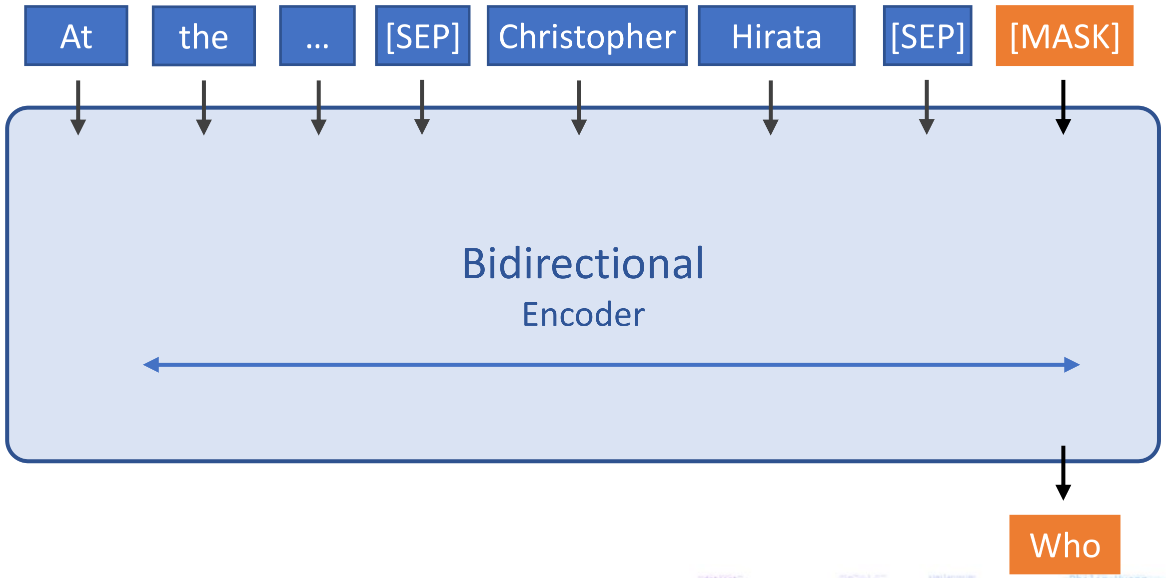
Input: Context and Answer

Context: At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars.

Answer: Christopher Hirata

Output: Question

Question: Who once worked on the project to conquer planet mars?



# Masked-LM Generation Architecture

- We follow QG model architecture (Chan, 2019) on Encoder Only Pre-trained LM model (e.g., BERT, RoBERTa, and DeBERTa)

