Applied Data Science Project

"Generation of video reviews based on textual descriptions"

Checkpoint III:

- Objective(s)
- Research question(s)
- Methods
- Experiments
- Conclusions

FLOWYGO

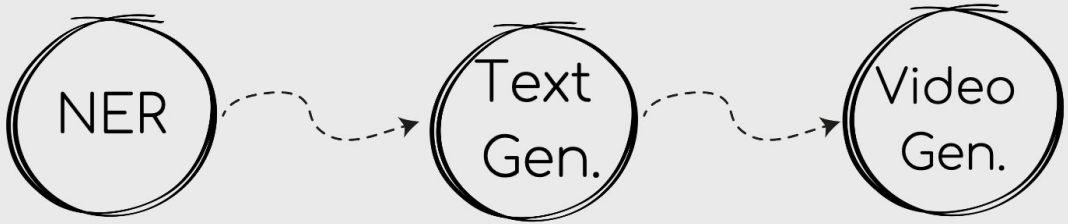
FLOWYGO

SUSTAINABLE DEVELOPMENT GOALS

12 RESPONSIBLE CONSUMPTION AND PRODUCTION

Project
(objectives)

- Goal: a system able to generate text and video reviews from a list of features of a given product
- The objective is to increase the chance of a better match between customer and products
- The system is divided in three main components:



FLOWYGO

SUSTAINABLE DEVELOPMENT GOALS

12 RESPONSIBLE CONSUMPTION AND PRODUCTION

Project
(research questions)

Main criticalities/research questions:

- How can we build a more advanced system for video generation to show features as soon as they are mentioned?
- How can we implement a module that recognizes when an attribute is being used?
- GPT-2 small vs medium: how do we choose the appropriate model considering capacity and cost?

We will further show:

- Description of how NER and GPT-2 are applied
- Description of how the video generator works
- Evaluation metrics for NER and GPT-2 models

Methods

NER

Named Entity Recognizer

spaCy

- We are using the spaCy's Named Entity Recognizer (NER).
- We customized the NER to adapt it to our task.
- It has the capability of identifying the features and the name of a product inside a review of such a product.
- In our case it is used to generate the training dataset for the GPT-2 fine tuning.

FLOWYGO

Model for the text generation

GPT-2 Medium
A larger model

- **345M Parameters**
- **1024 Model Dimensionality**

- For the textual reviews generation we are using the GPT-2 medium.
- GPT-2 Medium is the 355M parameter version of GPT-2, a transformer-based language model created and released by OpenAI.
- It is pre-trained on a set of about 40GB called WebText.

Name Entity Recognition

Croma Electronic Products Dataset

Amazon Phone Dataset

Amazon Product Info Sample

Merged dataset of reviews

INPUT

NER

OUTPUT

Dataset of tagged reviews

TRAINING

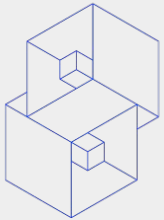
Data pre-processing

Manually Annotated Reviews

('Use Nikon Z 5 24.3MP Mirrorless Camera which with its high-definition videos, and simultaneous generation of time-lapse movies during interval-timer photography..... ,
{'entities': [(1069, 1085, 'ATTR'),
(454, 467, 'ATTR'),
(265, 294, 'ATTR'),
(1038, 1047, 'ATTR'),
.....]})

GPT-2

Output of the NER



Tokenization

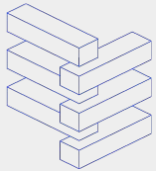
Proper Training Dataset

Generated Textual Reviews

"<OVERV_START> <NAME_START> " + name +" <NAME_END> <FEAT_START> " + " <NEXT_FEAT> ".join(attributes) + " <FEAT_END> <DESCR_START> " + \
description + " <DESCR_END> <OVERV_END>"

