T.C.

MARMARA UNIVERSITY

FACULTY of ENGINEERING

COMPUTER ENGINEERING DEPARTMENT

CSE4197 Engineering Project

Project Specification Document

Pics2Story: A.I. Storyteller

Group Members

Kerem Kosif	150119909	keremkosif@marun.edu.tr
Sinan Dumansız	150119812	sinandumansiz@marun.edu.tr
Nidanur Tekin	150119794	nidanurtekin@marun.edu.tr

Supervised by

Prof.Dr. ÇİĞDEM EROĞLU ERDEM

Cigdem Erdem

30/11/2022

Contents

1. Problem Statement

Pics2Story is an artificial intelligence project where users upload an image/photo and get a story in return for it. This project consists of a combination of two artificial intelligence algorithms. The first algorithm to be used in the project is to create labels by recognizing the objects in the photos. In the next stage, these tags will be used by the seq2seq algorithm to generate stories.

2. Problem Description and Motivation

Stories are used in many parts of our lives. Most popularly It's used in entertainment and educational industries. Even some news companies use story generation algorithms to generate news according to their subject. However, story generation is tied up with fundamental AI research problems. Algorithms should plan with the language and give communicative intent. Also, one should understand the language and have commonsense knowledge. However, these requirements are very challenging tasks even for artificial intelligence.

Our approach in this study field is that we will use images as a source of information for the generated story. Using images one paragraph story will be generated. The architecture will consist of an image captioning algorithm and a story generation algorithm. Image captioning is a sequence to sequence problem which generates descriptive explanations for a given image. COCO [8] is a popular dataset used in the training process of these algorithms. It also contains a wide range of image and caption pairs. In Figure 1, two images from the COCO dataset are shown together with their captions. On the other hand, the story generation algorithm also is a sequence to sequence algorithm. It takes small sentences as input and tries to generate meaningful stories related to the given target.

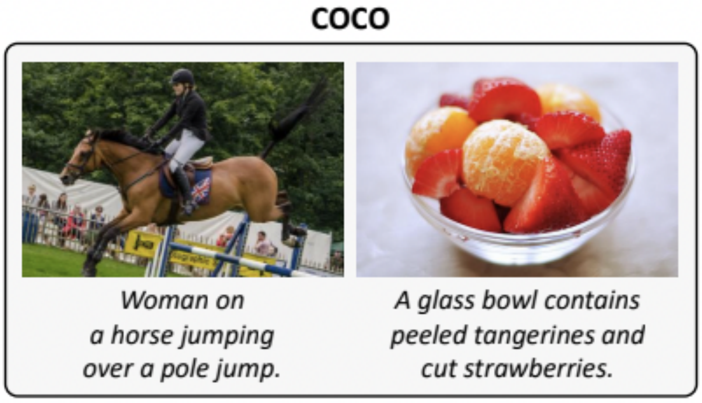

COCO
Woman on a horse jumping over a pole jump.
A glass bowl contains peeled tangerines and cut strawberries.

Figure 1: Example captioning from MS COCO dataset. We will generate similar captions in the first step [13].

There is no research on image to story generation that has not been done in the literature. So we are motivated to bring light to an unexplored field of artificial intelligence. Due to its structure, the story generation algorithm is a seq2seq model which can also be used in other fields besides storytelling. Some of these use cases are: It can try to capture the composition of an artwork [12] and give descriptive text about the painting. This model can give another perspective over human generated artwork compositions. The model could be trained for generating descriptive and detailed reports from CT and MRI scans. Finally, there can be other applications like using pictures of a certain machine to detect the defects and giving a report according to it.

3. Main Goal and Objectives

In this project our main goal is to create a multimodal deep learning model for story generation from images, which has not been done in the literature. First, we plan to fine-tune a pretrained captioning algorithm for the caption generation. Then, we will fine-tune a seq2seq algorithm to generate a story. Our objectives are as follows:

Objective 1: Implementing the Captioning Algorithm

In this project we will implement an algorithm to process the image and generate captions. We will test a couple of state of the art image captioning algorithms to find the most suitable for the pics2story model.

Objective 2: Design and Implementing the Story Generation Algorithm

We will design and implement a seq2seq model to generate stories. In this project we plan to fine tune one of the successful seq2seq algorithms like Google's T5 [4] or OpenAI's GPT-2 [2] alongside Fair's hierarchical neural story generation algorithm.

Objective 3: Creating the Multimodal Network

In the final step, the image captioning model and fine-tuned seq2seq model will be merged. For the final results we expect to generate one paragraph story from an image.

