

August 30, 2024 Grand Hyatt Hotel, Athens, Greece

# MANPU 2024

The 6th International IAPR Workshop on  
coMics ANalysis, Processing and Understanding

La Rochelle  
Université

# Toward accessible comics for blind and low vision readers

Christophe Rigaud, Jean-Christophe Burie, & Samuel Petit










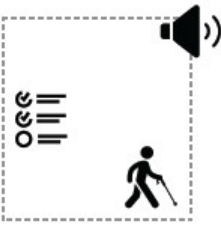
# Summary

1. Impairments
2. Challenges
3. State of the art
4. Proposition
  - 4.1 Text classification
  - 4.2 Character clustering
5. Conclusions

SAY, ER... LOOK HERE,  
I'D LIKE TO GO ALONG  
WITH YOU ON THIS  
AMAZING MANPU



# 1. Impairments

Visual impairment					Blindness		
Moderate	Blurry	Peripheral loss	Central loss	Spots	Partial	Near-total	Total
							

# 1. Impairments

## Motor disorders

Eye control

Paralysis

Tremors

## Cognitive disorders

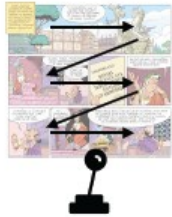
Dyslexia

Dyspraxia

Dysphasia

Attention disorders

Memory loss



## 2. Challenges

How to make comics accessible?



## 2. Challenges

### How to make comics accessible?



USER:

You are a storyteller for the blind. Please describe this comic book panel by naming characters.

ASSISTANT:

In this comic book panel, there are three characters. On the left, there is a character with blonde hair, wearing an orange sweater and blue pants, sitting on a bench and playing a guitar. Beside him is another character with dark hair, wearing a red shirt and black shorts, holding a tambourine, and sitting on the same bench. She is saying, "Look! There's Leonardo!" In the background, there is a character with blonde hair, wearing a purple top and blue pants, running away. The setting is in front of a café with a large window, and there is a sign of a musical note above the entrance. There is also a blue car parked on the street. A text box at the bottom right corner reads, "Leonardo has done nothing wrong, but what is bothering him? Who is he running from?"

## 2. Challenges

### How to make comics accessible?



USER:

You are a storyteller for the blind. Please describe this comic book panel by naming characters.

ASSISTANT:

In this comic book panel, there are **three characters**. On the left, there is a character with **blonde hair**, wearing an **orange sweater** and **blue pants**, sitting on a bench and playing a guitar. Beside him is another character with **dark hair**, wearing a **red shirt** and **black shorts**, holding a tambourine, and sitting on the same bench. She is saying, "Look! There's Leonardo!" In the background, there is a character with **blonde hair**, wearing a **purple top** and **blue pants**, running away. The setting is in front of a café with a **large window**, and there is a sign of a musical note above the entrance. There is also a **blue car** parked on the street. A text box at the bottom right corner reads, "Leonardo has done nothing wrong, but what is bothering him? Who is he running from?"



## 2. Challenges

### How to make comics accessible?



USER:

You are a storyteller for the blind. Please describe this comic book panel by naming characters.

ASSISTANT:

In this comic book panel, there are three characters. On the left, there is a character with blonde hair, wearing an orange sweater and blue pants, sitting on a bench and playing a guitar. Beside him is another character with dark hair, wearing a red shirt and black shorts, holding a tambourine, and sitting on the same bench. She is saying, **"Look! There's Leonardo!"** In the background, there is a character with blonde hair, wearing a purple top and blue pants, running away. The setting is in front of a café with a large window, and there is a sign of a musical note above the entrance. There is also a blue car parked on the street. A text box at the bottom right corner reads, **"Leonardo has done nothing wrong, but what is bothering him? Who is he running from?"**



## 2. Challenges

### How to make comics accessible?



USER:

You are a storyteller for the blind. Please describe this comic book panel by naming characters.

ASSISTANT:

In this comic book panel, there are three characters. On the left, there is a character with blonde hair, wearing an orange sweater and blue pants, sitting on a bench and playing a guitar. Beside him is another character with dark hair, wearing a red shirt and black shorts, holding a tambourine, and sitting on the same bench. She is saying, **"Look! There's Leonardo!"** In the background, there is a character with blonde hair, wearing a purple top and blue pants, running away. The setting is in front of a café with a large window, and there is a sign of a musical note above the entrance. There is also a blue car parked on the street. A text box at the bottom right corner reads, **"Leonardo has done nothing wrong, but what is bothering him? Who is he running from?"**

*a speech bubble is missing...*

## 2. Challenges

### How to make comics accessible?



USER:

You are a storyteller for the blind. Please describe this comic book panel by naming characters.