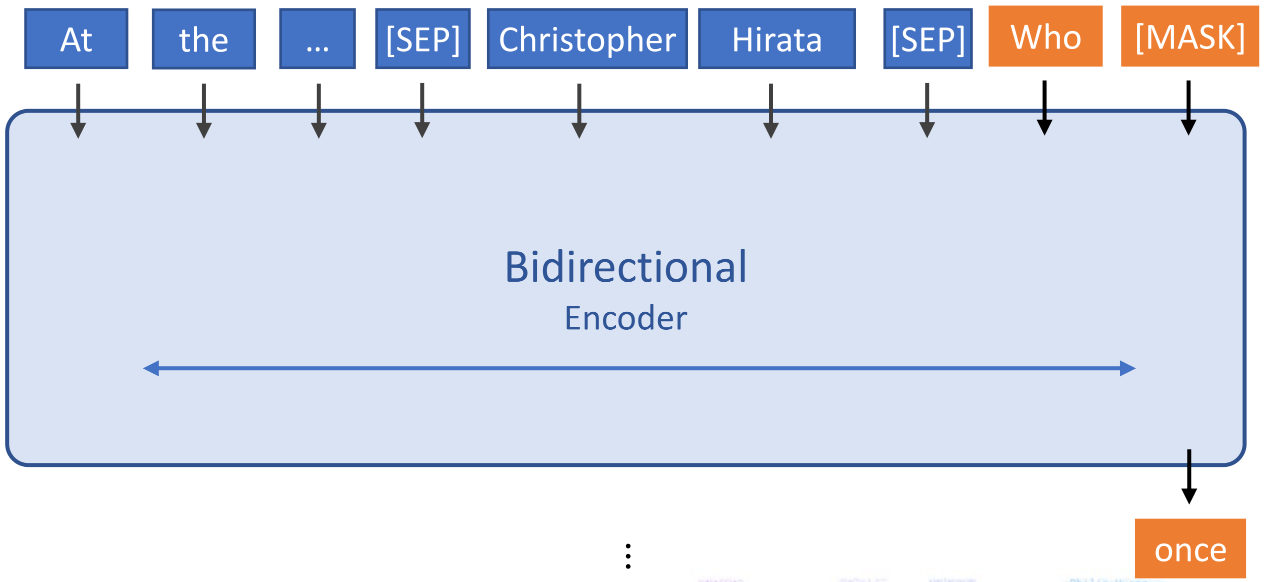
Input: Context and Answer

Context: At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars.

Answer: Christopher Hirata

Output: Question

Question: Who once worked on the project to conquer planet mars?



# Seq2Seq Generation Architecture

- We train QG tasks based on Seq2seq Pre-trained models (e.g. BART)

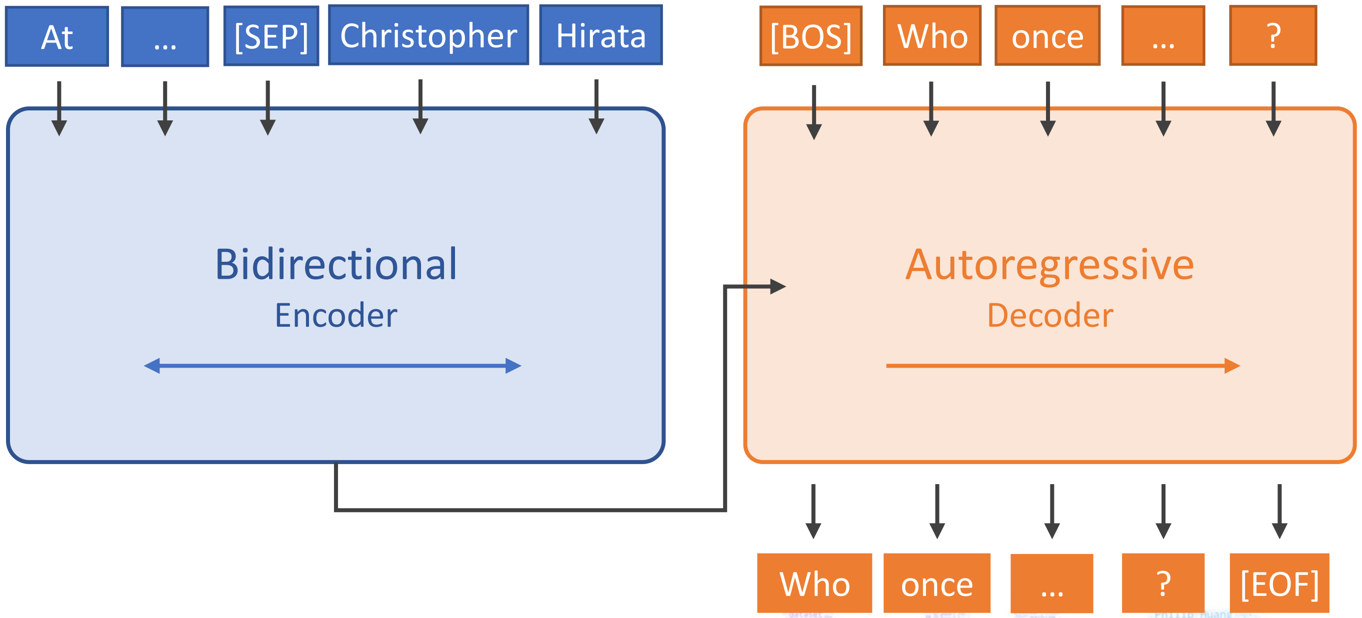
Input: Context and Answer

Context: At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars.

Answer: Christopher Hirata

Output: Question

Question: Who once worked on the project to conquer planet mars?



