CharMoral: A Character Morality Dataset for Morally Dynamic Character Analysis in Long-Form Narratives

Suyoung Bae   Gunhee Cho   Boyang Li†   Yungyung Cheong†

Sungkyunkwan University, South Korea
Nanyang Technological University, Singapore

COLING 2025

Suyoung Bae

sybae01@skku.edu

Sungkyunkwan University

Contents

1. Introduction
2. Method
3. Experiment & Analysis
4. Conclusion

1. Introduction

Our final goal is …

**Analyzing the dynamic moral behavior
of fictional characters**

1. Introduction (Cont.)

- Human Values and Fictional Stories
    - Fictional narratives often mirror societal norms, values, and dilemmas, providing a rich ground for exploring ethical behavior
    - Through automatic analyzing the ethical behavior of characters, ...
        - AI agents can effectively learn to align with human ethical standards.
        - We can use as a tool to analyze story engagement.
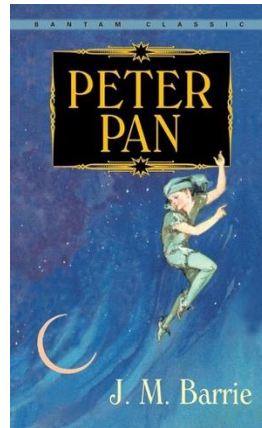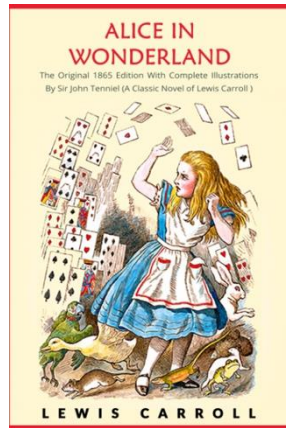
1. Introduction (Cont.)

Our task is extremely hard!

1. Moral reasoning of actions is highly dependent on context

2. Morality is not static, it changes

3. No existing datasets in long-form narratives

2. Method : Overview

1. Propose new dataset : CharMoral dataset
   - Characters' morality labels **across long-form narrative**
   - Morality labels **reflecting the context**
   - Can analyzing **moral dynamics**

2. To build dataset, we propose the **dataset construction framework**

3. We introduce a new score, **moral dynamic score** of a character

2. Method : CharMoral dataset

- Example of our CharMoral dataset
  - source dataset : IMDB Spoiler Dataset (https://www.imdb.com/)
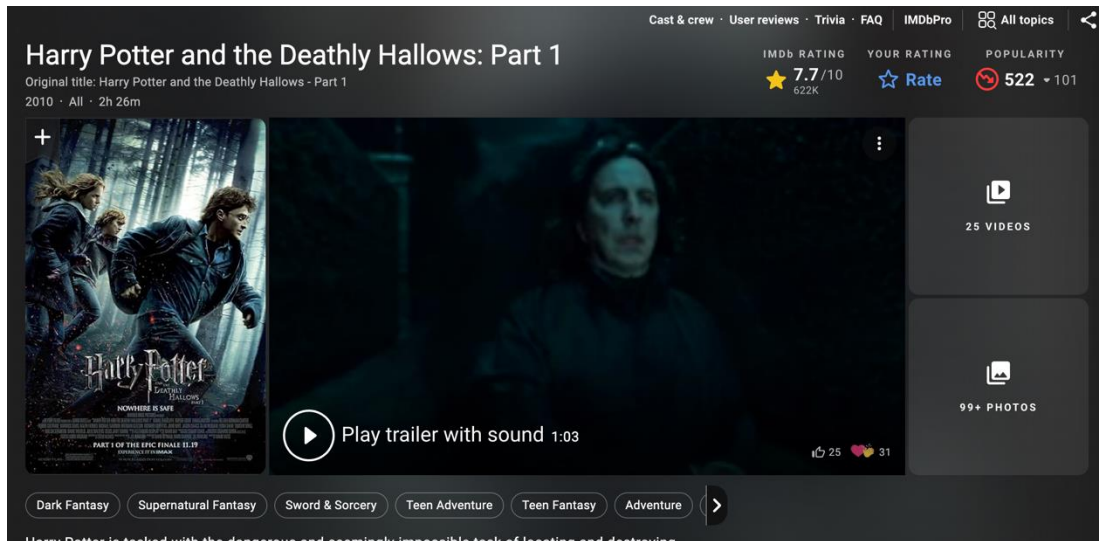
- story name: **Harry Potter and the Deathly Hallows**
- id: **tt0926084**
- rating: **7.7**
- story length: **1240**
- segment num : **54**
- genres: **'Adventure', 'Family', 'Fantasy', 'Mystery'**