ASSISTANT:

In this comic book panel, there are three characters. **On the left**, there is a character with blonde hair, wearing an orange sweater and blue pants, **sitting on a bench** and **playing a guitar**. Beside him is another character with dark hair, wearing a red shirt and black shorts, holding a tambourine, and **sitting on the same bench**. She is saying, "Look! There's Leonardo!" In the **background**, there is a character with blonde hair, wearing a purple top and blue pants, **running away**. The setting is **in front of a café** with a **large window**, and there is a sign of a musical note above the entrance. There is also a blue car parked **on the street**. A text box at the bottom right corner reads, "Leonardo has done nothing wrong, but what is bothering him? Who is he running from?"

## 2. Challenges

### How to make comics accessible?



USER:

You are a storyteller for the blind. Please describe this comic book panel by naming characters.

ASSISTANT:

In this comic book panel, there are three characters. **On the left**, there is a character with blonde hair, wearing an orange sweater and blue pants, **sitting on a bench** and **playing a guitar**. Beside him is another character with dark hair, wearing a red shirt and black shorts, holding a tambourine, and **sitting on the same bench**. She is saying, "Look! There's Leonardo!" In the **background**, there is a character with blonde hair, wearing a purple top and blue pants, **running away**. The setting is **in front of a café** with a **large window**, and there is a sign of a musical note above the entrance. There is also a blue car parked **on the street**. A text box at the bottom right corner reads, "Leonardo has done nothing wrong, but what is bothering him? Who is he running from?"

*character's names are missing...*

### 3. State of the art

2023 (Oct.)

[1] **Comics for Everyone: Generating Accessible Text Descriptions for Comic Strips.** Ramaprasad, R., Microsoft Research, arXiv preprint

- Single strip panel description on famous comics (Dilbert, Garfield and Peanuts)
- Identify the key elements of comic strips that are essential for generating good descriptions with LLM
- Some hallucination effects observed regarding extra details
- Training-free, prompt-based approach



The comic strip consists of three panels, each featuring a different scene and dialogue. In the first panel, we see two characters, Wally and the Boss, standing in an office setting. Wally is holding a cup of coffee and expressing his frustration with the performance review system. He says, "PERFORMANCE REVIEWS ARE UNFAIR TO UNDERPERFORMING EMPLOYEES SUCH AS MYSELF." In the second panel, Wally is shown alone, still holding his cup of Coffee...

### 3. State of the art

2023 (Oct.)

[1] **Comics for Everyone: Generating Accessible Text Descriptions for Comic Strips.** Ramaprasad, R., Microsoft Research, arXiv preprint

- Single strip panel description on famous comics (Dilbert, Garfield and Peanuts)
- Identify the key elements of comic strips that are essential for generating good descriptions with LLM
- Some hallucination effects observed regarding extra details
- Training-free, prompt-based approach

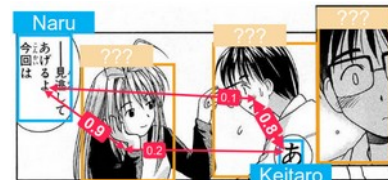


The comic strip consists of three panels, each featuring a different scene and dialogue. In the first panel, we see two characters, Wally and the Boss, standing in an office setting. Wally is holding a cup of coffee and expressing his frustration with the performance review system. He says, "PERFORMANCE REVIEWS ARE UNFAIR TO UNDERPERFORMING EMPLOYEES SUCH AS MYSELF." In the second panel, Wally is shown alone, still holding his cup of Coffee...

2024 (Apr.)

[2] **Zero-shot character identification and speaker prediction in comics via iterative multimodal fusion.** Li, Y., Hinami, R., Aizawa, K., Matsui, Y.: arXiv preprint

- OCR + LLM text analysis to obtain the list of character names
- Character name list is used as target labels for bubble speaker prediction and character classification
- Predict bubble speaker name from text only (LLM) and then combined with character classification output
- Character classification trained on 349 characters of Manga109



### 3. State of the art

2023 (Oct.)