# Experiment Results

- Datasets: EQG-RACE (Jia, 2020) a exam-like QA datasets based on RACE (Lai, 2017)

| Train | Test |
| --- | --- |
| 17445 | 950 |

- We implement the Masked-LM QG architecture with BERT, RoBERTa, and DeBERTa, and the Seq2Seq QG architecture with BART, all models are using the base size

- DeBERTa-QG advances the SOTA result from 11.96 to 20.19 in terms of BLEU 4 score

| Model | BLEU 1 | BLEU 2 | BLEU 3 | BLEU 4 | ROUGE-L | MOTEOR |
| --- | --- | --- | --- | --- | --- | --- |
| (Jia, 2020) | 35.10 | 21.08 | 15.19 | 11.96 | 34.24 | 14.94 |
| BERT-QG | 43.37 | 29.53 | 22.25 | 17.54 | 44.26 | 20.47 |
| RoBERTa-QG | 46.37 | 32.15 | 24.34 | 19.21 | 46.96 | 22.32 |
| DeBERTa-QG | 47.16 | 32.81 | 25.18 | 20.19 | 47.33 | 22.55 |
| BART-QG | 46.78 | 32.30 | 24.53 | 19.39 | 47.00 | 22.22 |

# Keyword Provision Question Generation

# KPQG Architecture

- We add the keyword tokens to the input sequence and insert the [MASK] tokens in between

Input: Context, Answer, and Keywords

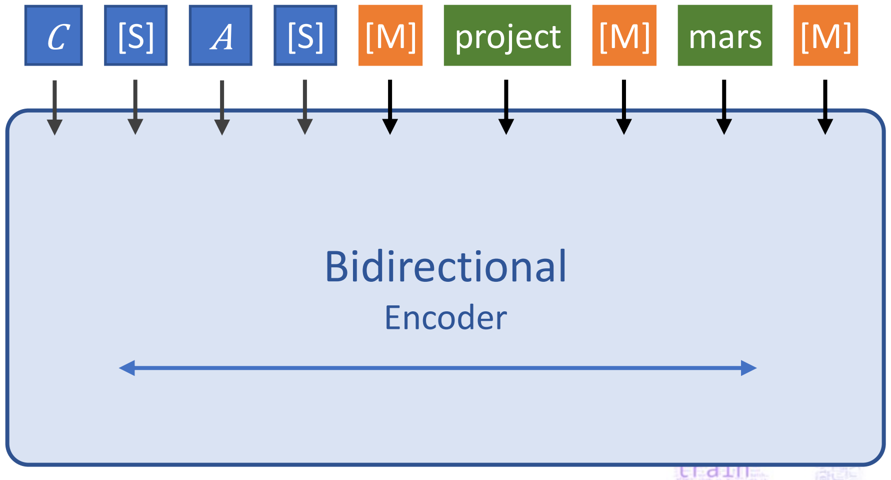
Context: At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars.

Answer: Christopher Hirata

Keywords: project, mars

Output: Question

Question: Who once worked on the project to conquer planet mars?



# KPQG Architecture

- We add the keyword tokens to the input sequence and insert the [MASK] tokens in between
- Leverage Masked-LM generation to predict suitably tokens for [MASK] tokens

Input: Context, Answer, and Keywords

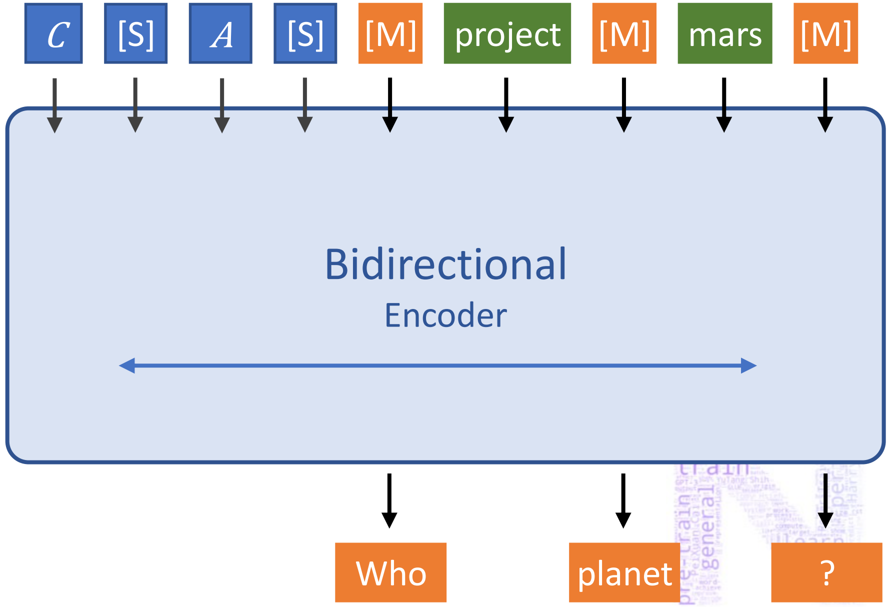
Context: At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars.

