

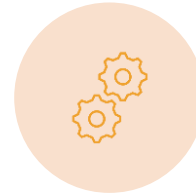
# **TEXT INFILLING**

## **MISSING TEXT GENERATION**

# We will cover...



Introduction



Background/Applications



Our Approach



Results and observations



Discussion

# INTRODUCTION

- Text infilling, Missing text generation[1]
- Predict spans of missing text
- Based on preceding and subsequent text
- Consistency[2]
- Cloze task[3]

[1] Yichen Huang, Yizhe Zhang, Oussama Elachqar, and Yu Cheng. 2019. Inset: Sentence infilling with inter-sentential transformer. arXiv preprint arXiv:1911.03892.

[2] Chris Donahue, Mina Lee, and Percy Liang. 2020. Enabling language models to fill in the blanks. arXiv preprint arXiv:2005.05339.

[3] Wilson L Taylor. 1953. “cloze procedure”: A new tool for measuring readability. Journalism quarterly, 30(4):415–433.

# BACKGROUND AND APPLICATIONS

- Mixed initiative[4] story generation[5]
  - Computer and human take the “initiative” to input
- Text revision[6]
  - Post edit text to refine
- Restoring ancient documents that are missing content[7]
- Mixed initiative [4] machine translation[8]

[4] J. E. Allen, C. I. Guinn, and E. Horvitz, “Mixed-initiative interaction,” IEEE Intelligent Systems and their Applications, vol. 14, no. 5, pp. 14–23, 1999.

[5] Daphne Ippolito, David Grangier, Chris Callison-Burch, and Douglas Eck. 2019. Unsupervised hierarchical story infilling. In Proceedings of the First Workshop on Narrative Understanding, pages 37–43.

[6] Yong-Siang Shih, Wei-Cheng Chang, and Yiming Yang. 2019. XI-editor: Post-editing sentences with xlnet. arXiv preprint arXiv:1910.10479.

[7] Yannis Assael, Thea Sommerschild, and Jonathan Prag. 2019. Restoring ancient text using deep learning: a case study on greek epigraphy. arXiv preprint arXiv:1910.06262.

[8] Yanling Xiao, Lemao Liu, Guoping Huang, Qu Cui, Shujian Huang, Shuming Shi, and Jiajun Chen. 2022. BiTIIMT: A Bilingual Text-infilling Method for Interactive Machine Translation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1958–1969, Dublin,

# OUR APPROACH

## SETUP

- GPT Neo[9]
  - 1.3 Billion
- FRW dataset\*
  - Dungeons and Dragons[10] (d&d)
  - Forgotten Realms[11][12]
  - Fandom Wikia
  - Multiple datasets – various degrees of pre-processing
- Finetune free text generator
  - Forgotten Realms domain

\* <https://huggingface.co/datasets/Akila/ForgottenRealmsWikiDataset>

[9] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. 2020. The pile: An 800gb dataset of diverse text for language modeling. arXiv preprint arXiv:2101.00027.

[10] Gary Gygax and Dave Arneson. 1974. dungeons & dragons, volume 19. Tactical Studies Rules Lake Geneva, WI

[11] Wizards RPG Team et al. 2014. Player's handbook (dungeons & dragons). Wizards of the Coast.

[12] Ed Greenwood. 2013. Dungeons & Dragons: Forgotten Realms Vol. 1. IDW Publishing

# OUR APPROACH

## PIPELINE



# OUR APPROACH

## MASKED DATASET

- We use the following tags
  - `<|sentence_block|>` the sentence mask
  - `<|answers|>` denotes the start of the answer section
  - `<|end|>` denotes the end of a section
- Probability of getting masked = 80%

*Following a year and a half of war against Tethyr, Calishite forces destroy Ithmong and nearly wipe out Clan Ithal. `</sentence_block/>` With rule over both nations, this is the beginning of the Shoon Imperium. `</answers/>` Following a year and a half of war against Tethyr, Calishite forces destroy Ithmong and nearly wipe out Clan Ithal. Tethyr is defeated. In Greengrass, King Amahl II of Tethyr swears allegiance to Syl-Pasha Akkabar el Shoon of Calimshan. With rule over both nations, this is the beginning of the Shoon Imperium. `</end/>`*

# RESULTS AND OBSERVATIONS

## Our prompt:

A Bard, A Wizard, and a Paladin walk into a tavern. `</sentence_block/>` And they agree to take on the quest for 1000GP.