[1] **Comics for Everyone: Generating Accessible Text Descriptions for Comic Strips.** Ramaprasad, R., Microsoft Research, arXiv preprint

- Single strip panel description on famous comics (Dilbert, Garfield and Peanuts)
- Identify the key elements of comic strips that are essential for generating good descriptions with LLM
- Some hallucination effects observed regarding extra details
- Training-free, prompt-based approach

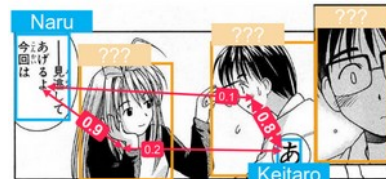


The comic strip consists of three panels, each featuring a different scene and dialogue. In the first panel, we see two characters, Wally and the Boss, standing in an office setting. Wally is holding a cup of coffee and expressing his frustration with the performance review system. He says, "PERFORMANCE REVIEWS ARE UNFAIR TO UNDERPERFORMING EMPLOYEES SUCH AS MYSELF." In the second panel, Wally is shown alone, still holding his cup of Coffee...

2024 (Apr.)

[2] **Zero-shot character identification and speaker prediction in comics via iterative multimodal fusion.** Li, Y., Hinami, R., Aizawa, K., Matsui, Y.: arXiv preprint

- OCR + LLM text analysis to obtain the list of character names
- Character name list is used as target labels for bubble speaker prediction and character classification
- Predict bubble speaker name from text only (LLM) and then combined with character classification output
- Character classification trained on 349 characters of Manga109



2024 (Jun.)

[3] **The Manga Whisperer: Automatically Generating Transcriptions for Comics.** Sachdeva, R., Zisserman, A.: CVPR'24

- Tackle the problem of speaker diarisation: who says what and when (layout + text transcript)
- Content extraction ResNet50 + 3-layered MLPs for character matching and speaker association
- End-to-end training (2x A40 GPUs) using Manga109, PopManga, Mangadex-1.5M





### 3. State of the art

2023 (Oct.)

[1] **Comics for Everyone: Generating Accessible Text Descriptions for Comic Strips.** Ramaprasad, R., Microsoft Research, arXiv preprint

- Single strip panel description on **famous** comics (Dilbert, Garfield and Peanuts)
- Identify the key elements of comic strips that are essential for generating good descriptions with LLM
- Some hallucination effects observed regarding extra details
- **Training-free, prompt-based approach**

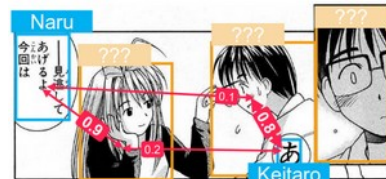


The comic strip consists of three panels, each featuring a different scene and dialogue. In the first panel, we see two characters, Wally and the Boss, standing in an office setting. Wally is holding a cup of coffee and expressing his frustration with the performance review system. He says, "PERFORMANCE REVIEWS ARE UNFAIR TO UNDERPERFORMING EMPLOYEES SUCH AS MYSELF." In the second panel, Wally is shown alone, still holding his cup of coffee...

2024 (Apr.)

[2] **Zero-shot character identification and speaker prediction in comics via iterative multimodal fusion.** Li, Y., Hinami, R., Aizawa, K., Matsui, Y.: arXiv preprint

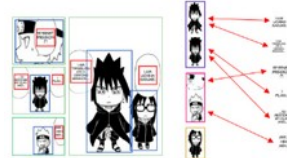
- **OCR + LLM text analysis to obtained the list of character names**
- Character name list is used as target labels for bubble speaker prediction and character classification
- Predict bubble speaker name from text only (LLM) and then combined with character classification output
- **Character classification trained** on 349 characters of Manga109



2024 (Jun.)

[3] **The Manga Whisperer: Automatically Generating Transcriptions for Comics.** Sachdeva, R., Zisserman, A.: CVPR'24

- **Tackle the problem of speaker diarisation: who says what and when (layout + text transcript)**
- Content extraction ResNet50 + 3-layered MLPs for character matching and speaker association
- **End-to-end training (2x A40 GPUs)** using Manga109, PopManga, Mangadex-1.5M



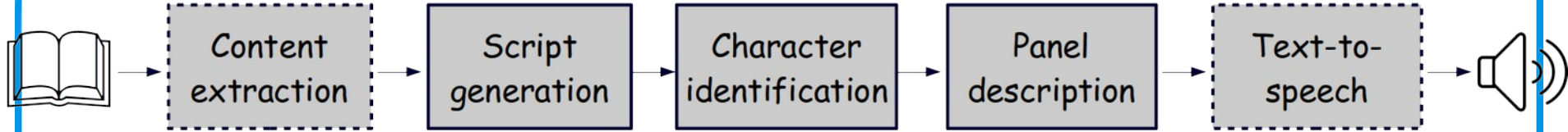
5

- 1x: As many people want a hunters license as there are stars in the sky.
- 1x: You think there're enough examiners with enough time to judge 'em all? Not hardly!
- 1x: That's why there are people like me, who not only "drive the bus," but also screen the applicants.
- 1x: The other passengers are all drop-outs, and have been reported as such to the committee.
- 1x: If any of them reach the site of the exam by different routes, they'll be turned away.
- 1x: Whether you three qualify is my call, too. Best you keep that in mind.
- 1x: Now, about my question...
- 2x: I'm...
- 2x: ...the last of the Kurta clan.