Answer: Christopher Hirata

Keywords: project, mars

Output: Question

Question: Who once worked on the project to conquer planet mars?



# KPQG Architecture

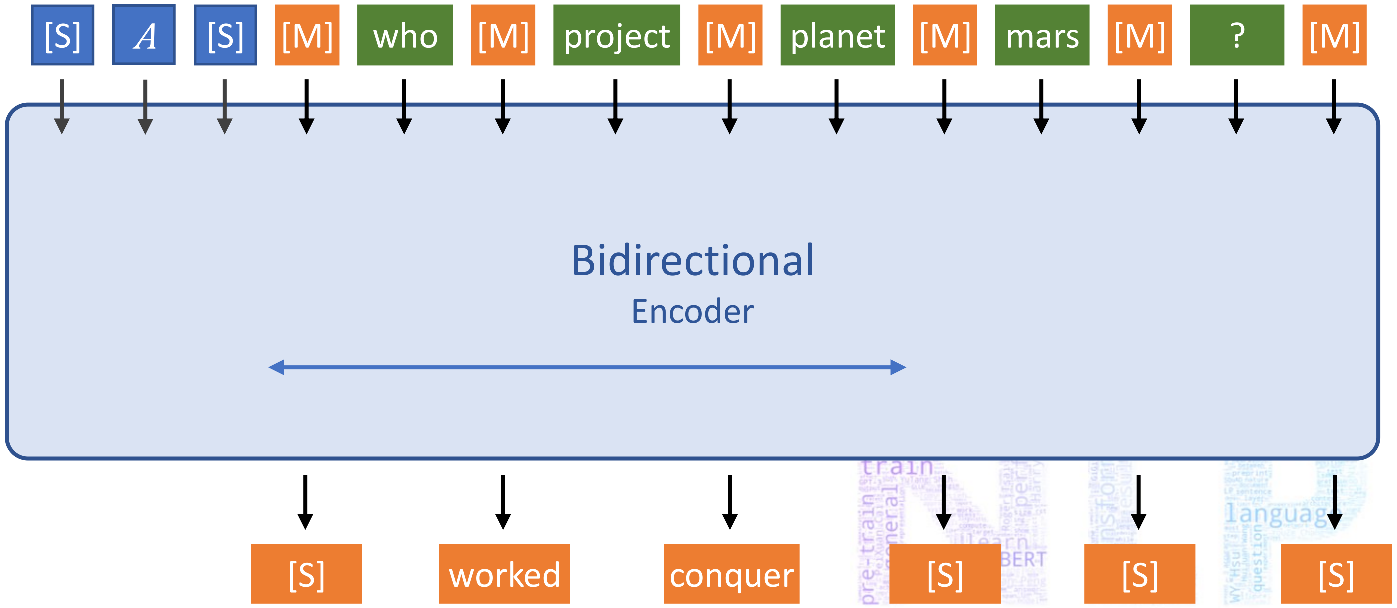
- We add the keyword tokens to the input sequence and insert the [MASK] tokens in between
- Leverage Masked-LM generation to predict suitably tokens for [MASK] tokens
- After the prediction, we recursively insert and predict the [MASK] tokens in the same manner. The iteration continues till all masked tokens becomes [SEP]

Input: Context, Answer, and Keywords

Context: At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars.

Answer: Christopher Hirata

Keywords: project, mars




Output: Question

Question: Who once worked on the project to conquer planet mars?