FINE-TUNING

Structured Data Products

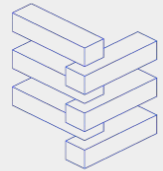


Data pre-processing

SEED

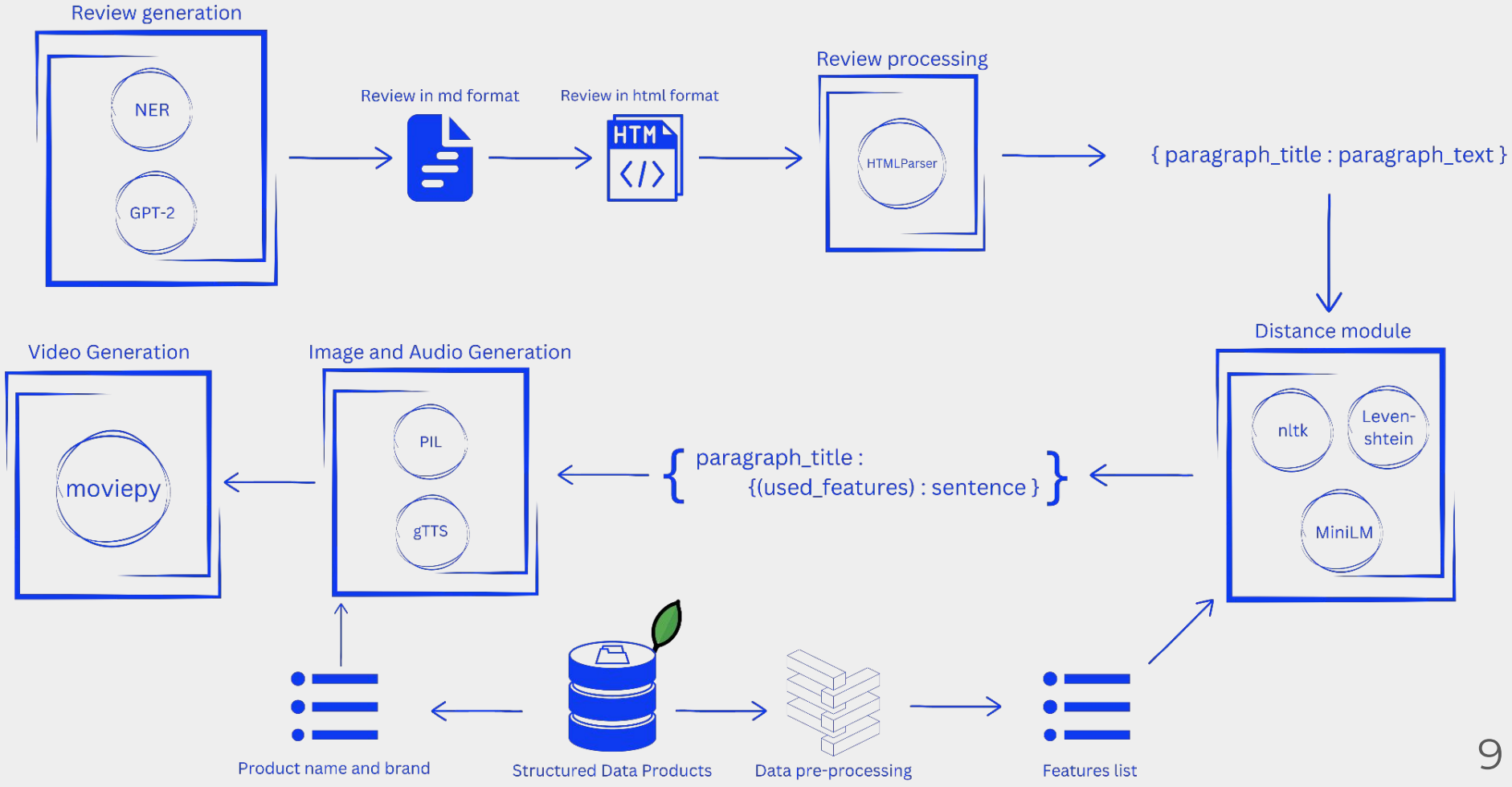
""<OVERV_START> <NAME_START> DELL PRECISION SERIE 7000 7560 ON7560WM01AU <NAME_END> <FEAT_START> Full HD Display <NEXT_FEAT> 11th gen Intel Core i7 <NEXT_FEAT> 15.6 inch <NEXT_FEAT> 1920 x 1080 pixel <FEAT_END>"

