



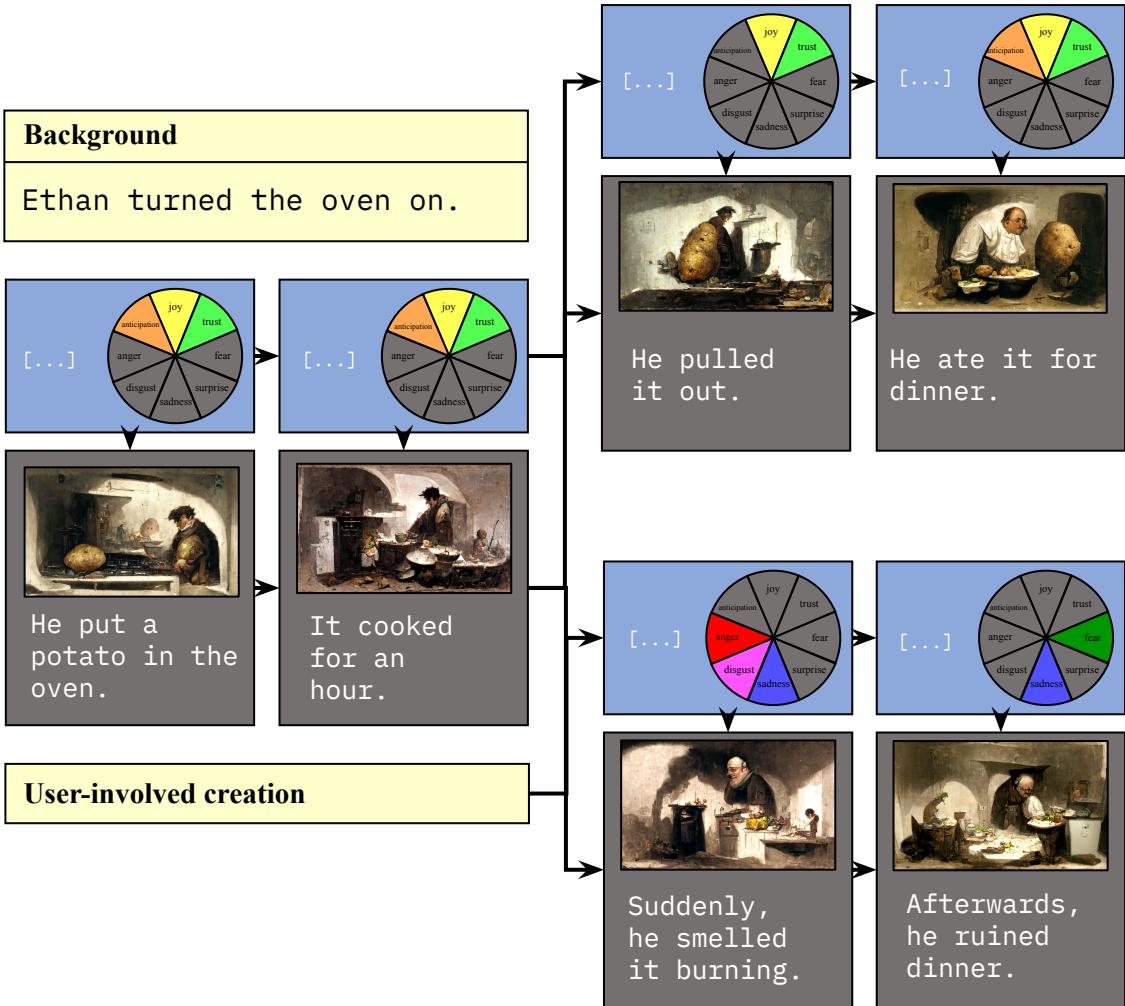
Rensselaer

why not change the world?®

Visual Story Generation Based on Emotional and Keyword Scheme

Yuetian Chen, Ruohua Li,
Bowen Shi, Peiru Liu, &
Mei Si

Rensselaer Polytechnic Institute

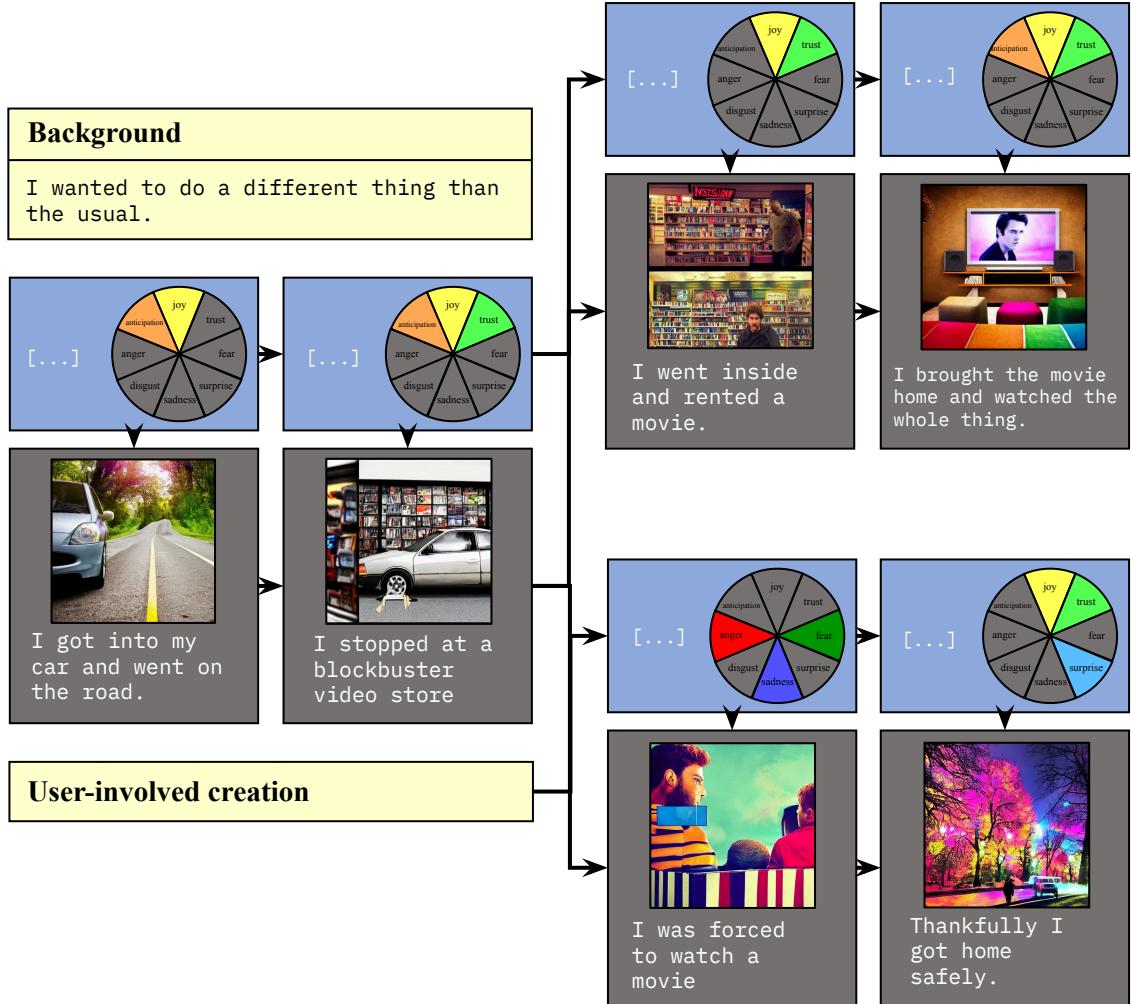




Visual Story Generation Based on Emotional and Keyword Scheme

Yuetian Chen, Ruohua Li,
Bowen Shi, Peiru Liu, &
Mei Si

Rensselaer Polytechnic Institute



Statement of the work

Interactive authoring system for story generation

- **Automatic story generation**

- **AI Story Generators**

- Lili Yao, Nanyun Peng, Ralph Weischedel, Kevin Knight, Dongyan Zhao, and Rui Yan. 2019.

Plan-and-write: Towards better automatic storytelling.

- Jasper: <https://www.jasper.ai/>

- Rytr: <https://rytr.me/>

- **Image Generation**

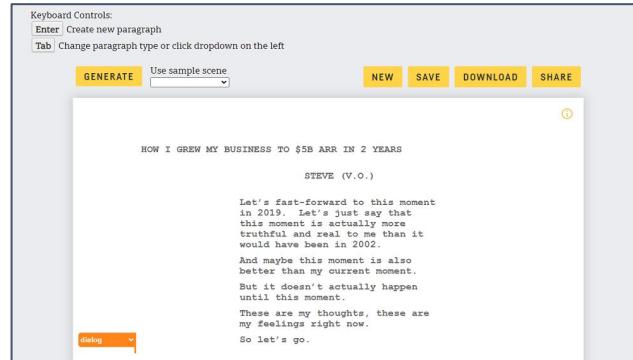
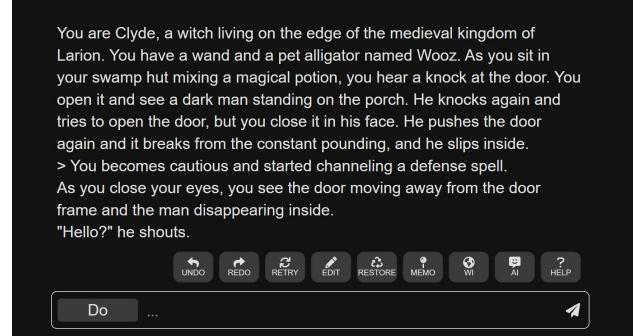
- Prafulla Dhariwal and Alex Nichol. 2021.

Diffusion models beat gans on image synthesis

Motivation

Interactive authoring system for story generation

- **Comercial AI text generators**
 - AI Dungeon: <https://play.aidungeon.io/>
 - Jasper: <https://www.jasper.ai/>
 - Rytr: <https://rytr.me/>
- **...**
- **No 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 sturytelling



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*.
 - **Fintuning language model**
 - Raffel, C.; et al. 2020.
Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer.
 - Francisca Adoma Acheampong; et al. 2020.
