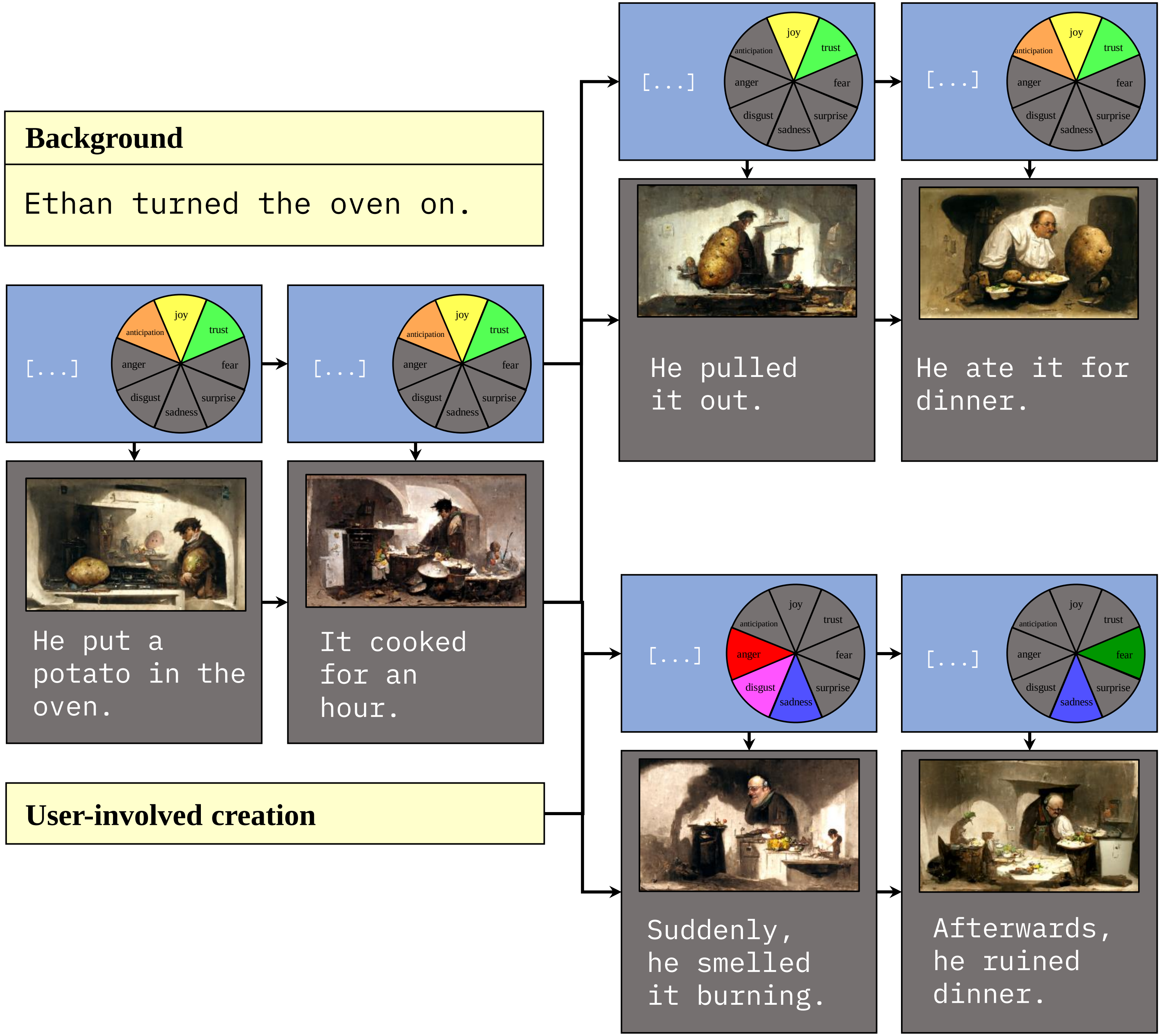
Rensselaer
why not change the world?®

Leveraging Large Language Models for Creative Story Generation

Yuetian Chen &
Brendan Capuzzo

CSCI 4968-01
Machine Learning and Optimization, Spring 2023

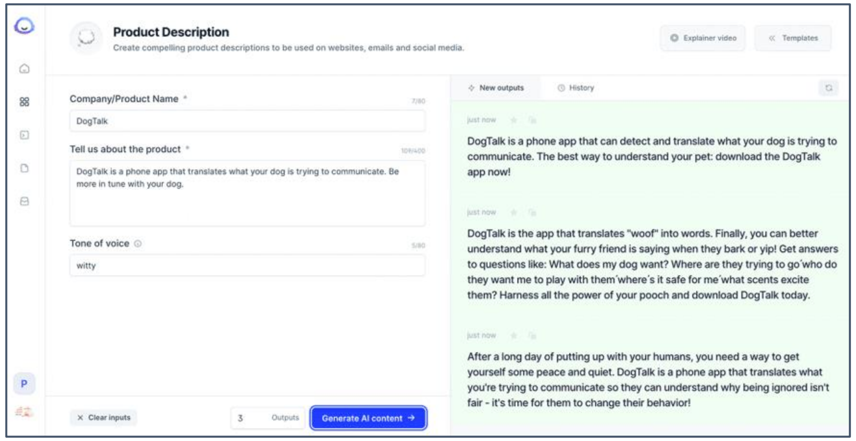


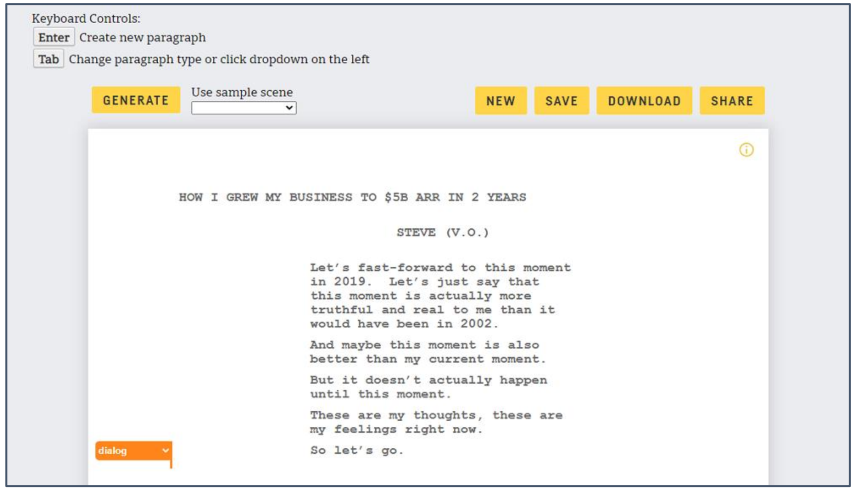
Motivation

Interactive authoring system for story generation

- **Commercial AI text generators**
  - → Jasper: https://www.jasper.ai/
  - → Rytr: https://rytr.me/
  - …
- **No direct control over the generated content**
- **No image generation**