2. Method : CharMoral dataset (Cont.)

- Example of our CharMoral dataset

Segment 0

After burying **[Dobby]** at the garden of the Shell cottage, ==**[Harry Potter]** convinces **[Griphook]**==

==to help them get to Lestrange's vault in **[Gringotts]**==, to retrieve one of **[Voldemort]**'s

==Horcruxes in exchange for Godric Gryffindor's Sword.== Meanwhile, Ollivander , the

[Wandmaker] warns **[Harry]** that he won't stand a chance with **[Voldemort]** who has the

Elder Wand. They arrived in **[Gringotts]**, Hermione disguised as Bellatrix , using a Polyjuice

Potion, **[Ron]** disguised as a random wizard while ==**[Harry]** and **[Griphook]** go under the==

==Invisibility Cloak.==

== == : Action

Segment 1

⋮

Segment 63

2. Method : CharMoral dataset (Cont.)

- Example of our CharMoral dataset

Segment 0

After burying **[Dobby]** at the garden of the Shell cottage, **[Harry Potter]** convinces **[Griphook]** to help them get to Lestrange's vault in **[Gringotts]**, to retrieve one of **[Voldemort]**'s Horcruxes in exchange for Godric Gryffindor's Sword. Meanwhile, Ollivander , the **[Wandmaker]** warns **[Harry]** that he won't stand a chance with **[Voldemort]** who has the Elder Wand. They arrived in **[Gringotts]**, Hermione disguised as Bellatrix , using a Polyjuice Potion, **[Ron]** disguised as a random wizard while **[Harry]** and **[Griphook]** go under the Invisibility Cloak.

: Action

Segment 1

...

Segment 63

Character:
Harry

Situation:
After burying Dobby at the garden of the Shell cottage, **[Harry]** Potter convinces Griphook to help them get to Lestrange's vault in Gringotts.

Intention:
[Harry] wants to retrieve one of Voldemort's Horcruxes in exchange for Godric Gryffindor's Sword.