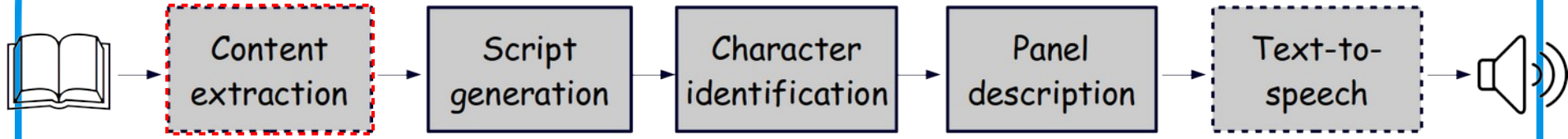
+ panel description on unknown comics AND complete script (layout + text + scene description)



## 4. Proposition



## 4. Proposition

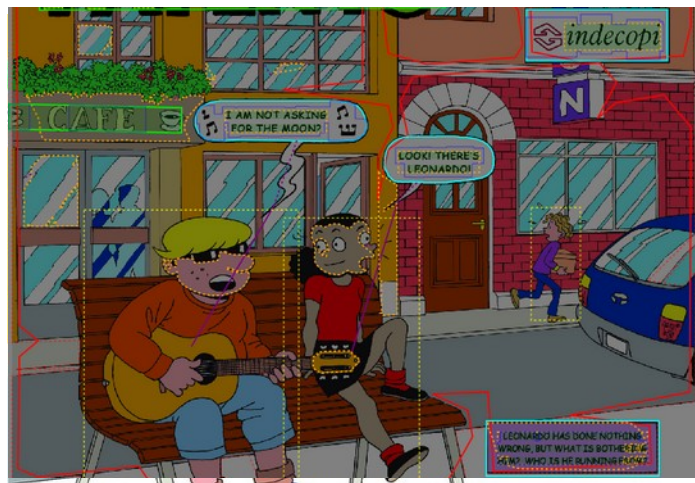
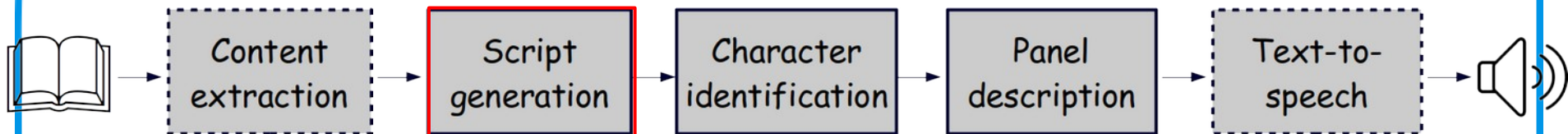


[4] **Image segmentation, classification and recognition methods for comics: A decade systematic literature review.** Sharma, R., Kukreja, V.: Engineering Applications of Artificial Intelligence 131, 107715 (2024)

[5] **Computational approaches to comics analysis.** Laubrock, J., Dunst, A.: Topics in cognitive science 12(1), 274–310 (2020)

[6] **A survey of comics research in computer science.** Augereau, O., Iwata, M., Kise, K.: Journal of Imaging 4(87) (2018)