# Token Importance

- The KPQG is trained to predict a masked token before/after the input/generated keyword tokens, and we propose the idea that the more important tokens have priority to be predicted
- Our idea is that if masking some token $q_i$ from a question sentence $[q_1, \ldots, q_{|Q|}]$ leads to decreased QA model performance, then $q_i$ is considered to be an important one with respect to the QA pairs

Question: Who once worked on the project to conquer planet mars?

masking "Who"

[PAD] once worked on the project to conquer planet mars?

Context: At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars.

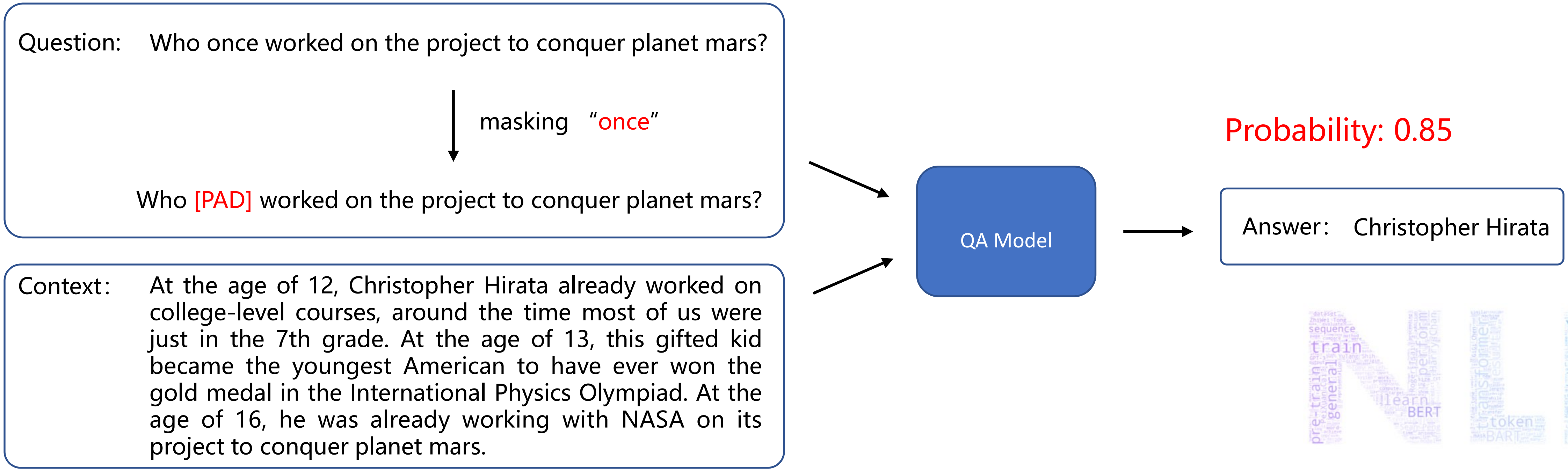
QA Model

Probability: 0.65

Answer: Christopher Hirata

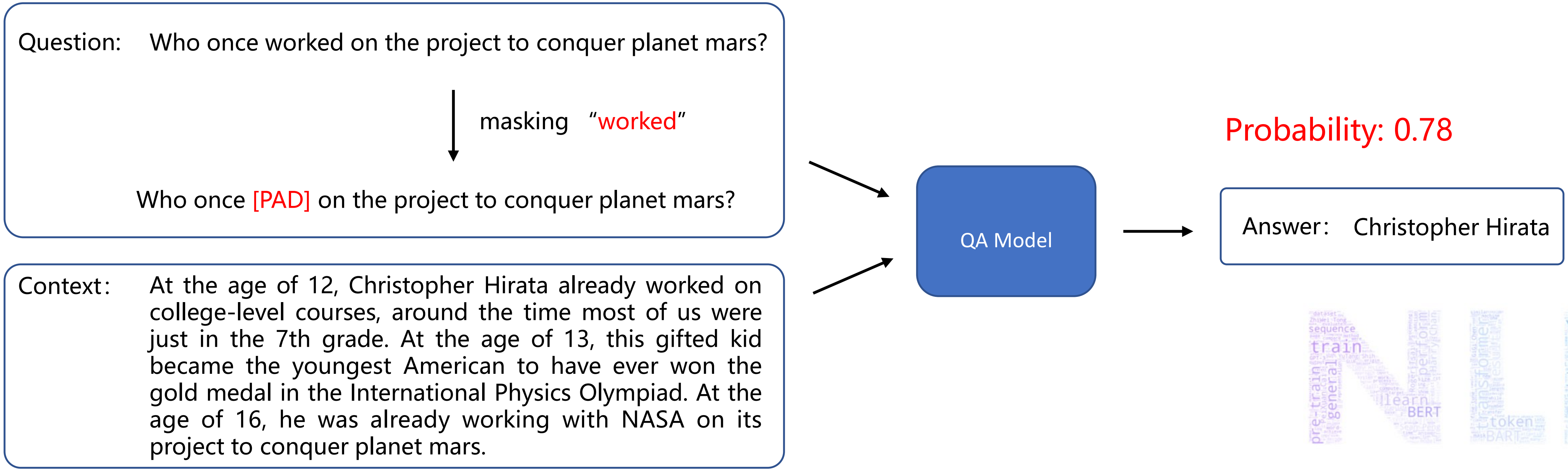
# Token Importance

- The KPQG is trained to predict a masked token before/after the input/generated keyword tokens, and we propose the idea that the more important tokens have priority to be predicted
- Our idea is that if masking some token $q_i$ from a question sentence $[q_1, ..., q_{|Q|}]$ leads to decreased QA model performance, then $q_i$ is considered to be an important one with respect to the QA pairs



# Token Importance

- The KPQG is trained to predict a masked token before/after the input/generated keyword tokens, and we propose the idea that the more important tokens have priority to be predicted
- Our idea is that if masking some token $q_i$ from a question sentence $[q_1, \ldots, q_{|Q|}]$ leads to decreased QA model performance, then $q_i$ is considered to be an important one with respect to the QA pairs



# Training Data For KPQG

- We create training data for KPQG based on the token importance

If the importance of a question sentence [q1, ..., q9] is [q4, q6, q2, q5, q3, q1, q9, q7, q8] (from high to low)

| | $X_i$ | **Labels for [M]** |
|---|---|---|
| i = 0 | $C$ [S] $A$ [M] | $q_4$ |