Text-based emotion detection: Advances, challenges, and opportunities.
- **Image generation and detection**
 - Joseph Redmon; et al. 2016. *You Only Look Once: Unified, Real-Time Object Detection*
 - Prafulla Dhariwal; et al. 2021. *Diffusion Models Beat GANs on Image Synthesis*

Related Work

Interactive authoring system for story generation

- Comercial AI text generators
 - Jasper: <https://www.jasper.ai/>
 - Rytr: <https://rytr.me/>
 - CopyAI: <https://www.copy.ai/>
 - AI Writer: <https://ai-writer.com/>

...

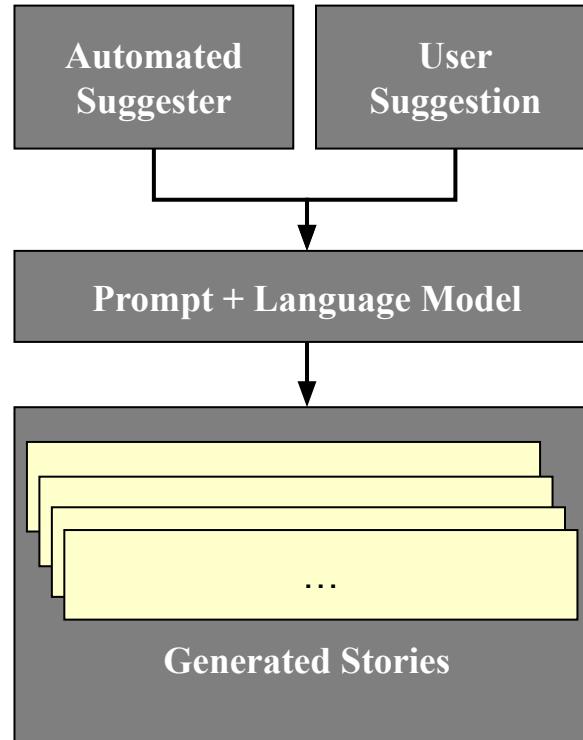
The screenshot shows the Rytr AI writing interface. At the top, there are dropdown menus for 'Select language' (set to 'us English') and 'Select tone' (set to 'Casual'). Below these are sections for 'Choose use case' (set to 'Blog Idea & Outline') and 'Primary keyword' (set to 'test'). A text input field contains the word 'test'. To the right, a preview window displays the generated content: 'The Comprehensive Guide to Taking a Test' followed by several paragraphs of text under headings like 'Introduction: What is a Test?' and 'How to Take a Good Test'.

This screenshot shows a different document in the Rytr interface. The title is 'HOW I GREW MY BUSINESS TO \$5B ARR IN 2 YEARS' attributed to 'STEVE (V.O.)'. The content includes a paragraph about fast-forwarding to 2019 and another about a moment in 2019 being better than the current one. At the bottom, there is a 'dialog' button.