## 4. Proposition



### PAGE 1 - 3 PANELS

#### PANEL 1

#### SOUND (and other)

3 CAFE &  
indecopi

#### DIALOGUE

**α:** I AM NOT ASKING FOR THE MOON?

**β:** LOOK! THERE'S LEONARDO!

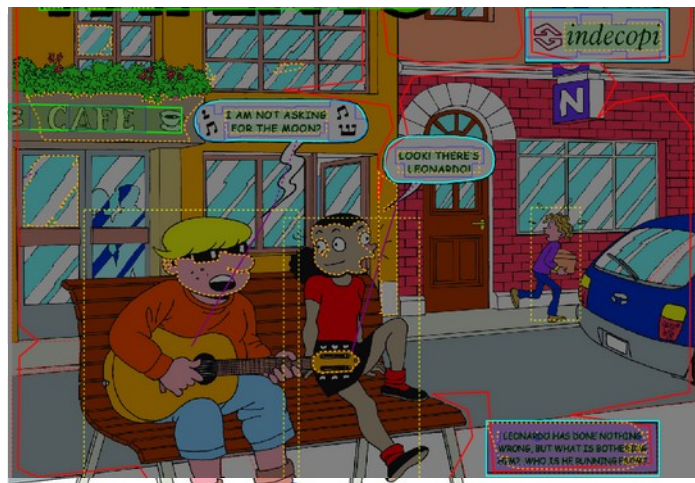
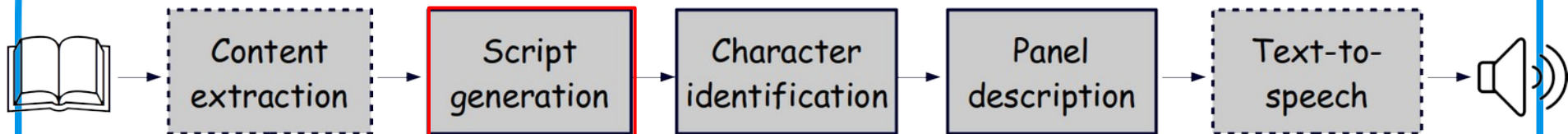
**γ:**

#### CAPTION

LEONARDO HAS DONE NOTHING  
WRONG, BUT WHAT IS BOTHERING HIM?  
WHO IS HE RUNNING FROM?

...

## 4. Proposition



Text  
classification

PAGE 1 - 3 PANELS

PANEL 1

**SOUND** (and other)

3 CAFE &  
indecopi

**DIALOGUE**

**α:** I AM NOT ASKING FOR THE MOON?

**β:** LOOK! THERE'S LEONARDO!

**γ:**

**CAPTION**

LEONARDO HAS DONE NOTHING  
WRONG, BUT WHAT IS BOTHERING HIM?  
WHO IS HE RUNNING FROM?

...

## 4. Proposition

### 4.1 Text classification

#### Sound effects (SFX)

Height > 2.5% x image\_width

OR

Min. slope > 10% vs horizontal



#### Dialogue

Balloon with tail (if any...)

① NOVELTY MEANS  
THAT THE INVENTION  
MUST NOT BE KNOWN.

#### Caption

Everything else.

#### Illustration

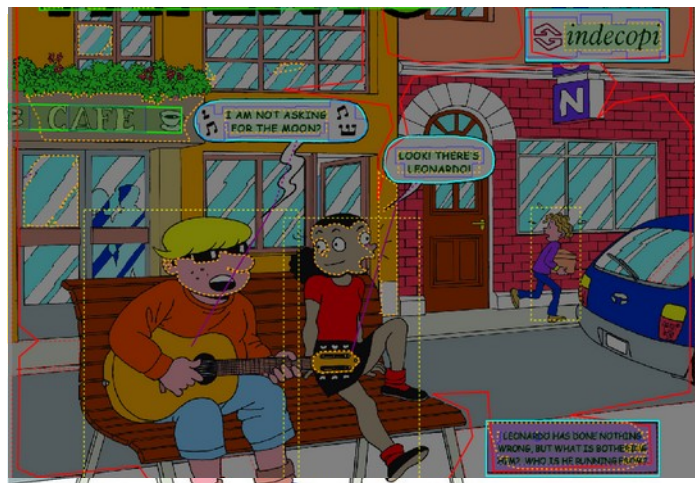
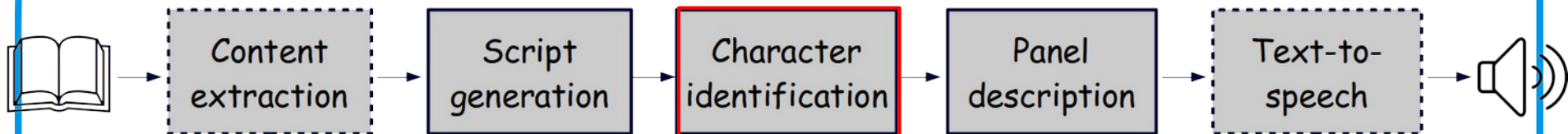
?