Consequence:
not exist
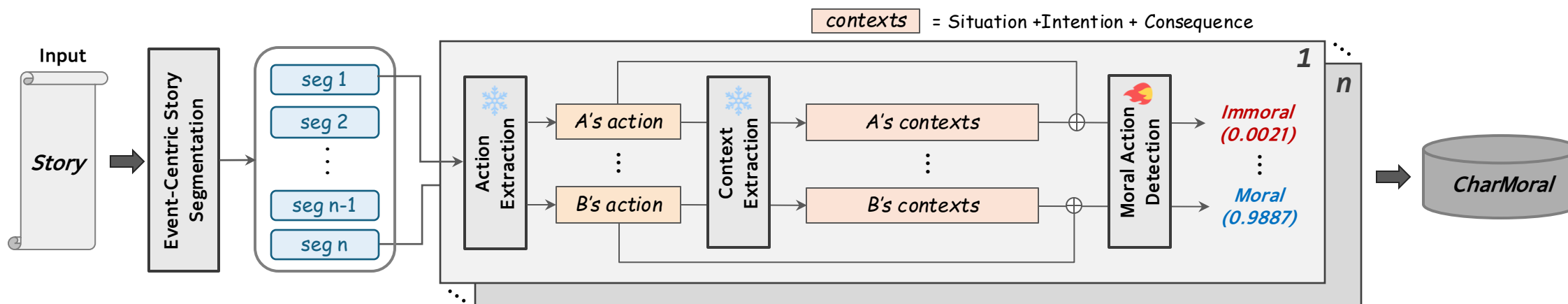
Label:
Moral

2. Method : CharMoral dataset (Cont.)

- Statistics of CharMoral dataset

After Segmentation	
#Story	1,337
#Segment	30,616
Story Mean length	1,665
Segment Mean length	85
After Context Extraction	
#Character	9,389
#Annotations	103,836
#Action exists	103,836 (100%)
#Situation exists	103,813 (99.98%)
#Intention exists	92,627 (89.2%)
#Consequence exists	82,076 (79.04%)
#All exists	75,724 (79.93%)
Label Distribution	
#Moral	50,717
#Immoral	53,119

Table 2: The statistics of *CharMoral*
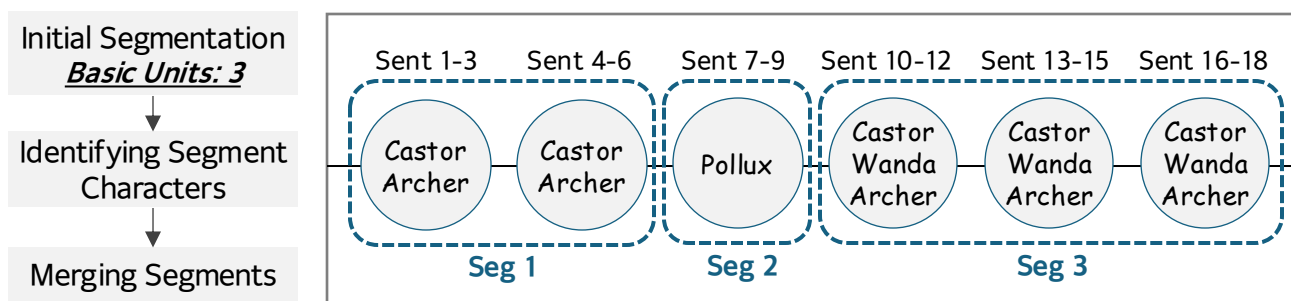
2. Method : Dataset construction