Statement of the work

Interactive authoring system for story generation

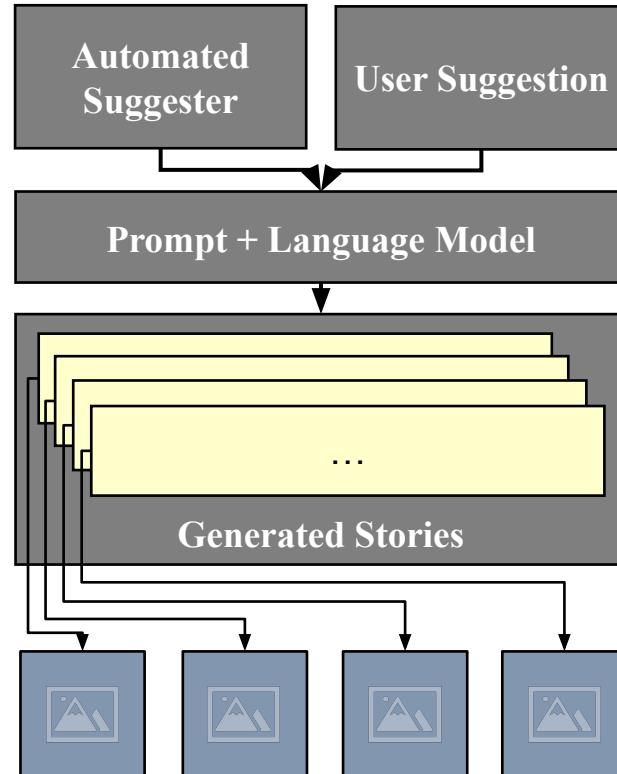
- **Separate authoring need from story composition need**
 - A balance between:
 - User immersion and controllability
 - Utilization of a priori knowledge of language models
- **Prompt-based learning for image generation**
 - Enhance the flexibility in generation
 - Create immersion for design process
 - Provide potential creation suggestions based on image content



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
- **Image generation prompted by text content**
 - Enhance the flexibility in generation
 - Create immersion for design process
 - Provide potential creation suggestions based on image content



Statement of the work

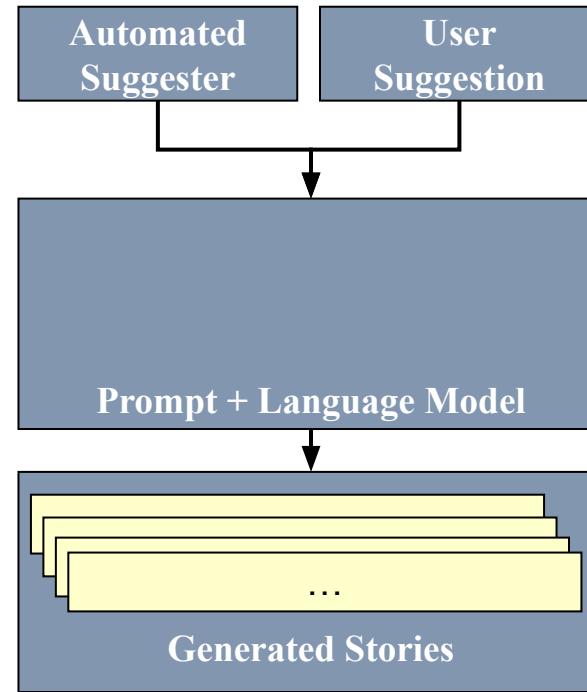
Interactive authoring system for story generation

- **Separate authoring need from story composition need**

- A balance between:

- User immersion and controllability
 - Utilization of a priori knowledge of language models

-



Current Situation of Short Narrative Creation

Fundamental Challenges of Automatic Context-based Generation

- **Consistency**

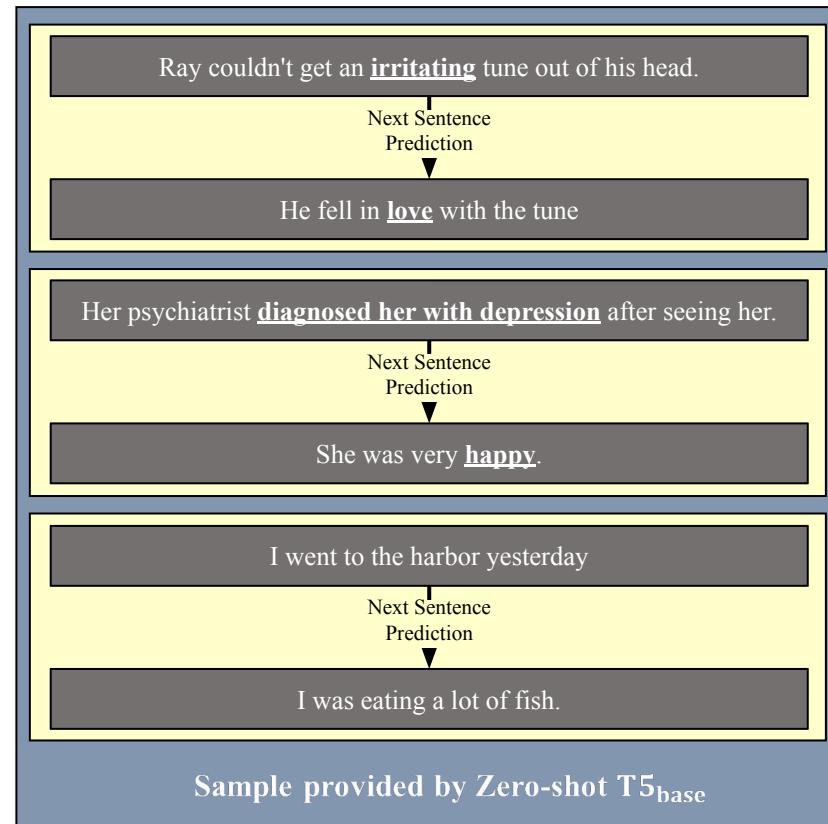
- No internal contradiction of story elements
 - properties & events of word & characters

- **Coherence**

- Events should be logically explainable
 - Help reader build up causal representation of events