  - → Even if invoking a separate image generation process
  - → Can't incorporate content from the generated images into future storytelling

- **Interactive authoring system**
  - → Control story with keywords and emotions
  - → Image generation
  - → Efficient knowledge distillation





Related Work

Interactive authoring system for story generation

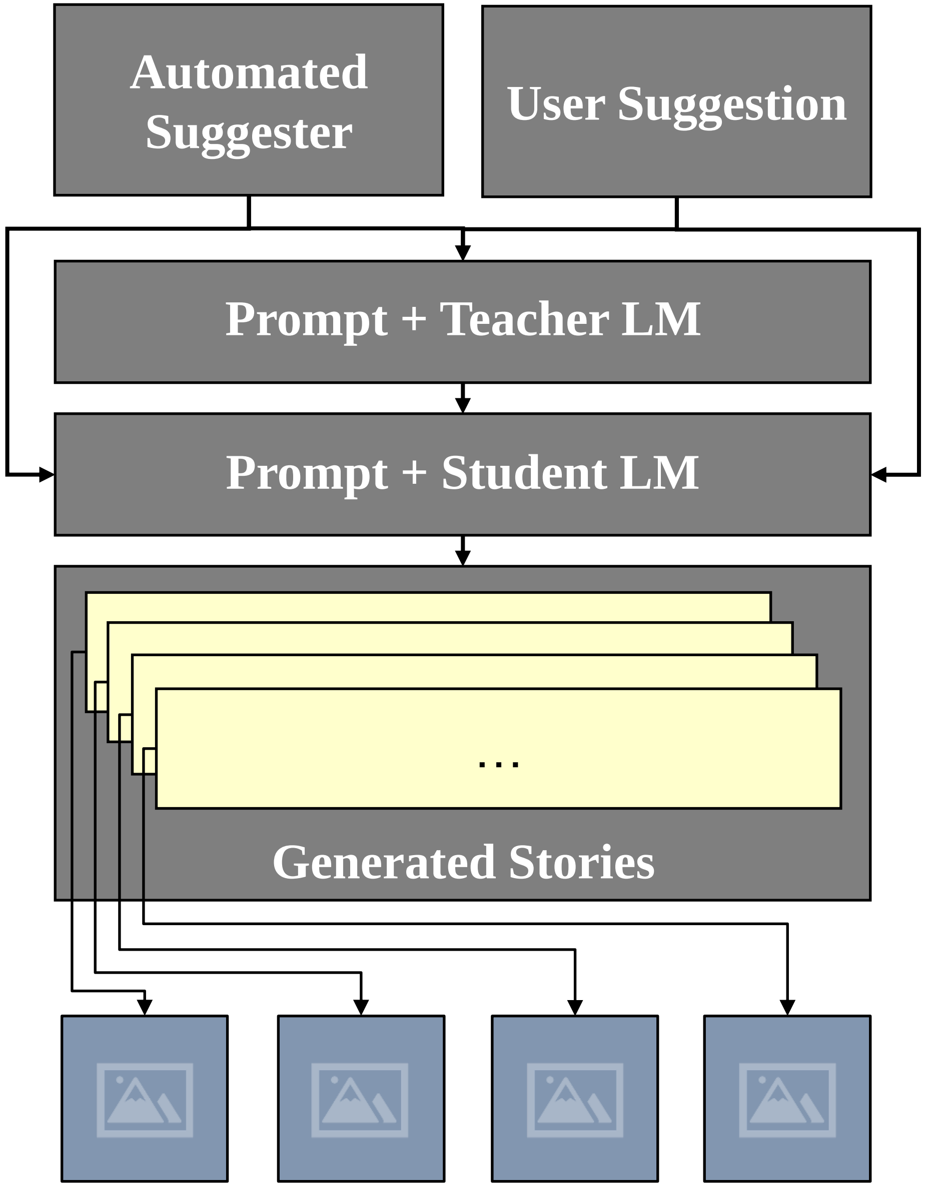
- **Neural-based story generation**
  - → Tambwekar, P.; et al. 2019. *Controllable Neural Story Plot Generation via Reward Shaping*.
  - → Lili Yao; et al. 2019. *Plan-and-Write: Towards Better Automatic Storytelling*.