## Result:

A Bard, A Wizard, and a Paladin walk into a tavern. A Bard asks the Paladin for advice and then tells them about a specific story about an arcane mage. While the Bard refers to the story as a legend, the Paladin tells them that there are many similar stories about an arcane mage and wonders how the Bard came to know about these stories. The Bard tells them that their story is too easy. He says that an arcane mage is always poor, always entering into ugly fights. Therefore, an arcane mage should instead find someone else to fight, such as a fighter, a rebirth warrior, a berserker, a warrior-for-hire, or a mercenary. They then discuss how they need to act in their lives on the other side of an arcane mage. The Paladin tells them that the solution to this problem is to go outside. He gives the Bard a bag of gold and asks for advice. And then the Bard tells the Paladin that an arcane mage never trains others to be good at wielding weapons. Therefore, they should steal other people's weapons and arm them with their knowledge. The Bard, A Wizard, and a Paladin then go to a tavern. They steal an arcane mage's book from the library and give it to the Paladin. The Bard wants to know if this helps. The Bard, A Wizard, and a Paladin walk into a tavern. They steal a black dragon's shed from an olive grove. They learn from the Paladin about an arcane mage's past. And they agree to take on the quest for 1000GP.



# RESULTS AND OBSERVATIONS

## Our prompt:

A Bard, A Wizard, and a Paladin walk into a tavern. *</sentence\_block/>* And they agree to take on the quest for 1000GP.

## Result:

A Bard, A Wizard, and a Paladin walk into a tavern. A Bard asks the Paladin for advice and then tells them about a specific story about an arcane mage. While the Bard refers to the story as a legend, the Paladin tells them that there are many similar stories about an arcane mage and wonders how the Bard came to know about these stories. The Bard tells them that their story is too easy. He says that an arcane mage is always poor, always entering into ugly fights. Therefore, an arcane mage should instead find someone else to fight, such as a fighter, a rebirth warrior, a berserker, a warrior-for-hire, or a mercenary. They then discuss how they need to act in their lives on the other side of an arcane mage. The Paladin tells them that the solution to this problem is to go outside. He gives the Bard a bag of gold and asks for advice. And then the Bard tells the Paladin that an arcane mage never trains others to be good at wielding weapons. Therefore, they should steal other people's weapons and arm them with their knowledge. The Bard, A Wizard, and a Paladin then go to a tavern. They steal an arcane mage's book from the library and give it to the Paladin. The Bard wants to know if this helps. The Bard, A Wizard, and a Paladin walk into a tavern. They steal a black dragon's shed from an olive grove. They learn from the Paladin about an arcane mage's past. And they agree to take on the quest for 1000GP.

# DISCUSSION

- Free text generator vs. infilling generator
- Uni-directional context awareness
- 1 mask covering multiple sentences
- 1 mask covering 1 sentence

# REFERENCES

- [1] Yichen Huang, Yizhe Zhang, Oussama Elachqar, and Yu Cheng. 2019. Inset: Sentence infilling with inter-sentential transformer. arXiv preprint arXiv:1911.03892.
- [2] Chris Donahue, Mina Lee, and Percy Liang. 2020. Enabling language models to fill in the blanks. arXiv preprint arXiv:2005.05339.
- [3] Wilson L Taylor. 1953. “cloze procedure”: A new tool for measuring readability. *Journalism quarterly*, 30(4):415–433.
- [4] J. E. Allen, C. I. Guinn, and E. Horvitz, “Mixed-initiative interaction,” *IEEE Intelligent Systems and their Applications*, vol. 14, no. 5, pp. 14–23, 1999.
- [5] Daphne Ippolito, David Grangier, Chris Callison-Burch, and Douglas Eck. 2019. Unsupervised hierarchical story infilling. In *Proceedings of the First Workshop on Narrative Understanding*, pages 37–43.
- [6] Yong-Siang Shih, Wei-Cheng Chang, and Yiming Yang. 2019. Xl-editor: Post-editing sentences with xlnet. arXiv preprint arXiv:1910.10479.
- [7] Yannis Assael, Thea Sommerschield, and Jonathan Prag. 2019. Restoring ancient text using deep learning: a case study on greek epigraphy. arXiv preprint arXiv:1910.06262.
- [8] Yanling Xiao, Lemao Liu, Guoping Huang, Qu Cui, Shujian Huang, Shuming Shi, and Jiajun Chen. 2022. BiTIIMT: A Bilingual Text-infilling Method for Interactive Machine Translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1958–1969, Dublin, Ireland. Association for Computational Linguistics.
- [9] Leo Gao, Stella Biderman, Sid Black, Laurence Gold- ing, Travis Hoppe, Charles Foster, Jason Phang, Ho- race He, Anish Thite, Noa Nabeshima, et al. 2020. The pile: An 800gb dataset of diverse text for lan- guage modeling. arXiv preprint arXiv:2101.00027.
- [10] Gary Gygax and Dave Arneson. 1974. *dungeons & dragons*, volume 19. Tactical Studies Rules Lake Geneva, WI
- [11] Wizards RPG Team et al. 2014. *Player’s handbook (dungeons & dragons)*. Wizards of the Coast.
- [12] Ed Greenwood. 2013. *Dungeons & Dragons: Forgotten Realms Vol. 1*. IDW Publishing

THANK YOU