OUTPUT



Data post-processing

Video Generation



Metrics

A first evaluation

- We can identify two types of match:
  - perfect match
  - semantic match

- On a first evaluation, we obtained:
  - sentence embedding sees approximately a micro average loss in perfect match of 6.7%, and a macro of 12.5%
  - sentence embedding allows for semantic matches, e.g.:
    - slim flat screen
    - greater viewing area 65 inches
    - remote is easy to operate and the remote button is easy to press remote one touch access

Experiments

NER
Evaluation

Evaluation method:

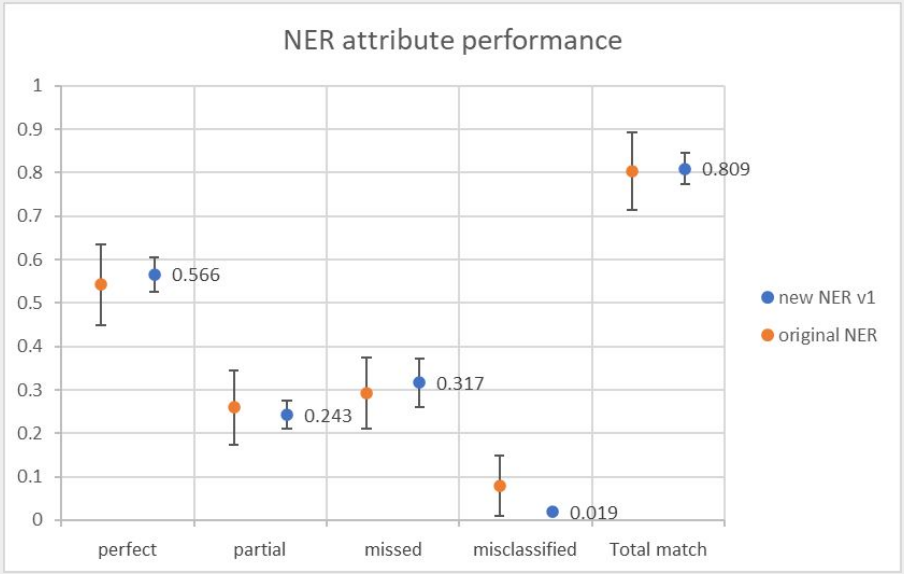
- K-Fold with 10 split on the manually annotated reviews

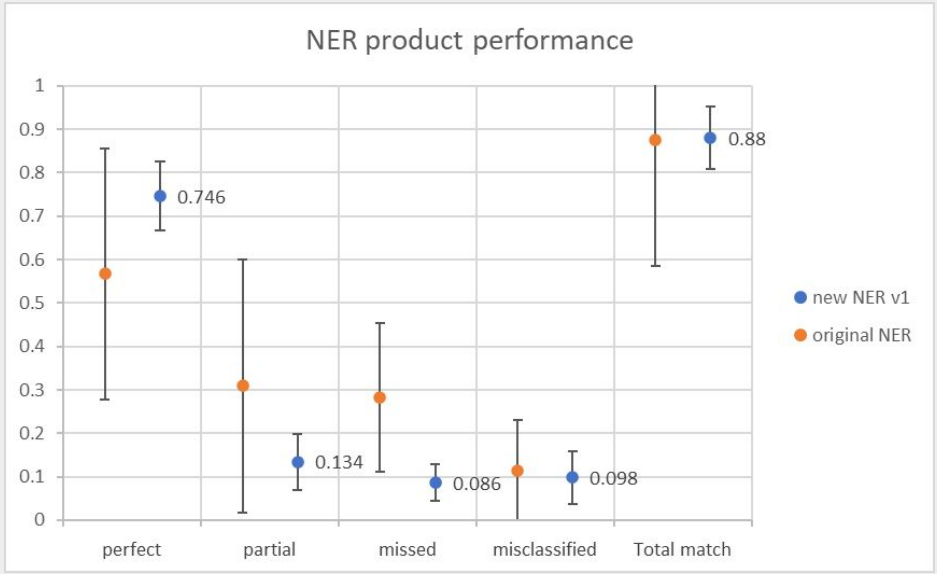
Metrics for evaluation :