- **Knowledge Distillation**
  - → Zhen Huang; et al. 2021. *Revisiting knowledge distillation: An inheritance and exploration framework*
  - → Yoon Kim; et al. 2016. *Sequence-level knowledge distillation*
  - → Victor Sanh; et al. 2020. *Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter*

Statement of the work

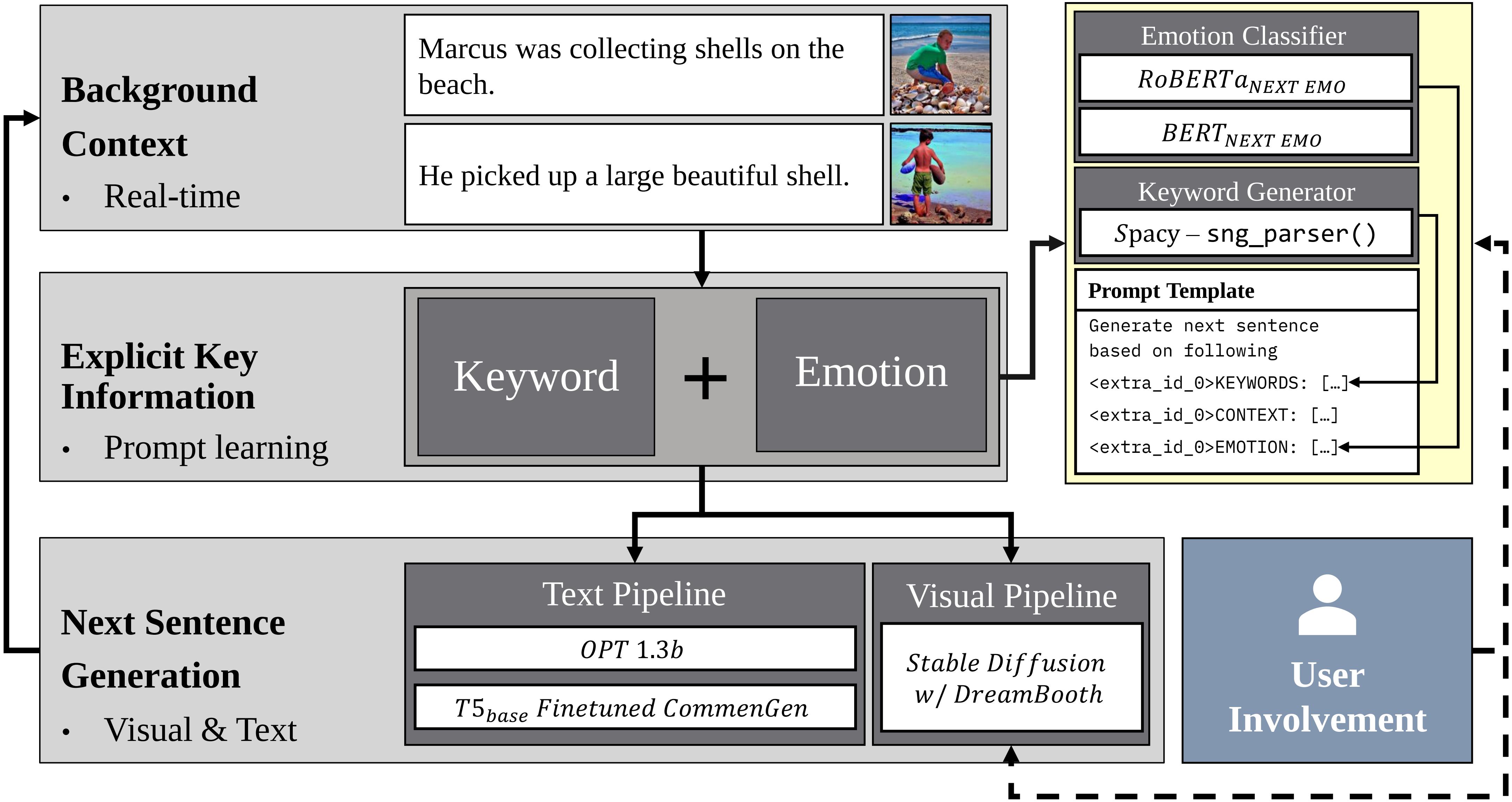
Interactive authoring system for story generation

- **Separate authoring need from story composition**
  - → User-controlled story generation with:
    - User's intention on key information
    - Priori knowledge of language models
- **Knowledge distillation in finetuning procedure**
  - → Transfer knowledge to a smaller student model
  - → Resulting model is more efficient with reduced computational resources
- **Image generation prompted by text content**
  - → Enhance the flexibility in generation
  - → Create immersion for design process
  - → Provide potential creation suggestions based on image content



Bring Language Model to Story Generation

Basic Principle: Visual + Text Interactive Generation



Implementation of an End-to-end Pipeline

Dataset Format & Story Generation Process

- **ROCStories (Story Cloze Test)**
  → A dataset evaluating narrative structure learning
    - ***98,159*** stories

  → Each story has a similar structure
    - Short narrative with ***five*** sentences
    - Logic & progression between sentences

- **Writing Prompts**
  → ***303,358*** pairs of writing prompts and human-written stories
  → Sourced from Reddit's `r/WritingPrompts` forum
  → Diverse topics, lengths, and ideas

index	Sentence
0	Marcus was collecting shells on the beach.
1	He picked up a large beautiful shell.
2	He put it in his pocket to save for later.
3	Suddenly he felt a sharp pinch.
4	A crab was inside the shell pinching his leg.

index	Sentence
0	Oswald decided to write a novel.
1	He worked on his book for several months.
2	Oswald took his book to a publisher.
3	The publisher rejected his book almost immediately.
4	Oswald decided to give up on writing.