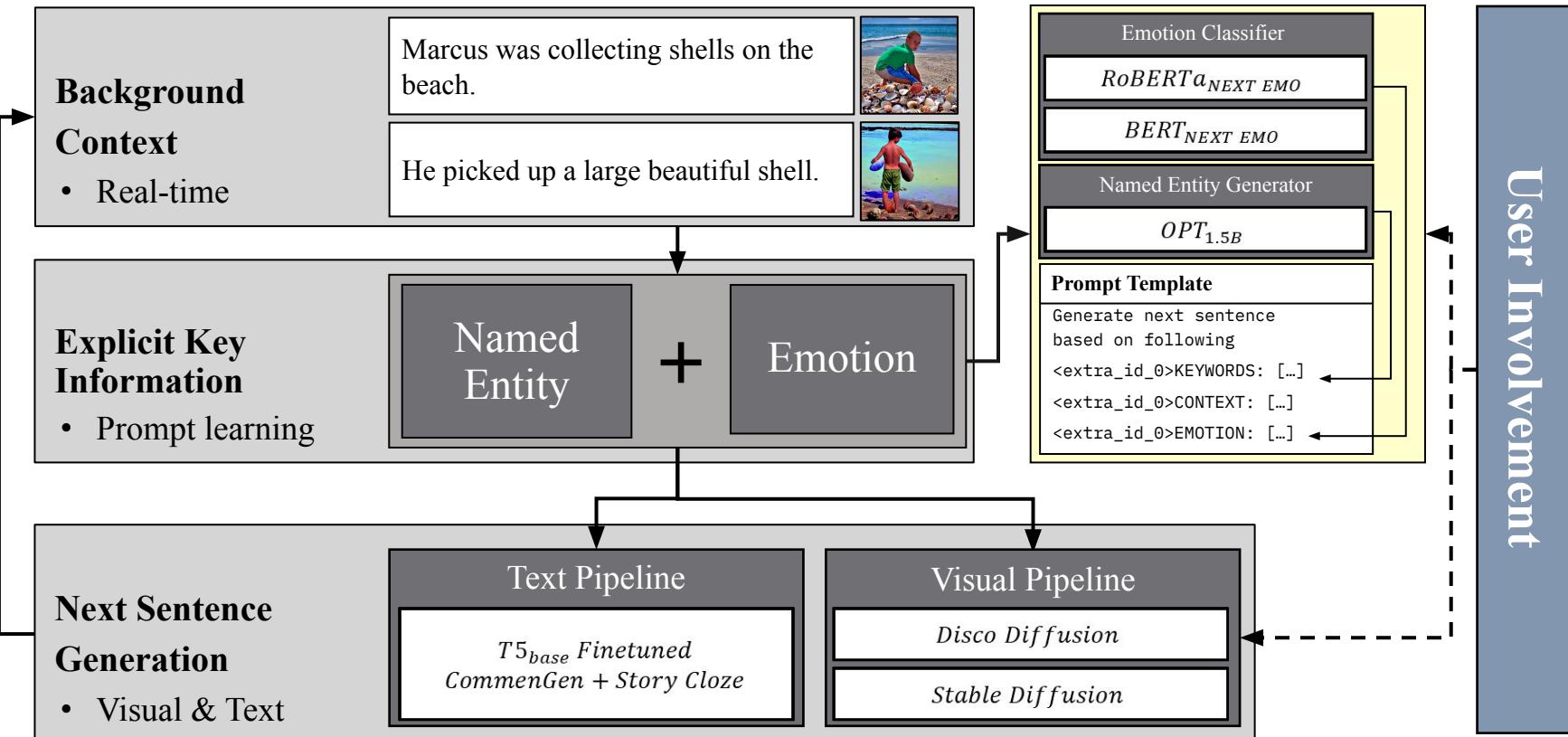
- **Motivation**

- Story ≠ Predicting behavior
 - Justified surprise is fine
 - Author provides action & make actions reasonable to reader