- Perfect match
- Partial match
- Missed
- Misclassified
- Total match

FLOWYGO

NER Performance : baseline vs final model





Main focus:

- In attribute performance, drastically decreased the error of misclassifying an attribute with product
- In product performance, we increased the overall performance both in accuracy and stability

GPT-2

Evaluation metrics

Evaluation method:

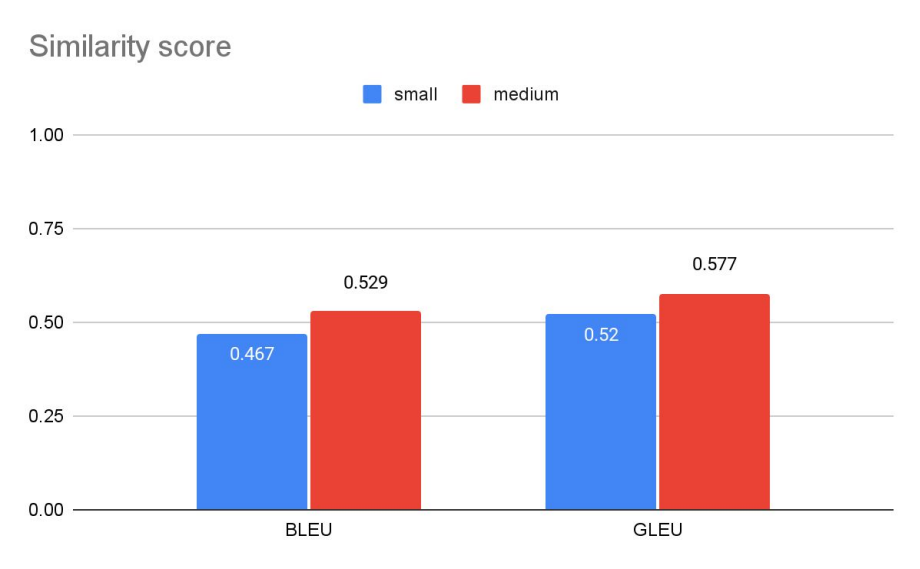
- comparison between original description and generated review for each test-set product

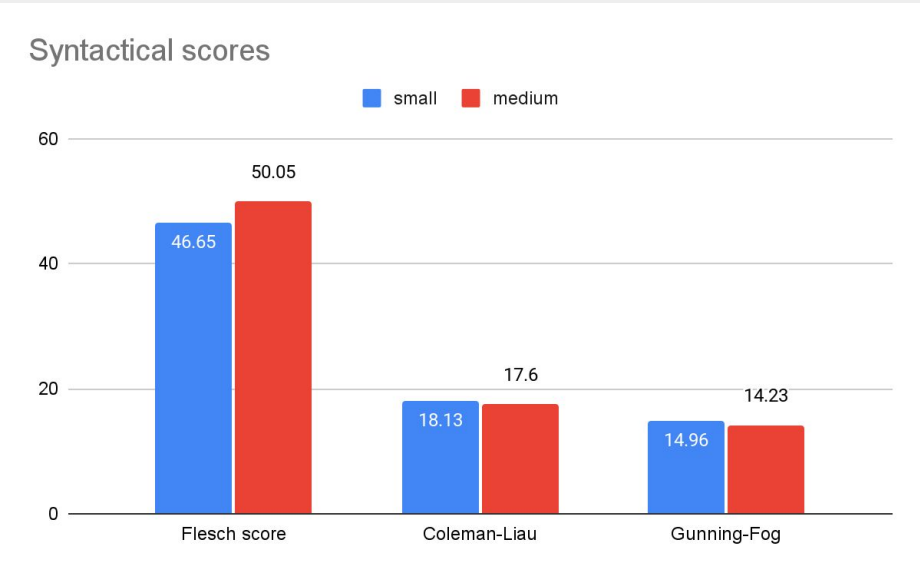
Metrics for evaluation :