contexts = Situation + Intention + Consequence

2. Method : Dataset construction (Cont.)

contexts = Situation + Intention + Consequence

- Input
- Story
- Event-Centric Story Segmentation
- seg 1, seg 2, ..., seg n-1, seg n
- Action Extraction
- A's action, B's action
- Context Extraction
- A's contexts, B's contexts
- Moral Action Detection
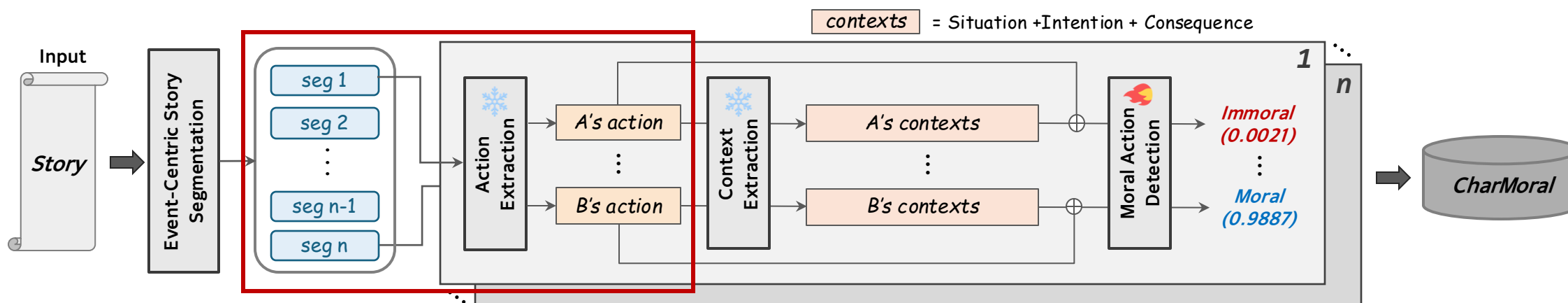- Immoral (0.0021), Moral (0.9887)
- CharMoral
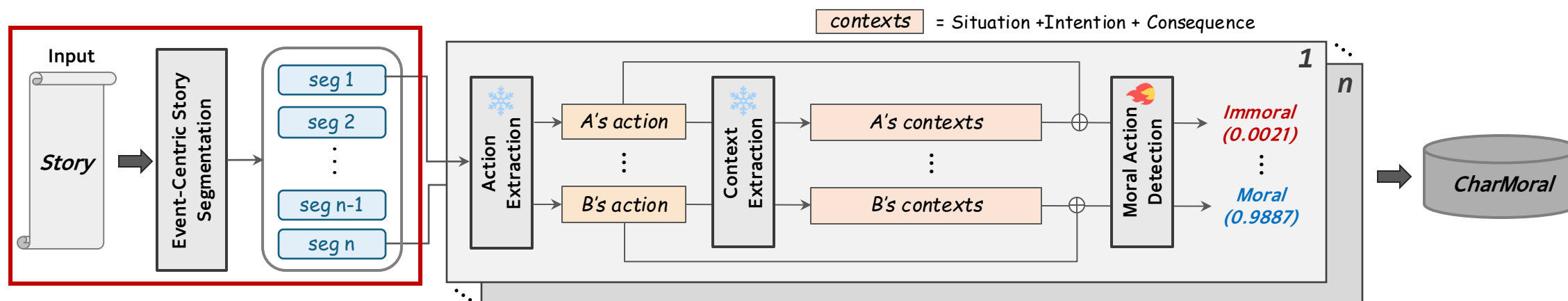
(A) Event-Centric Story Segmentation

- To divide the long story into key events.
- This method draws upon our earlier work: *A Two-Stage Summarization Model using Scene Attributes (Kim et al., CreativeSumm 2022)*

Initial Segmentation *Basic Units: 3* → Identifying Segment Characters → Merging Segments

Sent 1-3 | Sent 4-6 | Sent 7-9 | Sent 10-12 | Sent 13-15 | Sent 16-18

Castor Archer | Castor Archer | Pollux | Castor Wanda Archer | Castor Wanda Archer | Castor Wanda Archer

Seg 1 | Seg 2 | Seg 3

2. Method : Dataset construction (Cont.)

contexts = Situation + Intention + Consequence

(B) Action Extraction

- Extract characters' action sentences in each segment using GPT-4o

Action extraction prompt

instruction
In the "text", the character names are enclosed in "[" and "]".
From the text, the list of characters is *{character list}*.
Extract all the actions of each character name.
If there is no action, print "no action".
…
Input:
text: *{segment}*

<start>
[Archer]: "[Archer] is taking his six-year-old son …",
[Archer]: "[Archer] runs his hand over [Michael]'s face and smiles.",
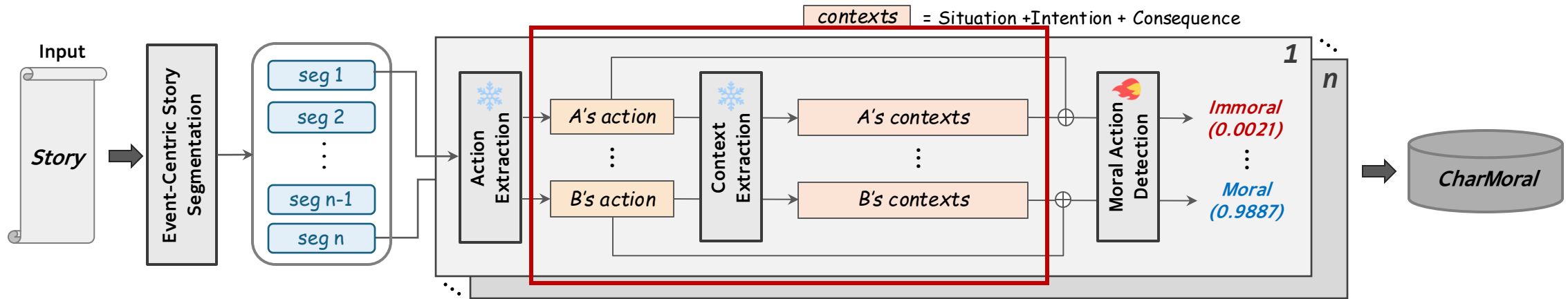**[Michael]:** "no action",
[Castor]: "[Castor] sets up a suppressed sniper rifle on the hill overlooking the carousel.",