Sample narratives quoted from Story Cloze

Implementation of an End-to-end Pipeline

Text pipeline

- **Encode emotion information as the input of model**

→ $\forall \vec{C} \in D, \vec{C} = \begin{bmatrix} e_1 \\ \vdots \\ e_8 \end{bmatrix}, e \in [0, 1]$

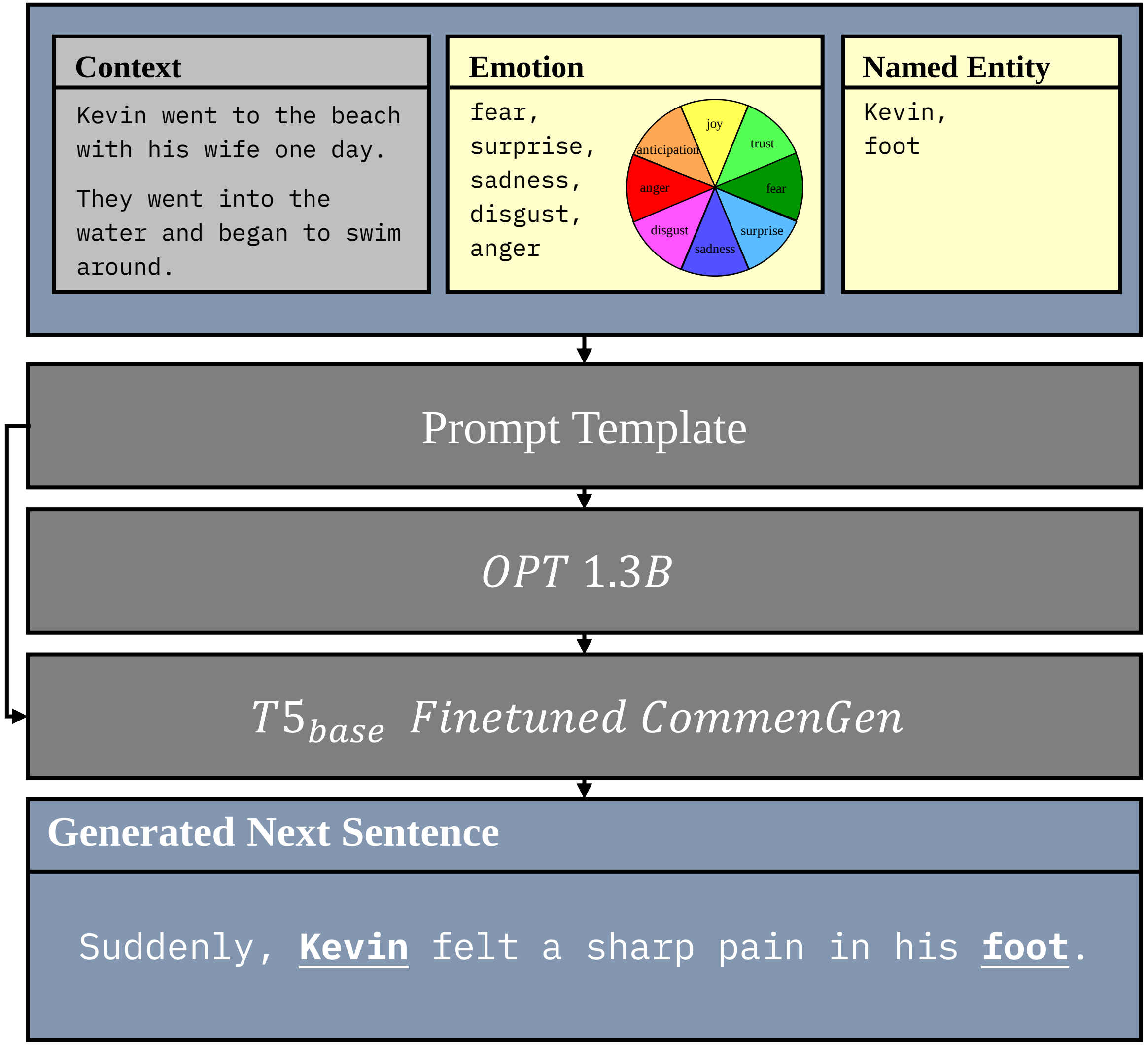
→ 8-label prediction using $RoBERTa_{large}$
  - ***78.3%*** Macro ROC AUC

→ *Objective:*
  - generate confidence level of each emotion categories based on input sentence

- **Generate a set of entities for next sentence**

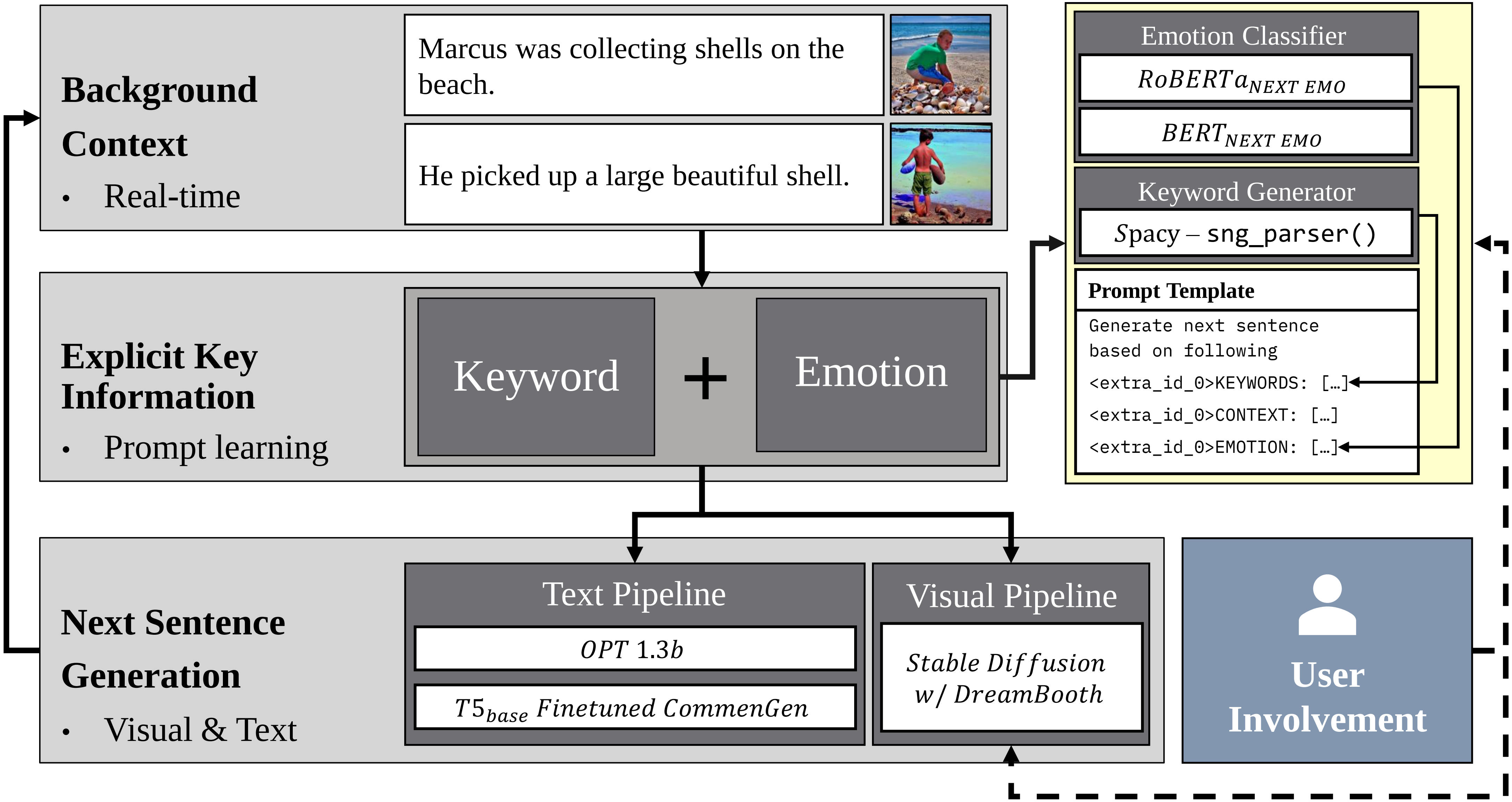
→ Named entity extraction using `sng_parser`