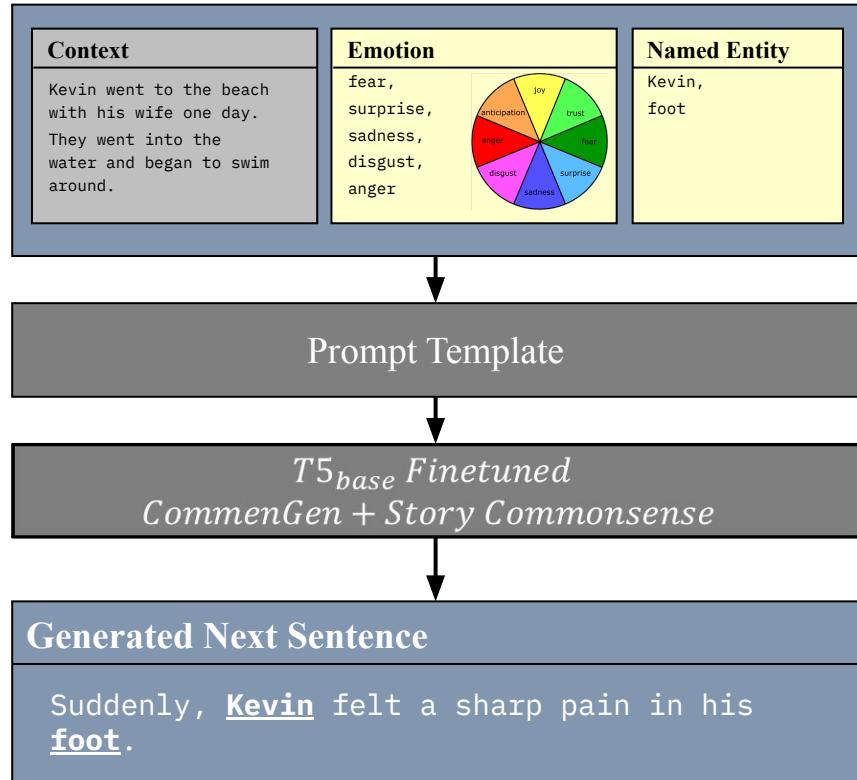
Statement of the work

Basic Principle: Visual + Text Interactive Generation



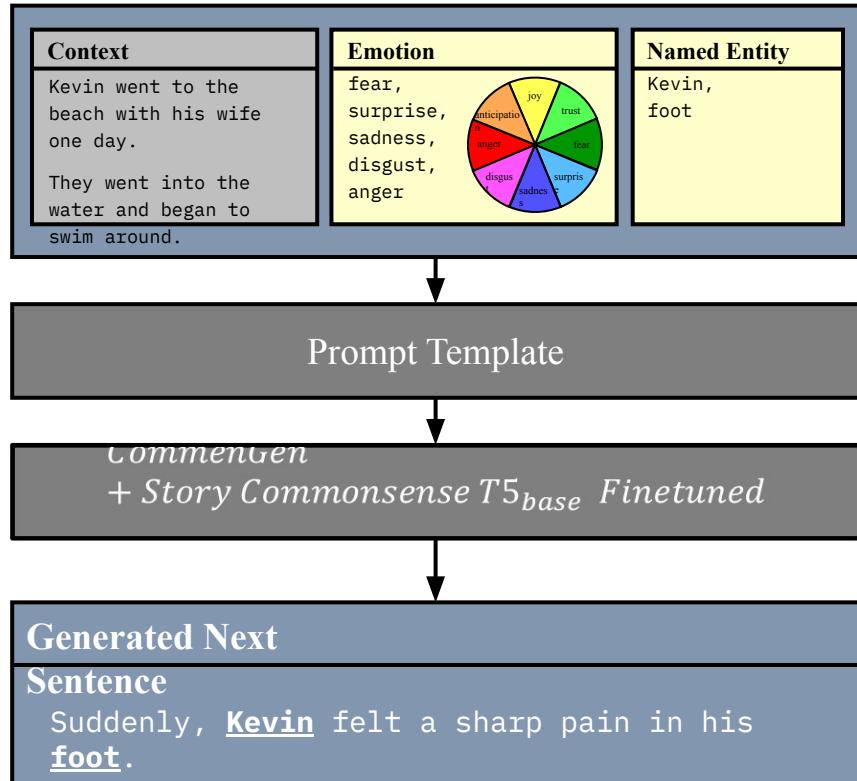
Implementation of an End-to-end Pipeline

Text pipeline



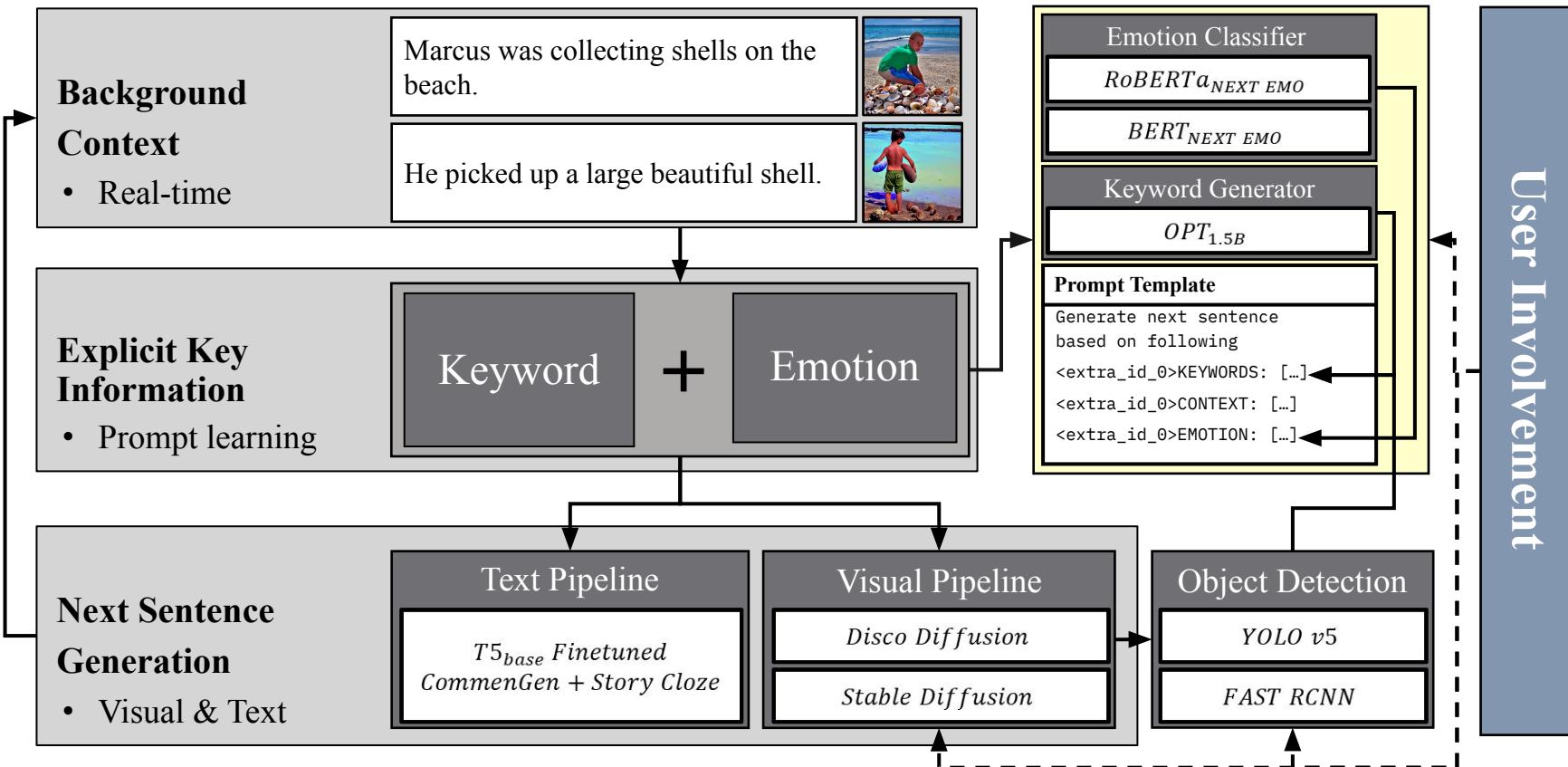
Implementation of an End-to-end Pipeline

Text pipeline



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
- Each story has a similar structure
 - Consists of five sentences
 - Contain twists of events in the middle stages

- **Story Commonsense (dataset)**

- Use subset of Story Cloze as the basis
 - 15,000 stories
- Provide categorical annotation for each plot
 - joy, trust, fear, surprise, sadness, anticipation, anger & disgust

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.