- ① Die Daten sind nicht bekannt
- ② Die Daten sind nicht bekannt
- ③ Die Daten sind nicht bekannt
- ④ Die Daten sind nicht bekannt



## 4. Proposition



Character  
clustering+**nam**ing

PAGE 1 - 3 PANELS

PANEL 1

SOUND (and other)

3 CAFE &  
indecopi

DIALOGUE

**α**: I AM NOT ASKING FOR THE MOON?

**β**: LOOK! THERE'S LEONARDO!

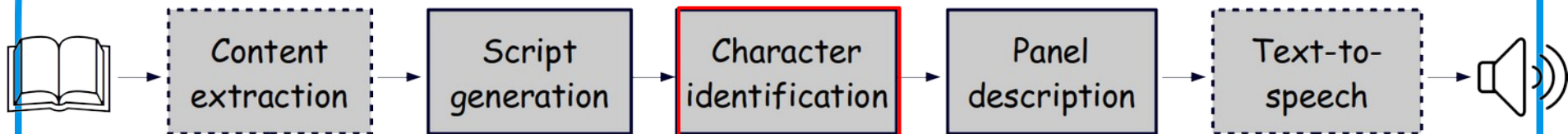
**γ**:

CAPTION

LEONARDO HAS DONE NOTHING  
WRONG, BUT WHAT IS BOTHERING HIM?  
WHO IS HE RUNNING FROM?

...

## 4. Proposition



Character  
clustering+ naming

PAGE 1 - 3 PANELS

PANEL 1

SOUND (and other)

3 CAFE &  
indecopi

DIALOGUE

**Marco:** I AM NOT ASKING FOR THE MOON?

**Babi:** LOOK! THERE'S LEONARDO!

**Leonardo:**

CAPTION

LEONARDO HAS DONE NOTHING WRONG,  
BUT WHAT IS BOTHERING HIM? WHO IS HE  
RUNNING FROM?

...



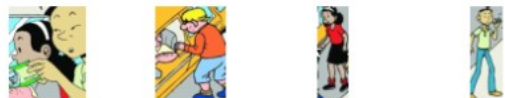
# 4. Proposition

## 4.2 character clustering



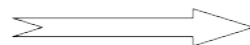
### 1. Image feature vector

From CLIP (ViT version)



### 2. Vector size reduction

UMAP (768  $\rightarrow$  8 features) [7]



### 3. Vector clustering

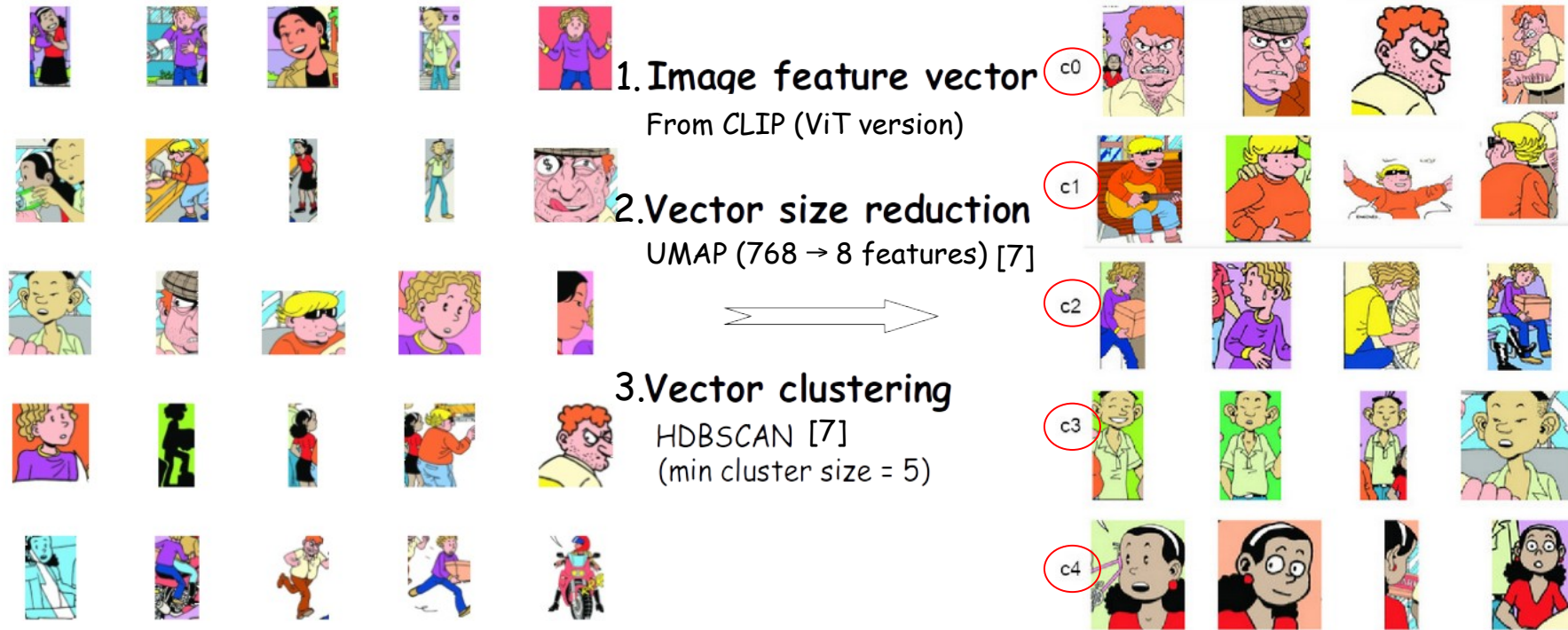
HDBSCAN [7]  
(min cluster size = 5)