# Human Evaluation

- We use the DeBERTa-base model for KPQG and use the RACE QA model from huggingface,[1] which has 84.9% accuracy on RACE datasets
- 300 context paragraphs and the corresponding answers were randomly selected from the test set of EQG-RACE
- We invited 30 evaluators, each given 10 contextual paragraphs, and asked to use the KPQG model to provide keywords to generate questions to compare the difference between QG and KPQG
- Score [0,1,2] in Likert scale based on the Fluency, Expectedness, and Answerability

| **Model** | **Fluency** | **Expectedness** | **Answerability** |
|---|---|---|---|
| DeBERTa-QG | 1.60 | 0.86 | 1.20 |
| DeBERTa-KPQG | 1.60 | 1.37 | 1.44 |

[1] https://huggingface.co/LIAMF-USP/roberta-large-finetuned-race

# Case Study

| **Example 1** | |
|---|---|
| Context | At the age of 12, Christopher Hirata already worked on college-level courses, around the time most of us were just in the 7th grade. At the age of 13, this gifted kid became the youngest American to have ever won the gold medal in the International Physics Olympiad. At the age of 16, he was already working with NASA on its project to conquer planet mars. |
| Answer | Christopher Hirata |
| Gold Question | Who once worked on the project to conquer planet mars? |
| DeBERTa-QG | Who was the youngest American to have ever won the gold medal in the International Physics Olympiad? |

# Case Study

| **Example 2** | |
|---|---|
| Context | Brazil like the French, Brazilians usually eat a light breakfast. Lunch, the largest meal of the day, usually consists of meat, rice, potatoes, beans and vegetables. between 6:00 p.m. and 8:00 p.m., people enjoy a smaller meal with their families. Brazilians don' t mind eating a hurried or light meal and sometimes buy food from street carts. but they always finish eating before walking away. |
| Answer | Brazil |
| Gold Question | In which country do people consider lunch the largest meal? |
| DeBERTa-QG | Which country has a light breakfast? |