→ *Objective:*
  - generate a set of entities based on context



Bring Language Model to Story Generation

Basic Principle: Visual + Text Interactive Generation



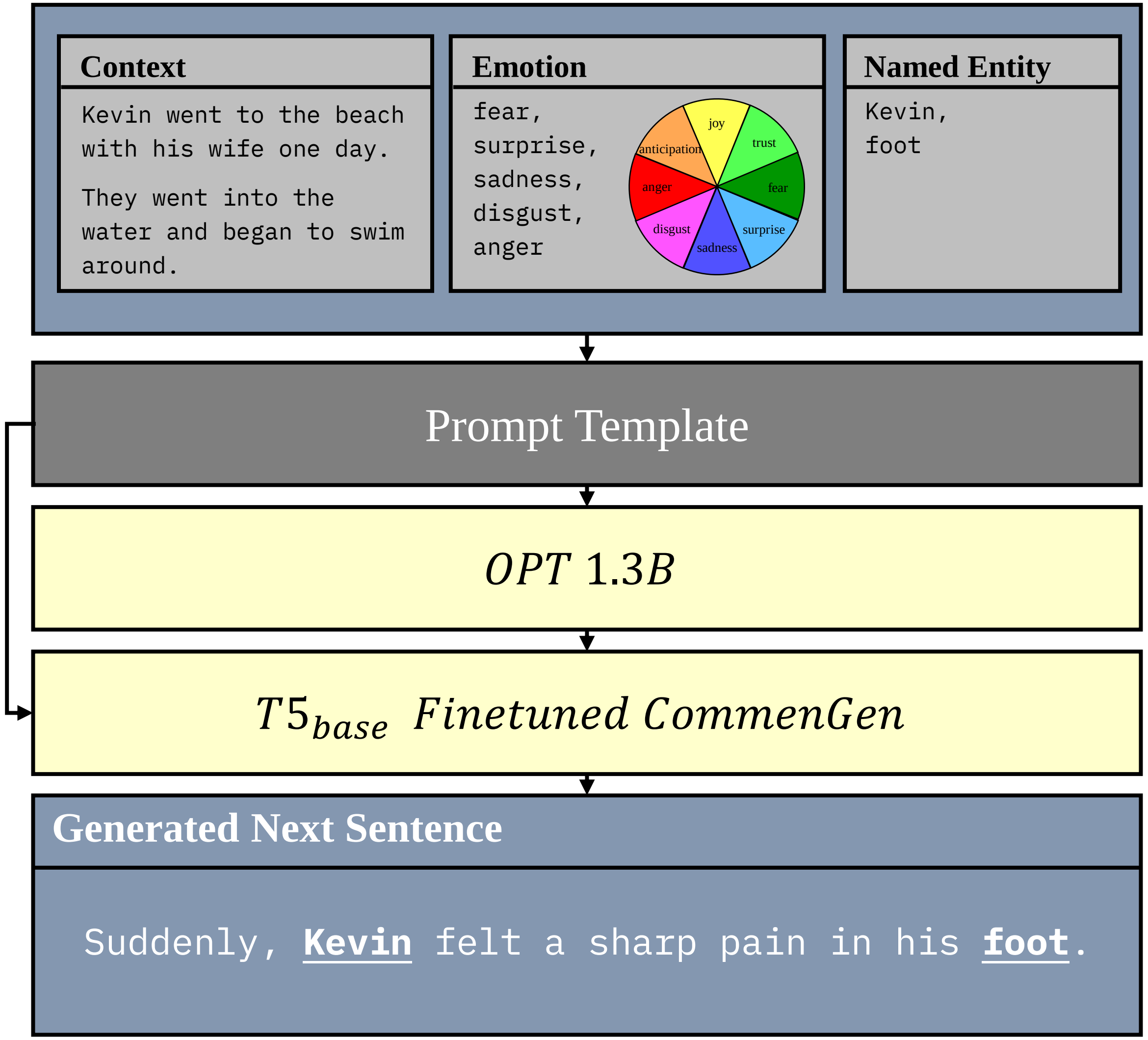
Implementation of an End-to-end Pipeline