- Similarity scores:
  - BLEU
  - GLEU (Google BLEU)
- Syntactical scores:
  - Flesch reading ease score
  - Coleman-Liau index
  - Gunning-Fog index

FLOWYGO

Text Generation: baseline vs final model





Main focus:

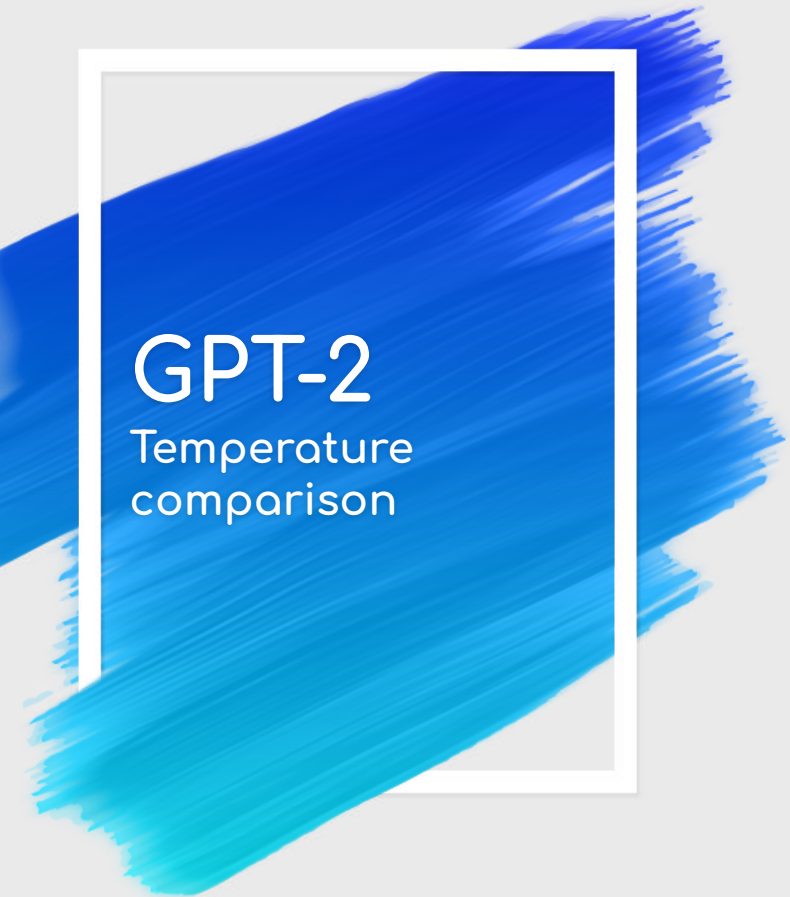
- Improvement in both BLEU and GLEU performance
- Medium model presents more readable text compare to small model

GPT-2

Temperature comparison (HISENSE H55B7500 TV)

With a higher level of temperature

Display your Android and iOS TV and its content conveniently using the H5500's remote, easily controlling media on it. The Remote is also compatible with smartphones and tablets and works well with most of them. The remote allows you to easy use your devices without any hassles. Besides, it is easy to pack and operate as it comes with USB type-a Usb Connector Type, Parental Controls and Usb type-a Usb Connector Type Cable.



With a lower level of temperature

The most advanced TV with advanced features. The all new and powerful HISENSE H55B7500 TV with advanced features makes it the perfect companion for your home. It comes with a powerful processor that provides the best 4K UHD experience. It comes with a powerful storage capacity and a powerful storage capacity. It is also equipped with Usb type-a Usb Connector Type. It is equipped with Audio Return Channel Arc. It comes with Parental Controls. It has Game Mode. It has Subtitles. It has High Dynamic Range Hdr 1000. It has Extended Pvr.