2. Method : Dataset construction (Cont.)

contexts = Situation + Intention + Consequence

(C) Context Extraction

- Extract action's context information to assess morality accurately using GPT-4o

Action's context extraction prompt

instruction
Your work is to extract three sentences in the "Segment", corresponding Situation, Intention, and Consequence related to the *{character name}*'s given "Action".
…
Input:
Segment: *{segment}*
Action: *{action}*

<start>
[Situation]: "In September 1991, at the Griffith …" ,
[Intention]: "To carry out a sniper attack, possibly targeting [Archer]." ,
[Consequence]: "The sniper shot intended for [Archer] …" ,
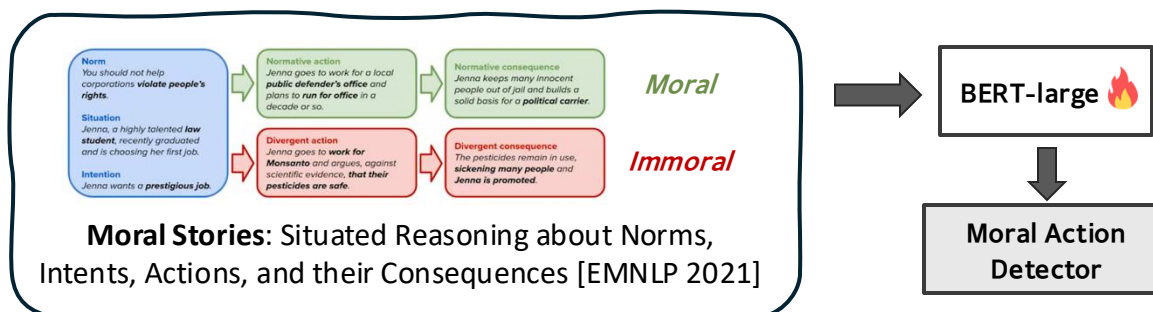
2. Method : Dataset construction (Cont.)
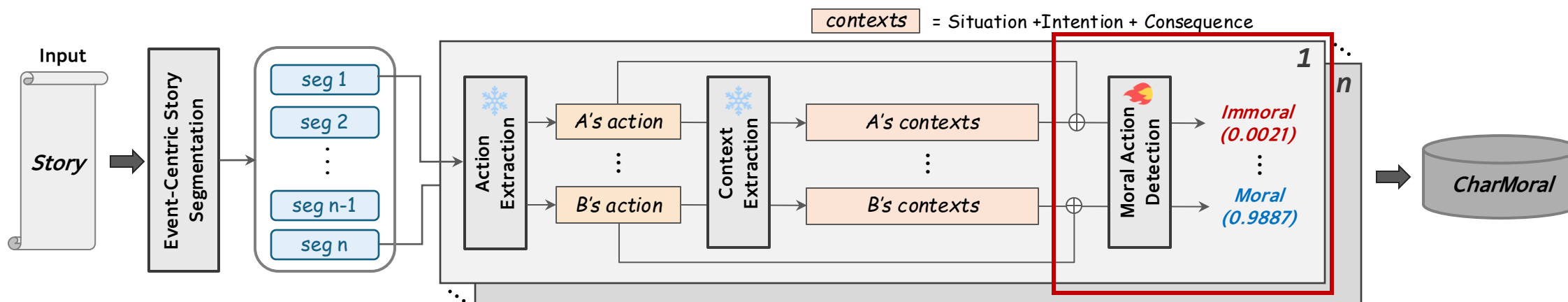
contexts = Situation + Intention + Consequence