A sample narrative quoted from Story Cloze

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

- **Story Commonsense (dataset)**

- Use subset of Story Cloze as the basis
 - 15,000 stories
- Provide categorical annotation for each plot
 -



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

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

\rightarrow 8-label prediction using *RoBERTa_{large}*

- **67.8%** Macro ROC AUC

\rightarrow Objective: generate confidence level of each emotion categories based on input sentence

- Generate a set of entities for next sentence

\rightarrow Generated based on context

\rightarrow Training under the full ROCStories dataset

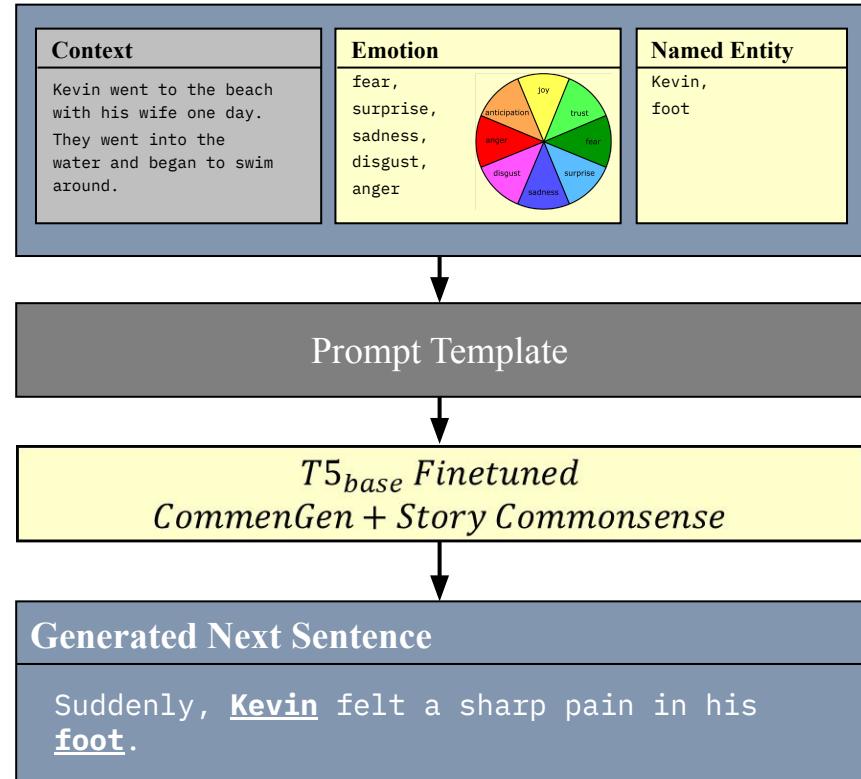
- 101,903 valid narrative samples

- Named entity extraction using `sng_parser`

\rightarrow Generated using *OPT_{1.3B}*

\rightarrow Objective:

- generate a set of entities based on context



Implementation of an End-to-end Pipeline

Text pipeline

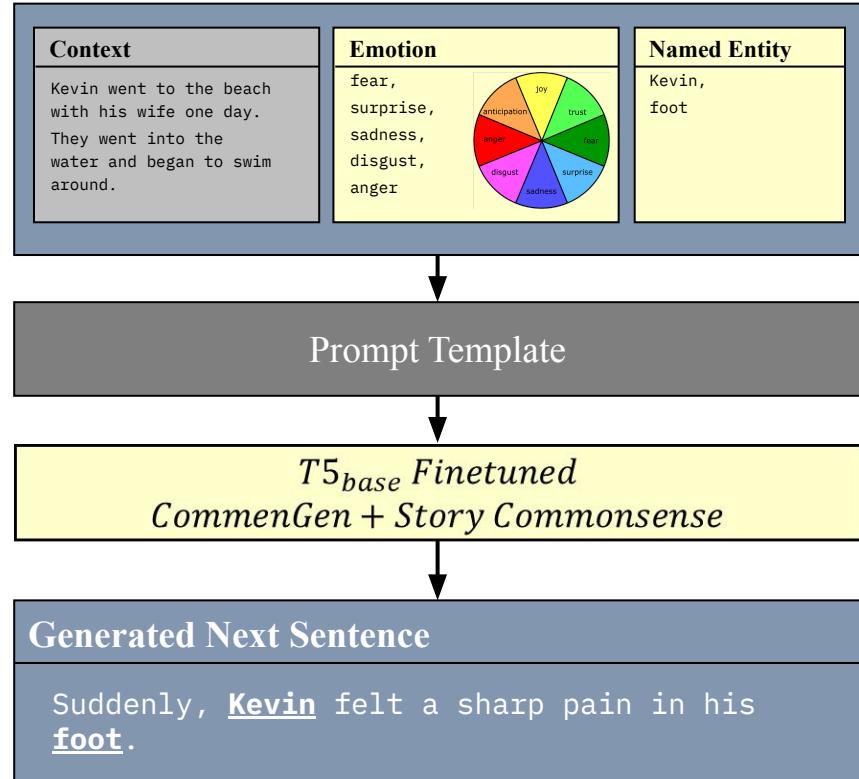
- Encode emotion information as the input of model

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

- 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

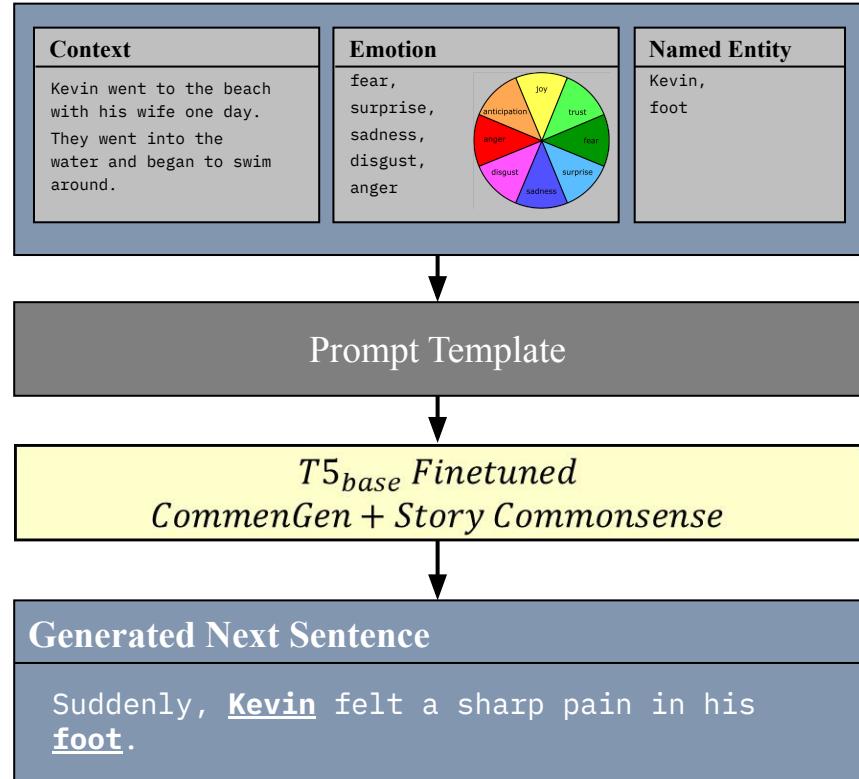
- Generated based on context
- Training under the full ROCStories dataset
 - 101,903 valid narrative samples
 - Named entity extraction using *sng_parser*
- Generated using *OPT_{1.3B}*
- *Objective*:
 - generate a set of entities based on context