[7] **Automatic classification of manga characters using density-based clustering.** Yanagisawa, Hideaki, et al., International Workshop on Advanced Imaging Technology (IWAIT) 2020. Vol. 11515. SPIE, 2020.

# 4. Proposition

## 4.2 character clustering



[7] **Automatic classification of manga characters using density-based clustering.** Yanagisawa, Hideaki, et al., International Workshop on Advanced Imaging Technology (IWAIT) 2020. Vol. 11515. SPIE, 2020.

# 4. Proposition

## 4.2 character name inference

### PAGE 1 - 3 PANELS

#### PANEL 1

#### SOUND (and other)

3 CAFE &

indecopi

#### DIALOGUE

**c1:** I AM NOT ASKING FOR THE MOON?

**c4:** LOOK! THERE'S LEONARDO!

**c2:**

#### CAPTION

LEONARDO HAS DONE NOTHING  
WRONG, BUT WHAT IS BOTHERING HIM?  
WHO IS HE RUNNING FROM?

...

USER: This is the script of a comic book:  
[SCRIPT]. Please list all character's names  
with corresponding unique identifiers.  
Itemize and explain your reasoning.

ASSISTANT llama-3-70b-instruct:

After analyzing the script, I've identified  
the following comic characters and their  
corresponding identifiers:

1. **\*\*Leonardo\*\*** (c2)

\* Reasoning: Leonardo is mentioned as  
the inventor...

2. **\*\*Marco\*\*** (c1)

\* Reasoning: Marco is introduced as a  
friend of Leonardo's

3. **\*\*Babi\*\*** (c4)

\* Reasoning: Babi is...

4. **\*\*Ping Song\*\*** (c3)

\* Reasoning: Ping Song is introduced  
as a friend of

5. **\*\*Chuli\*\*** (c6)

\* Reasoning: Chuli is introduced as a  
friend who can help...

# 4. Proposition

## 4.2 character name inference

### PAGE 1 - 3 PANELS

#### PANEL 1

#### SOUND (and other)

3 CAFE &  
indecopi

#### DIALOGUE

**(c1)**: I AM NOT ASKING FOR THE MOON?

**(c4)**: LOOK! THERE'S LEONARDO!

**(c2)**:

#### CAPTION

LEONARDO HAS DONE NOTHING  
WRONG, BUT WHAT IS BOTHERING HIM?  
WHO IS HE RUNNING FROM?

...

USER: This is the script of a comic book:  
[SCRIPT]. Please list all character's names  
with corresponding unique identifiers.  
Itemize and explain your reasoning.

ASSISTANT llama-3-70b-instruct:

After analyzing the script, I've identified  
the following comic characters and their  
corresponding identifiers:

1. **\*\*Leonardo\*\*** ← **(c2)**

\* Reasoning: Leonardo is mentioned as  
the inventor...

2. **\*\*Marco\*\*** ← **(c1)**

\* Reasoning: Marco is introduced as a  
friend of Leonardo's

3. **\*\*Babi\*\*** ← **(c4)**

\* Reasoning: Babi is...