(D) Action morality Extraction

- Employ an expert model (= Moral Action Detector) to accurately annotate whether the actions of the characters are morally justified based on their contextual information.
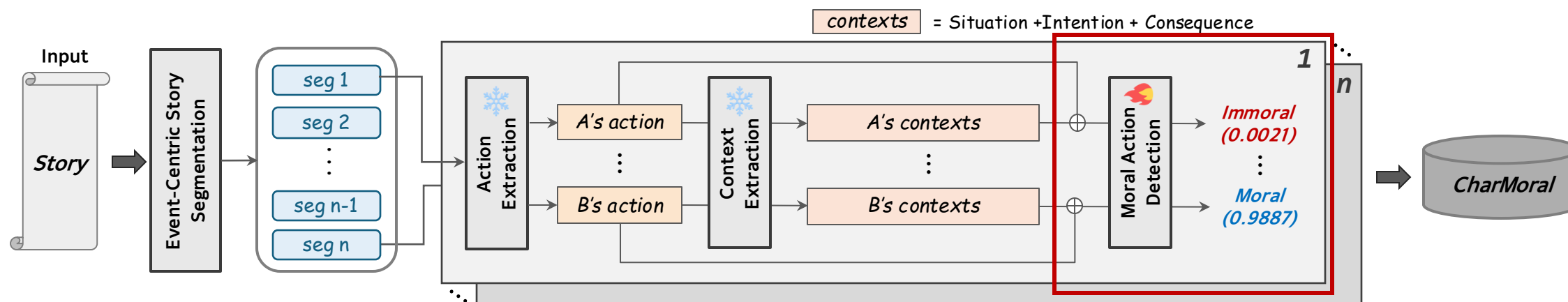
1. Training

Moral Stories: Situated Reasoning about Norms, Intents, Actions, and their Consequences [EMNLP 2021]

BERT-large → Moral Action Detector

Metric	Acc	F1
Score	0.977	0.977

Test accuracy of fine-tuning MAD

contexts = Situation + Intention + Consequence
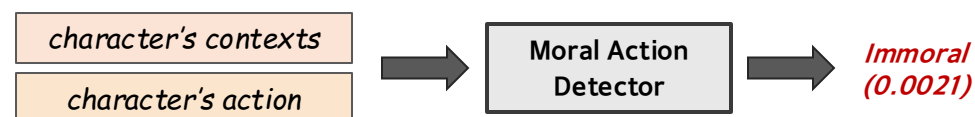
(D) Action morality Extraction

- Employ an expert model (= Moral Action Detector) to accurately annotate whether the actions of the characters are morally justified based on their contextual information.

2. Inference

3. Experiment & Analysis : Human evaluation

- Effectiveness of our framework by comparing its predictions with human assessments
  - The results of human assessments by comparing the moral action classification performance of MAD and LLMs.

Model	MAD	GPT 3.5	GPT 4
Acc	**0.92**	0.78	0.86

3. Experiment & Analysis : Moral dynamic score

- Moral Dynamic Score

$$score_c = \frac{(\#pass_c)}{(\#segment_c) - 1}$$

The number of times a character's morality logit score sequences crosses the 0.5 threshold.