Conclusions

Conclusions

summary

Named-entity recognition (NER):

- Increased dataset manually annotated
- Better and more stable model performance

Text-generation-model (GPT-2):

- The better NER model, The better the training data.
- Larger text generation model: GPT-2 medium
- New metric to evaluate the quality of generated text

Text-to-speech and text-to-video:

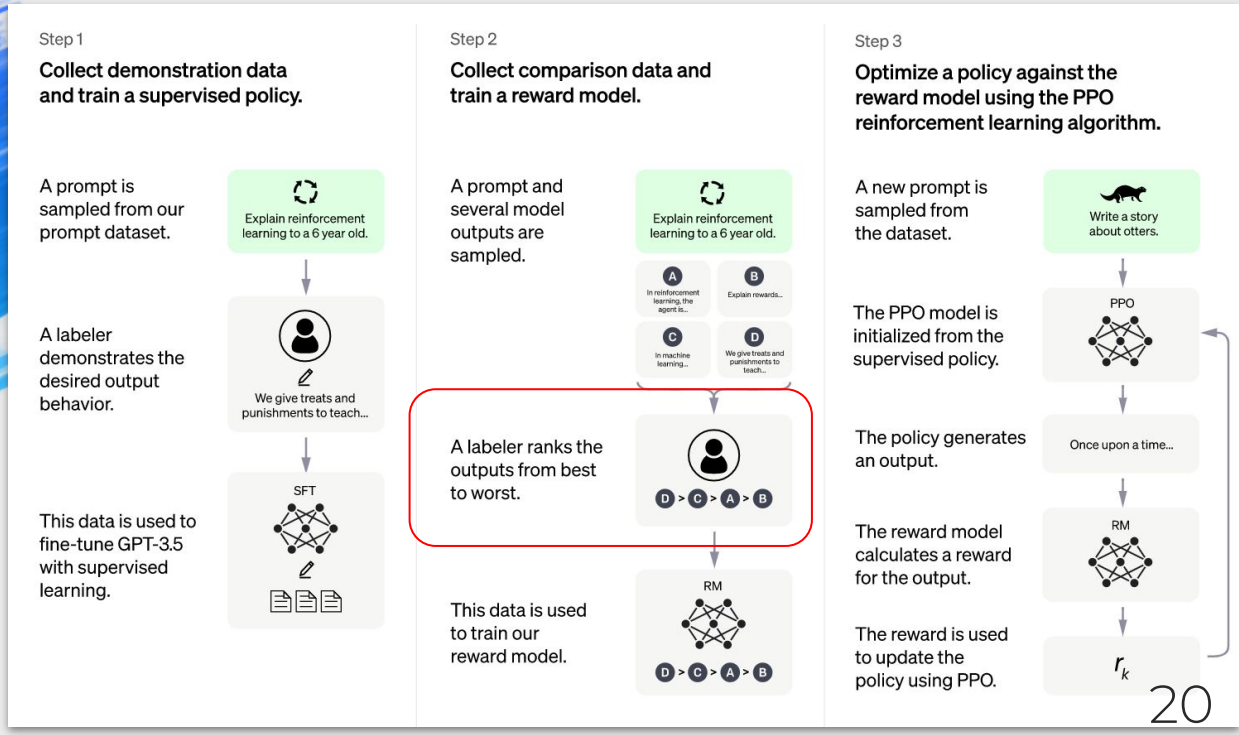
- Text-to-speech: gtts.
- Slides generation: PIL.
- Video generation: moviepy.
- The distance model.
- This system overall gives satisfactory results.

FLOWYGO

Conclusions

Evaluation always becomes a pain point:

- How good is the training dataset prepared by ner?
- How many people like the generated overview?
- How good is our video?



Conclusions

Future work

Future work:

- Make the GPT2 model more robust in different domains.
- Make the generated reviews adapt to more different groups of people.
- Make the final video more diverse and humanized

Hisense h55b7500

Hisense

Thanks you for your attention. Questions?

Contacts:
Luca Agnese: lucaagnese@hotmail.it:
Matteo Donadio: matteodonadio00@gmail.com
Tianming Qu: qutianming0930@gmail.com
Flavio Spuri: spuri.flavio@gmail.com
Jiahao Zhang: giacomino99217@gmail.com