Implementation of an End-to-end Pipeline

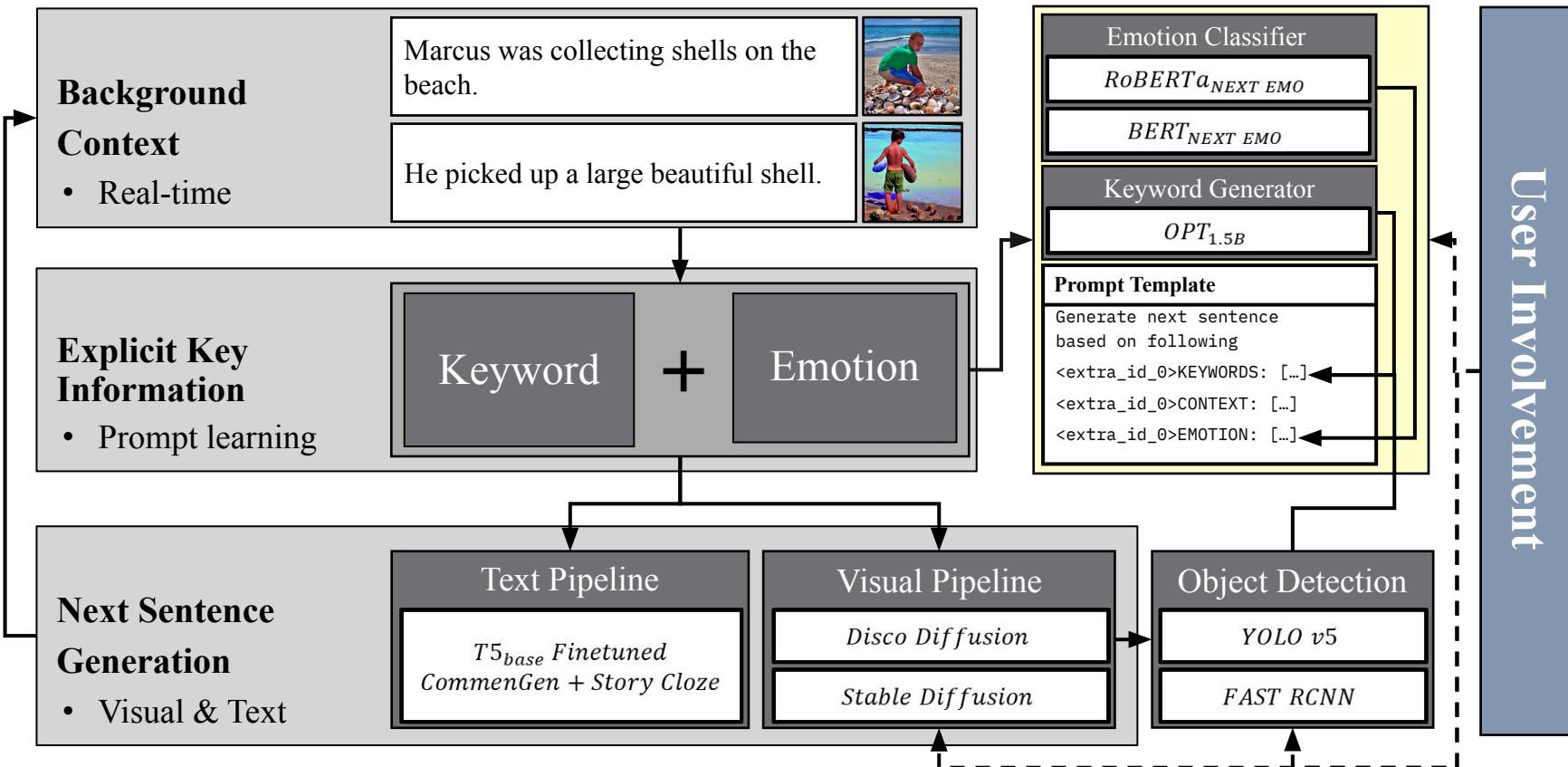
Text pipeline

- Multi-task training on multiple datasets
 - CommenGen
 - 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
- Interactive short narrative generation
 - Based on user-provided story background
 - Generate next sentence based on a triad of data
 - Context
 - Named entity
 - Plutchik's wheel of emotions label



Bring Language Model to Story Generation

Basic Principle: Visual + Text Interactive Generation



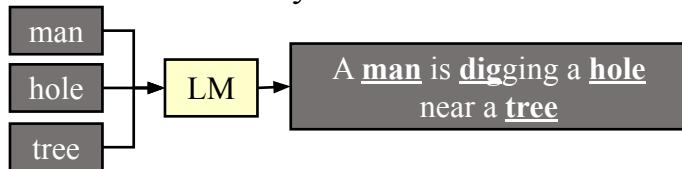
Implementation of an End-to-end Pipeline

Text pipeline

- Multi-task training on multiple datasets

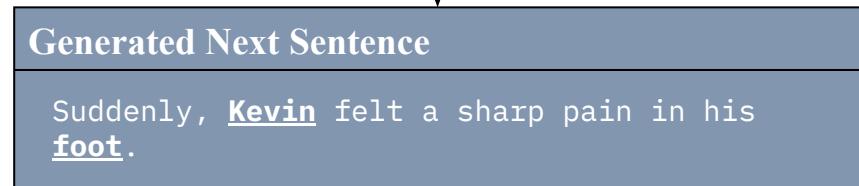
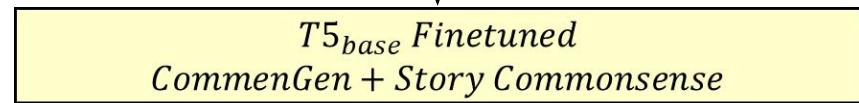
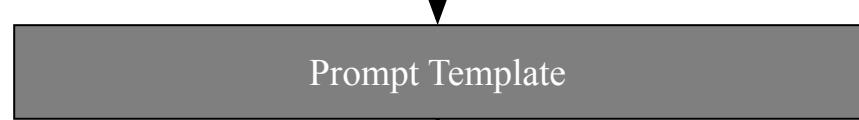
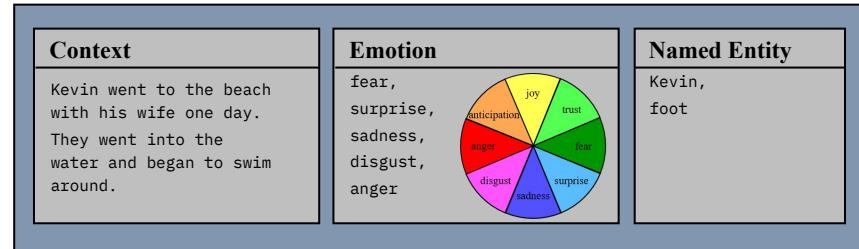
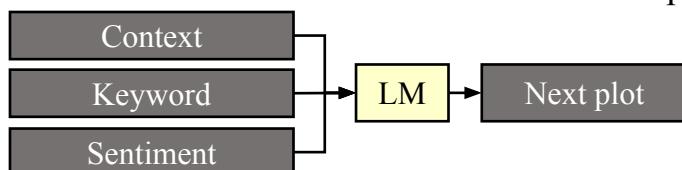
- CommenGen