# Case Study

| Example 3 | |
|---|---|
| Context | Three Central Texas men were honored with the Texas department of public safety’ s director’ s award in a Tuesday morning ceremony for their heroism in saving the victims of a fiery two car accident. the accident occurred on March 25 when a vehicle lost control while traveling on a rain-soaked state highway 6 near Baylor camp road. it ran into an oncoming vehicle, leaving the occupants trapped inside as both vehicles burst into flames. Bonge was the first on the scene and heard children screaming. he broke through a back window and pulled Mallory Smith, 10, and her sister, Megan Smith, 9, from the wreckage. |
| Answer | Bonge |
| Gold Question | Who saved Megan Smith from the damaged car? |
| DeBERTa-QG | Who was the first on the scene and heard children screaming? |

# Conclusion

- We report a state-of-the-art exam-like QG model by advancing the current best model from 11.96 to 20.19 (in terms of BLEU 4 score) on EQG-RACE dataset
- We present KPQG model for guiding QG generation, so as to address the problem of controllability
- By providing keyword information, we can generate results that are closer to the user's expectation. We believe that our method is more practical to the realization of educational QG system applications

# Reference


- Bao, H.; Dong, L.; Wei, F.; Wang, W.; Yang, N.; Liu, X.; Wang, Y.; Piao, S.; Gao, J.; Zhou, M.; et al. 2020. Unilmv2: Pseudo-masked language models for unified language model pre-training. arXiv preprint arXiv:2002.12804.
- Chan, Y.-H.; and Fan, Y.-C. 2019. A recurrent BERT-based model for question generation. In Proceedings of the 2nd Workshop on Machine Reading for Question Answering, 154–162.
- Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Dong, L.; Yang, N.; Wang, W.; Wei, F.; Liu, X.; Wang, Y.; Gao, J.; Zhou, M.; and Hon, H.-W. 2019. Unified language model pre-training for natural language understanding and generation. In Advances in Neural Information Processing Systems, 13042–13054.
- Du, X.; Shao, J.; and Cardie, C. 2017. Learning to ask: Neural question generation for reading comprehension. arXiv preprint arXiv:1705.00106.
- He, P.; Liu, X.; Gao, J.; and Chen, W. 2020. Deberta: Decoding-enhanced bert with disentangled attention. arXiv preprint arXiv:2006.03654.
- Jia, X.; Zhou, W.; Sun, X.; and Wu, Y. 2020. EQG-RACE: Examination-Type Question Generation. arXiv preprint arXiv:2012.06106.
- Lai, G.; Xie, Q.; Liu, H.; Yang, Y.; and Hovy, E. 2017. Race: Large-scale reading comprehension dataset from examinations. arXiv preprint arXiv:1704.04683.

# Reference


- Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; and Zettlemoyer, L. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.
- Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
- Rajpurkar, P.; Zhang, J.; Lopyrev, K.; and Liang, P. 2016. Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.
- Trischler, A.; Wang, T.; Yuan, X.; Harris, J.; Sordoni, A.; Bachman, P.; and Suleman, K. 2016. Newsqa: A machine comprehension dataset. arXiv preprint arXiv:1611.09830.
- Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; Davison, J.; Shleifer, S.; von Platen, P.; Ma, C.; Jernite, Y.; Plu, J.; Xu, C.; Scao, T. L.; Gugger, S.; Drame, M.; Lhoest, Q.; and Rush, A. M. 2020. HuggingFace's Transformers: Stateof- the-art Natural Language Processing. arXiv:1910.03771.
- Zhao, Y.; Ni, X.; Ding, Y.; and Ke, Q. 2018. Paragraph-level Neural Question Generation with Maxout Pointer and Gated Self-attention Networks. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 3901–3910.
- Zhou, Q.; Yang, N.; Wei, F.; Tan, C.; Bao, H.; and Zhou, M. 2017. Neural question generation from text: A preliminary study. In National CCF Conference on Natural Language Processing and Chinese Computing, 662–671. Springer.


Yao-Chung Fan
Department of Computer Science and Engineering
National Chung Hsing University, Taiwan
Contact: yfan@nchu.edu.tw

Code: https://github.com/Harry-Chan/keyword-provision-qg