The number of segments in which a character appears

- Moral Dynamic Character

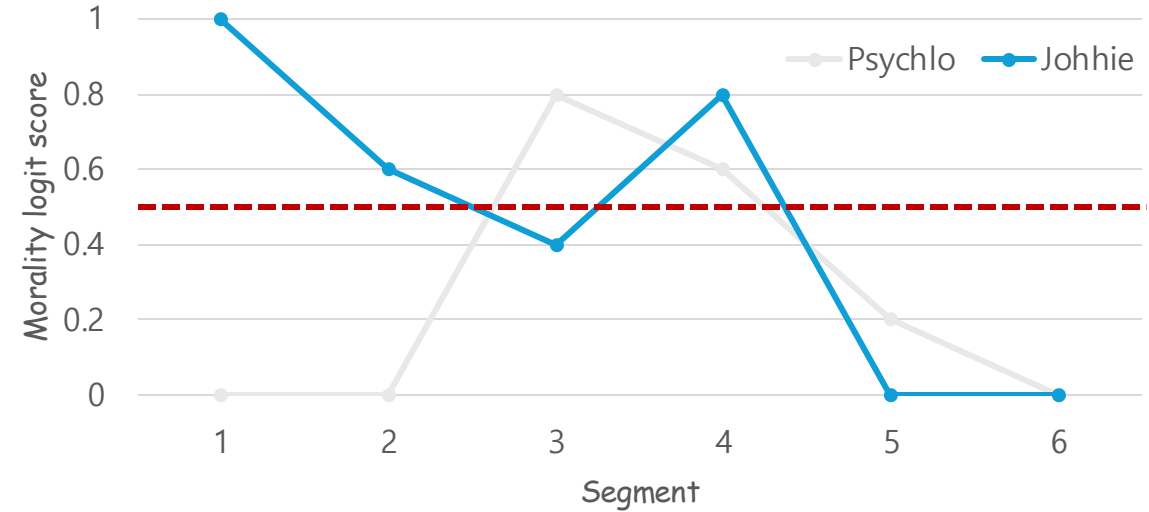$$dynamics_c = \begin{cases} dynamic & \text{if } score_c \geq 0.5, \\ static & \text{otherwise} \end{cases}$$

- Moral Dynamic Score

$$score_c = \frac{(\#pass_c)}{(\#segment_c) - 1}$$

- Moral Dynamic Character

$$dynamics_c = \begin{cases} dynamic & \text{if } score_c \geq 0.5, \\ static & \text{otherwise} \end{cases}$$
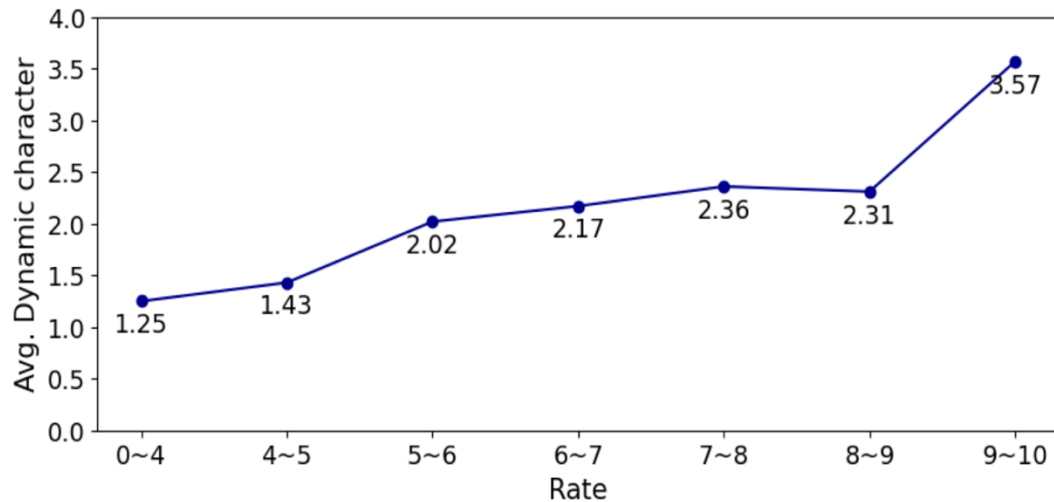