Text pipeline – Teacher Model

- **Teacher network: OPT 1.3B**
  → Source of knowledge for the student model
  → Large, pre-trained language model
  → Trained on massive datasets; capture intricate

Model	# Parameters	Comparison
OPT 1.3B	1,300,000,000 (1.3b)	OPT is ***5.9x*** more than T5
T5-Base	220,000,000 (220m)	

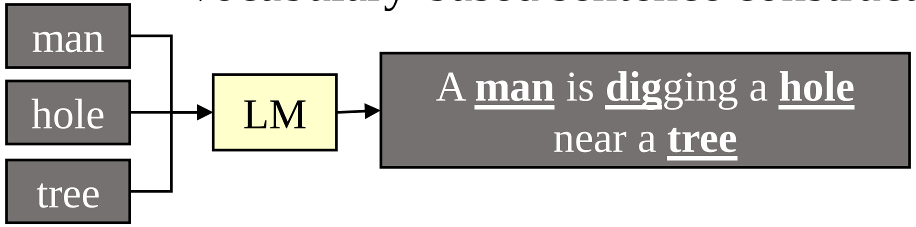
- **Architectural enhancements**
  → Different positional encoding for processing large inputs
  → Contributes to coherent, contextually appropriate text generation



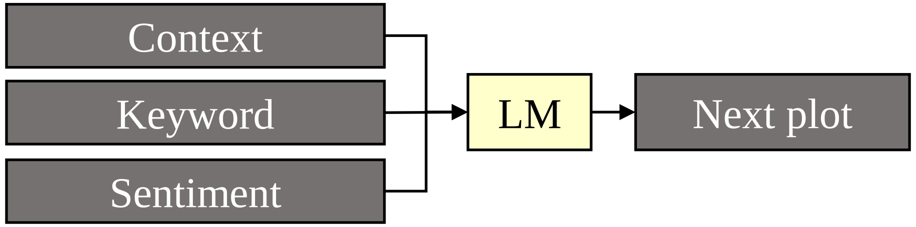
Implementation of an End-to-end Pipeline

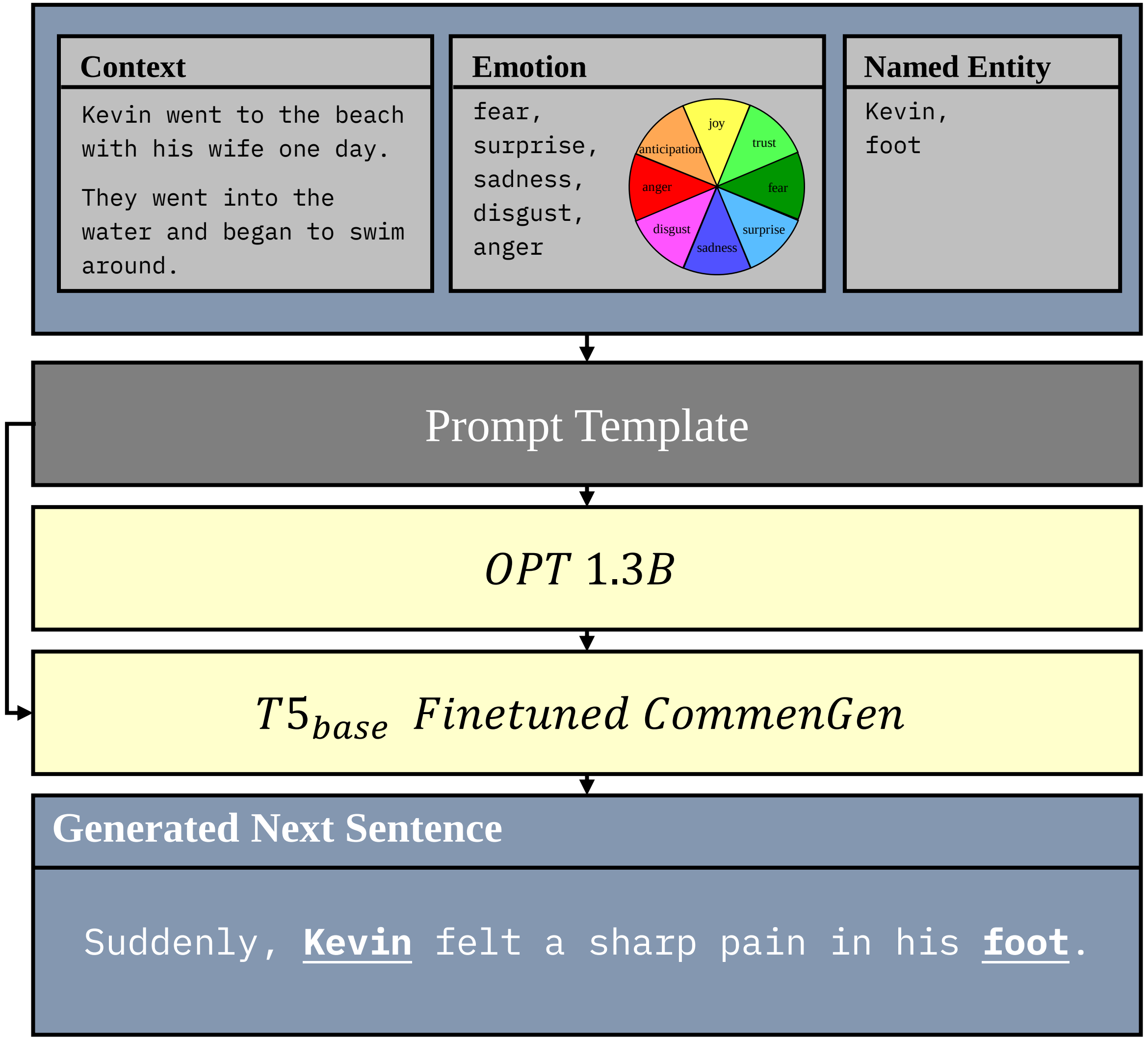
Text pipeline – Student Model

- **Multi-task pre-training on multiple datasets**
  - → CommenGen
    - `mrm8488/t5-base-finetuned-common_gen`
    - Constrained text generation task
    - *Objective:*
      - Vocabulary-based sentence construction