- 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



Advantages on Narrative Co-generation

Visual pipeline

- **Image generation from pure narratives**
 - Enhance the flexibility in generation
 - Create immersion for design process
 - Provide potential creation suggestions based on image content

Advantages on Narrative Co-generation

Visual pipeline

- **Image generation from pure narratives**

- Prompt template

```
"a story about [CONTEXT]"
```

- Disco Diffusion

- “*Carl Spitzweg*”

- Stable Diffusion

- “*Colored Photo*”

- **Image feature extraction by target detection**

- Results combining from YOLOv5 & Faster R-CNN

- Confidence level as a recommendation strength metric



“Ray gathered his friends to tell them a funny joke he heard.”

Advantages on Narrative Co-generation

Visual pipeline

- Parallel to the text pipeline
 - Enhance the flexibility in generation
 - Create immersion for design process
- Image generation from pure narratives
 - *Disco Diffusion*
 - Produce relevant figures with artistic quality
 - *Carl Spitzweg* splash-art style
 - *Stable Diffusion*
 - More robust drawing for character structure
 - Realistic photography style
- Image feature extraction by target detection
 - Results combining from YOLOv5 & Faster R-CNN
 - Provide potential creation suggestions based on image content



“Ray gathered his friends to tell them a funny joke he heard.”

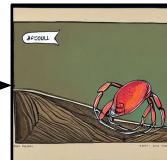
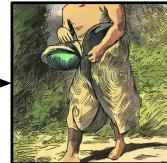
Demo

<http://int-2022-visual-story-gen.uw.r.appspot.com/>

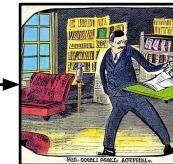
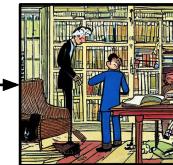
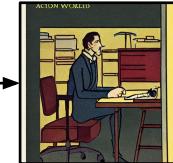
Advantages on Narrative Co-generation

Demo - sample generation results

#	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.



#	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.



...

Future Works & Conclusion

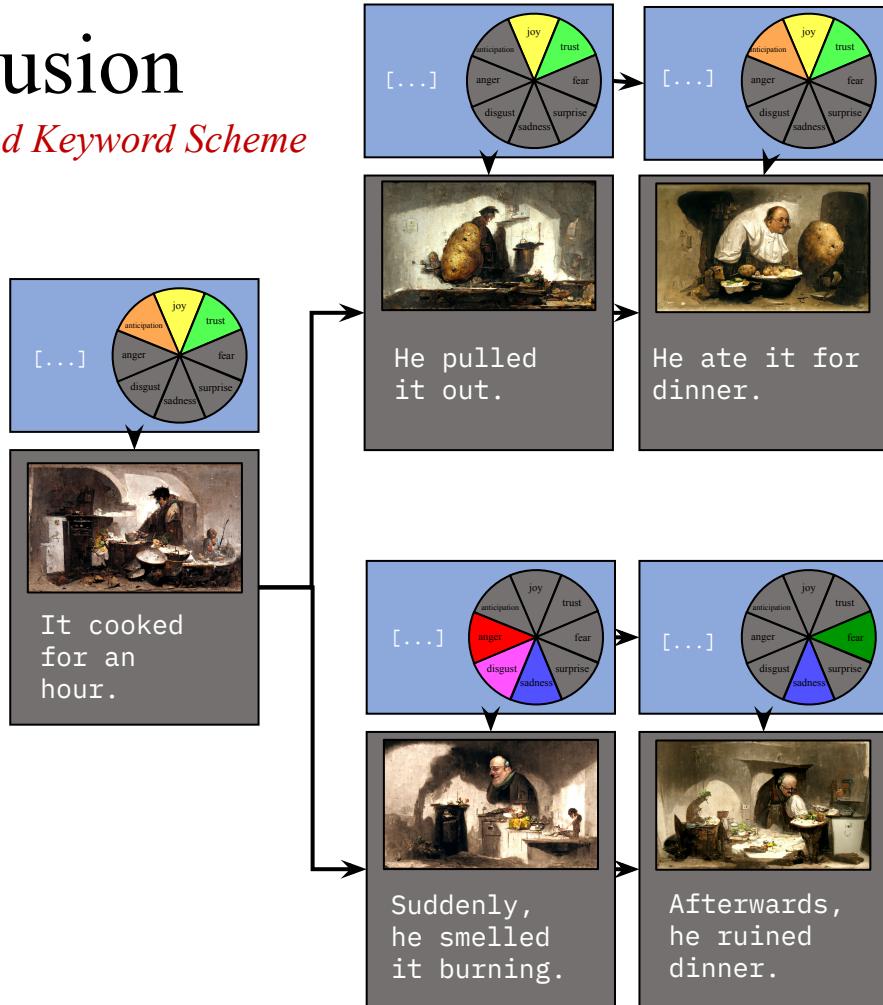
Visual Story Generation Based on Emotional and Keyword Scheme

- **Whats next?**

- Increase the scale of generated stories
- Enhance the text and visual coupling

- **In conclusion**

- Visual story co-creation experience
 - Narrative generation
 - Image generation



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 – Text pipeline
 - *Named Entities + Sentiments Label + Context = Controlled story...*
- Part 2 – Image pipeline
 - Generated stylized illustrations using *Stable Diffusion, Disco Diffusion ...*

4. Advantages on Narrative Co-generation

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

5. Future Improvement on Prompting and Framework Structure

- Scale increasing & prompting optimization...