$\#pass_{Johhie} = 3 \qquad \#segment_{Johhie} = 6$
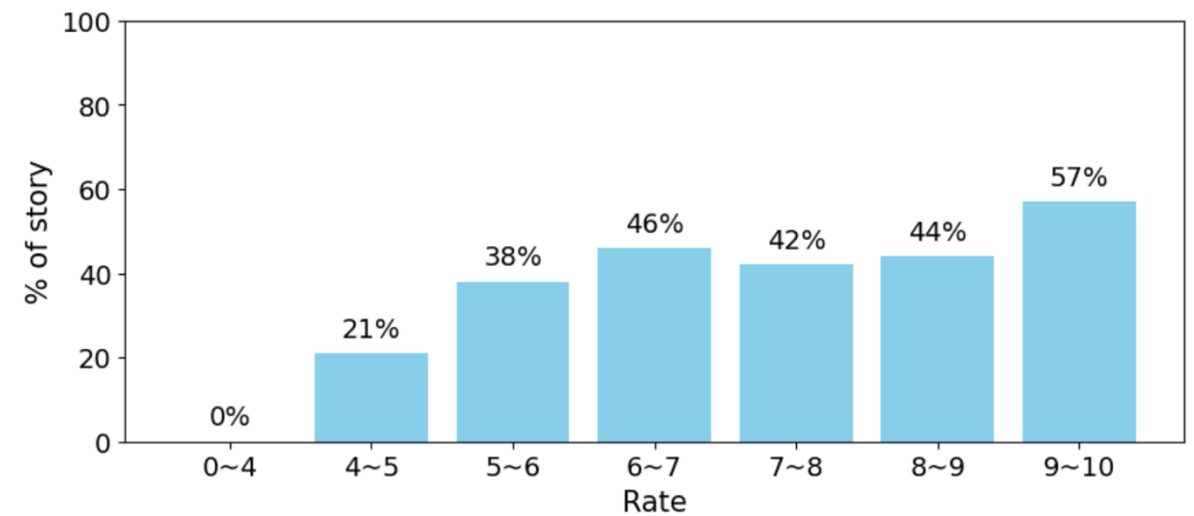
$$\#score_{Johhie} = \frac{3}{6-1} = \frac{3}{5} = 0.6$$

Dynamic Character

3. Experiment & Analysis : Morally dynamic character analysis

- **A1:** A greater number of morally dynamic characters positively contributes to the story's overall interest

- **A2:** The moral evolution of the main character plays a key role in increasing the story's overall interest.

3. Experiment & Analysis : Effectiveness in moral reasoning tasks

1. Performance of Fine-tuning moral action classification task

Domain	In-domain		Cross-domain			
Train -> Test	Ours -> Ours		Ours -> Moral Stories		Ours -> Social Chemistry	
Metric	*Acc*	*F1*	*Acc*	*F1*	*Acc*	*F1*
Action	0.692	0.712	0.689	0.716	**0.823**	**0.778**
Action w/ Context	**0.947**	**0.947**	**0.969**	**0.969**	0.707	0.655

2. Performance of Zero-shot and Few-shot context-aware moral action classification task

Model	GPT-3.5		GPT-4o	
Metric	*Acc*	*F1*	*Acc*	*F1*
zero-shot	0.617	0.685	0.667	0.680
1-shot	0.619	0.688	0.667	0.690
3-shot	**0.627**	0.703	**0.684**	**0.693**
5-shot	0.622	**0.708**	0.683	0.692

4. Conclusion

- We introduce **CharMoral, a novel dataset** designed to analyze the moral dynamics of characters in long-form narratives.

- We propose a **four-stage framework** that leverages LLMs

- We introduce a novel score, **moral dynamic score** of a character to analyze whether the morally dynamic characters play a key role in increasing story engagement.

Thank you for listening!

Contact: sybae01@skku.edu

📄 Paper   Code & Dataset