  - → Story Commonsense
    - Bring keywords and prompt template to task
    - *Objective:*
      - Generate next sentence based on input





Implementation of an End-to-end Pipeline

Text pipeline – Implementation & Evaluation

- **Knowledge distillation process**

  → $\mathcal{L}_{ce} = KL\left(\sigma_m\left(\frac{logits_{student}}{T}\right) \mid\mid \sigma_m\left(\frac{logits_{teacher}}{T}\right)\right)$

  → $\mathcal{L}_{mlm} = \ell_{cross\ entropy}(logits_{student},\ input\ ids)$

  → $\mathcal{L}_{cos} = \ell_{cosine\ loss}(h_{student}, h_{teacher}, target)$

  → $\mathcal{L}_{total} = \alpha_{ce}\mathcal{L}_{ce} + \alpha_{mlm}\mathcal{L}_{mlm} + \alpha_{cos}\mathcal{L}_{cos}$

  → Optimal: $T = 2,\ \alpha_{ce} = 0.4, \alpha_{mlm} = 0.5, \alpha_{cos} = 0.1.$

- **Evaluation metrics for generated content**

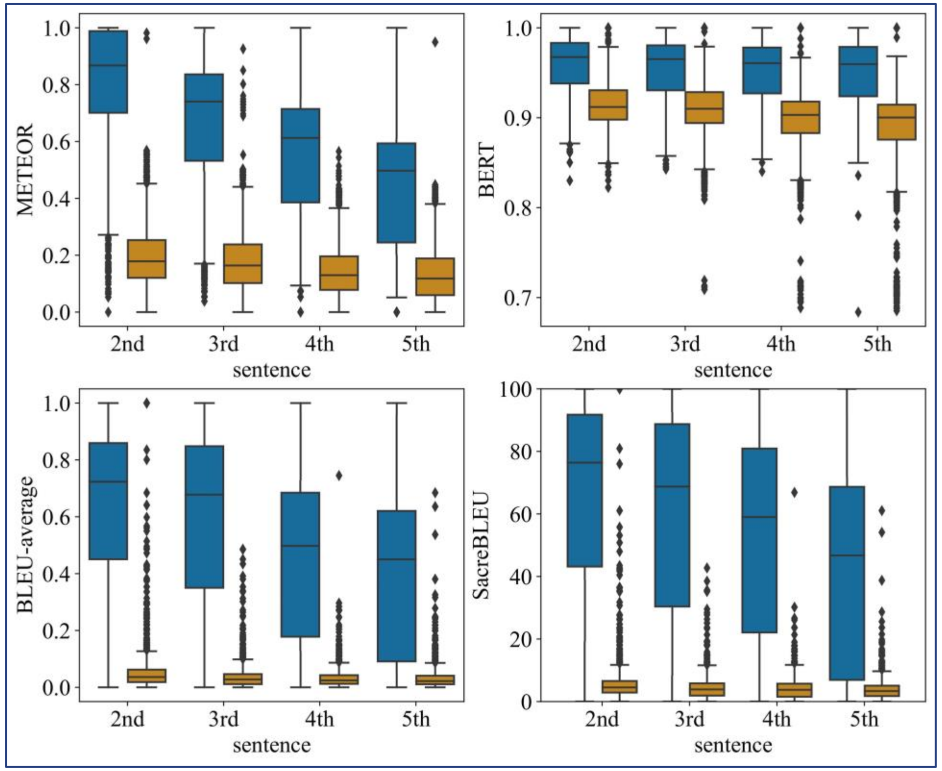
  → BLEU, BERT-score, METEOR, SacreBLEU

  → Measures exact word matching, stemming, synonym matching

- **Experimental results**

  → Improvements in evaluation metrics compared to baseline model

  → Emotions, keywords, and knowledge distillation enhance story generation quality



Demo

http://vsg-ek.herokuapp.com/

Future Works & Conclusion

Visual Story Generation Based on Emotional and Keyword Scheme

- **Two key directions**

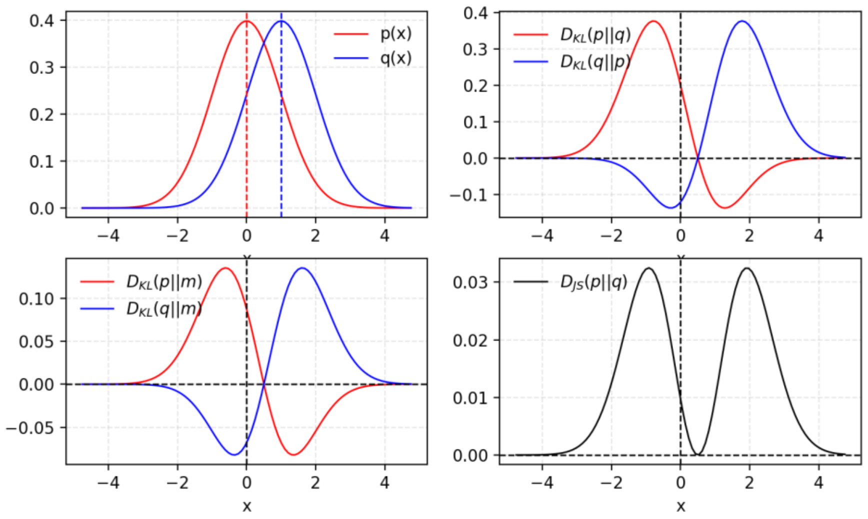
  → Replace KL divergence with *Jensen-Shannon divergence*