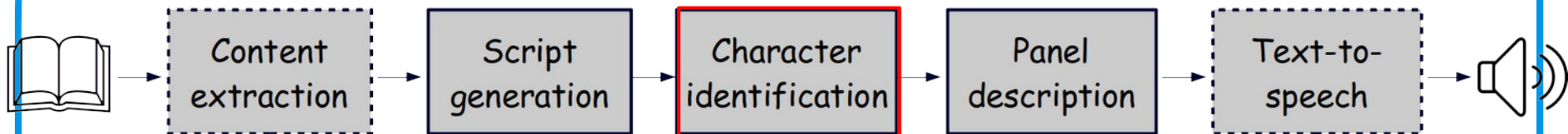
4. **\*\*Ping Song\*\*** (c3)

\* Reasoning: Ping Song is introduced  
as a friend of

5. **\*\*Chuli\*\*** (c6)

\* Reasoning: Chuli is introduced as a  
friend who can help...

## 4. Proposition



Character naming

PAGE 1 - 3 PANELS

PANEL 1

SOUND (and other)

3 CAFE &  
indecopi

DIALOGUE

**Marco:** I AM NOT ASKING FOR THE  
MOON?

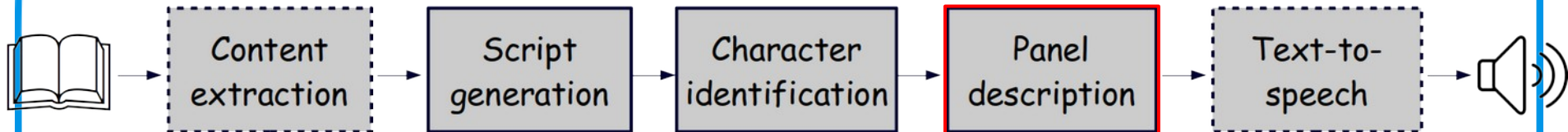
**Babi:** LOOK! THERE'S LEONARDO!

**Leonardo:**

CAPTION

LEONARDO HAS DONE NOTHING  
WRONG, BUT WHAT IS BOTHERING HIM?  
WHO IS HE RUNNING FROM?

## 4. Proposition



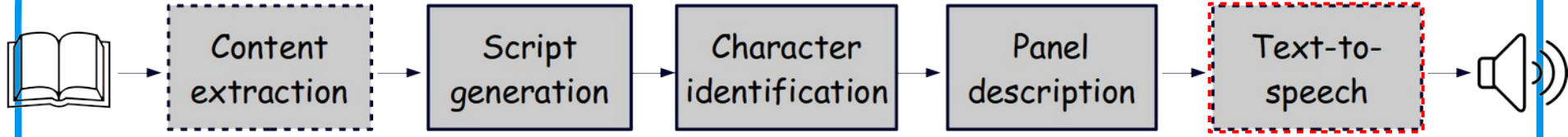
USER: ... [SCRIPT].

ASSISTANT GPT-4o:

**Marco**, a young person with blond hair, is sitting on a bench playing a guitar and **singing**, "I am not asking for the moon?" Next to him on the bench is **Babi**, who is wearing a red shirt and black skirt. She looks **surprised** and is **pointing towards a boy named Leonardo**, who is **running away** with a worried expression on his face. **Leonardo** is depicted running past a **blue car** and a **door**, while **Babi exclaims**, "Look! There's Leonardo!" The **caption** at the bottom of the panel adds to the **mystery**, **stating**: "Leonardo has done nothing wrong, but what is bothering him? Who is he running from?"



## 4. Proposition



||ElevenLabs



# 5. Conclusions

## Contributions

- Text classification baseline
- Automatic character's names inference
- Automatic script generation
- Context-aware panel description
- Training-free, prompt-based approach!

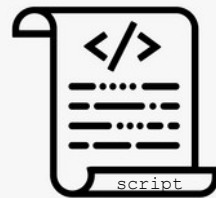
## Perspectives

- Extend text classes (electronic, screens, graffiti, document, signs, shop...)
- Add panel-to-panel transition descriptions (action-to-action, time-passage, zoom...)
- Evaluation quantitative [9] and qualitative (text-to-image?)
- Validation with specialists of visual impairment



"Patents Comic Book" WIPO publication

image-to-text



TTS

text-to-image



Generated by FLUX.1 [schnell]

La Rochelle  
Université



PDF preprint

<https://arxiv.org/abs/2407.08248>



Sequential Art Image Laboratory  
<https://sail.univ-lr.fr>

### Contacts:

christophe.rigaud  
jean-christophe.burie  
@univ-lr.fr  
samuel.petit  
@demarque.com



Generated by FLUX.1 [schnell]



Data

<https://gitlab.univ-lr.fr/crigau02/publication/-/tree/main/2024/MANPU/>



Avec les aveugles et les malvoyants