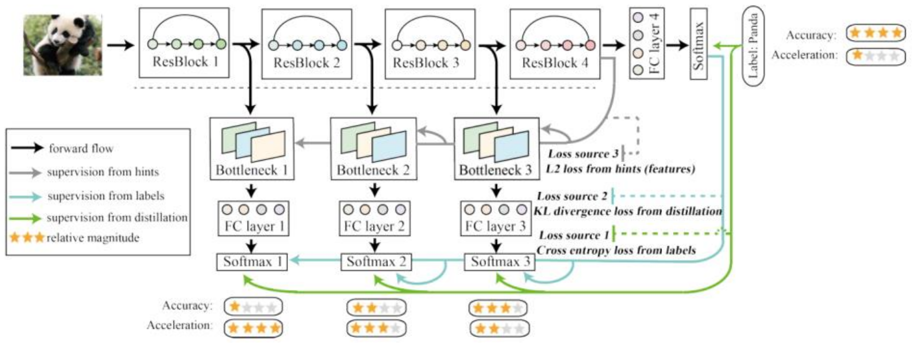
$$JS(P,Q)=\frac{1}{2}KL(P\,||\,M)+\frac{1}{2}KL(Q\,||\,M),M=\frac{1}{2}(P+Q)$$

  - ***Symmetry:*** consistent distance measure
  - ***Smoothness:*** robust to noise and local minima

  → Explore *self-distillation* for language generation

  - ***Regularization:*** better generalization, reduced overfitting
  - ***Simplification*** of implementation: reduced computational overhead





Reference

- Banerjee, S., and Lavie, A. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the ACL Work- shop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, 65–72. Ann Arbor, Michigan: Association for Computational Linguistics.
- Fan, A.; Lewis, M.; and Dauphin, Y. 2018. Hierarchical neural story generation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 889–898. Melbourne, Australia: Association for Computational Linguistics.
- Gerv ́as, P.; D ́ıaz-Agudo, B.; Peinado, F.; and Herv ́as, R. 2005. Story plot generation based on cbr. In Knowl. Based Syst.
- Ibrahim, B. I. E.; Eyharabide, V.; Le Page, V.; and Billiet, F. 2022. Few-shot object detection: Application to medieval musicological studies. Journal of Imaging 8(2).
- Kiros, R.; Zhu, Y.; Salakhutdinov, R. R.; Zemel, R.; Urta- sun, R.; Torralba, A.; and Fidler, S. 2015. Skip-thought vectors. In Cortes, C.; Lawrence, N.; Lee, D.; Sugiyama, M.; and Garnett, R., eds., Advances in Neural Information Processing Systems, volume 28. Curran Associates, Inc.
- Li, B.; Lee-Urban, S.; Johnston, G.; and Riedl, M. 2013. Story generation with crowdsourced plot graphs. Proceed- ings of the 27th AAAI Conference on Artificial Intelligence, AAAI 2013 598–604.
- Lin, C.-Y., and Och, F. J. 2004. ORANGE: a method for evaluating automatic evaluation metrics for machine trans- lation. In COLING 2004: Proceedings of the 20th Interna- tional Conference on Computational Linguistics, 501–507. Geneva, Switzerland: COLING.
- Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C. L.; and Doll ́ar, P. 2014. Microsoft coco: Common objects in context.
- Lin, T.-Y.; Doll ́ar, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2016. Feature pyramid networks for object detection.
- Lin, B. Y.; Zhou, W.; Shen, M.; Zhou, P.; Bhagavatula, C.; Choi, Y.; and Ren, X. 2019. Commongen: A constrained text generation challenge for generative commonsense rea- soning. arXiv preprint arXiv:1911.03705.
- Liu, D.; Li, J.; Yu, M.-H.; Huang, Z.; Liu, G.; Zhao, D.; and Yan, R. 2020. A character-centric neural model for auto- mated story generation. Proceedings of the AAAI Confer- ence on Artificial Intelligence 34(02):1725–1732.
- Meehan, J. R. 1977. Tale-spin, an interactive program that writes stories. In IJCAI.
- Montfort, N. 2007. Generating narrative variation in inter- active fiction. Dissertations available from ProQuest.
- Mostafazadeh, N.; Chambers, N.; He, X.; Parikh, D.; Batra, D.; Vanderwende, L.; Kohli, P.; and Allen, J. 2016a. A corpus and evaluation framework for deeper understanding of commonsense stories. arXiv preprint arXiv:1604.01696.
- Mostafazadeh, N.; Grealish, A.; Chambers, N.; Allen, J.; and Vanderwende, L. 2016b. CaTeRS: Causal and temporal re- lation scheme for semantic annotation of event structures. In Proceedings of the Fourth Workshop on Events, 51–61. San Diego, California: Association for Computational Linguis- tics.
- Onega, S., and Landa, J. A. G. 2014. Narratology: an intro- duction. Routledge.
- Post, M. 2018. A call for clarity in reporting BLEU scores. In Proceedings of the Third Conference on Machine Trans- lation: Research Papers, 186–191. Belgium, Brussels: As- sociation for Computational Linguistics.
- Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning transferable visual models from natural language supervision.

Q & A

1. State-of-the-art Work of Short Narrative Creation

- Separate authoring need from story composition need
- Prompt-based learning for image generation

2. Statement of the work

- Basic Principle: Visual + Text Interactive Generation

3. Implementation of an End-to-end Pipeline

- Part 1 – Suggester & Prompt Learning

  → *Named Entities + Sentiments Label + Context = Controlled story…*

- Part 2 – Knowledge distillation

  → $\mathcal{L}_{total} = \alpha_{ce}\mathcal{L}_{ce} + \alpha_{mlm}\mathcal{L}_{mlm} + \alpha_{cos}\mathcal{L}_{cos}$

4. Demo - http://vsg-ek.herokuapp.com/

- Modify suggested keywords for real-time plot changes based on requirements

5. Future Improvement & Conclusion

- *Jensen-Shannon divergence* & *self-distillation…*