Objective 4: Creating a Web Application/Mobile App

We plan to reach the end user to create a web/mobile system where membership-based photos are uploaded and a story is received as a result.

4. Related Work

Story generation and image captioning are very popular research topics. However, not much work in recent years has been done on image to story generation. Below, we briefly review the image captioning methods and story generation methods in the literature: Neural-Storyteller [5] is the first image to story multimodal research in literature. The story generator was created using the researchers previous works. Generated stories were somehow related to the image but it lacked many components of human-written stories. Main reason for that is that developments in deep learning and natural language processing weren't enough for creating a meaningful context. There have been some studies to create a plot from several images [7]. The visual storytelling algorithm takes 5 sequences of images and tries to generate a series of captions which eventually forms a plot. However it's different from our model because we are using a single image and our purpose is to inspire the story using an image. Whereas the visual story generation model's purpose is more oriented in figuring out what is happening in several pictures. And finally, “Explain Me the Painting” [12] is about generating artwork descriptions. In this research tabular data of the artist and his related artwork is merged into a fusion model. This study is closer to our model however they are fusing tabular data to specialize strictly on famous paintings of famous painters. Even though there are some similar works, there are not recent studies for image to story generation that utilizes state of the art algorithms.

5. Scope

The main purpose of the project is to develop an algorithm that creates meaningful stories from images. With the caption generated from an image, it is aimed to connect the semantic integrity of the story with the given image. Inorder to generate more creative stories personalized captions will be used to fine-tune the captioning algorithms. The resulting story is expected to be 110-175 words long. The language of the created story will be English.

Constraints:

- Created story doesn’t have to represent the image perfectly. Instead the story should be inspired from the image.
- There shouldn’t be repeated words in the generated story.
- The captioning algorithm will be trained on MS COCO dataset and fine-tuned using InstaPics dataset. Finally “WritingPrompts” will be used to train the story generation algorithm.
- Implementations will be done using Python language
- Cloud GPU services will be used for fine-tuning and training of deep learning models.

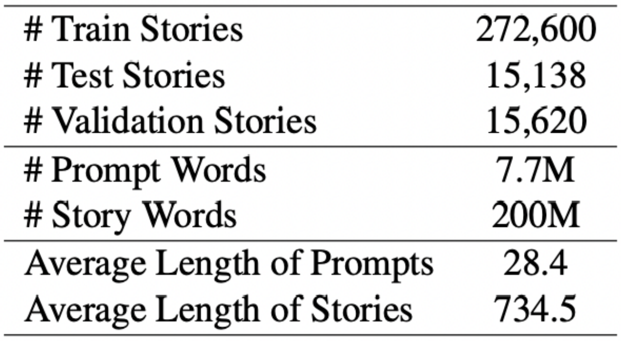
6. Methodology and Technical Approach

A. Introduction

Story generation from an image is a very complex problem. The main reason is that the architecture is composed with state of the art algorithms. Both recurrent and convolutional networks will be used to generate stories. So, our story generation model consists of image captioning and story generation parts. First, an image captioning algorithm will be created. Later, a fine-tuned seq2seq model will be implemented and trained. Finally, these two models will be merged and ready to generate stories from given images.

B. Writing Prompts Dataset

In this project we will be using FAIR's dataset "WritingPrompts" [1] to train the story generation algorithm. "WritingPrompts" is a community where online users inspire each other by submitting story premises or prompts, and other users are free to respond. Each prompt can contain multiple story answers. Prompts come in different subjects, lengths, and details. Stories must be at least 30 words, avoid general profanity or inappropriate content, and be inspired by the given prompts. In this step the WritingPrompts dataset will be preprocessed and reorganized for the training story generation algorithm. There are approximately 300.000 stories in the dataset and 5% of the prompts for a validation set and 5% for a test set will be reserved accordingly. In Table 1 statistical information of the dataset is shown.



# Train Stories	272,600
# Test Stories	15,138
# Validation Stories	15,620
# Prompt Words	7.7M
# Story Words	200M
Average Length of Prompts	28.4
Average Length of Stories	734.5

Table 1: Statistics of the WritingPrompts dataset [1].

C. The Image Captioning

The first step of creating the image story model is fine-tuning the image captioning model. The captioning algorithm consists of two parts: a visual encoder part and a language model. For the visual encoder, there are several methods but we decided to use attention over visual regions because of its high performance. In this approach an object detector is in charge of proposing image regions. This is then linked with a mechanism that learns to weigh each region for each word prediction. For object detection, Faster R-CNN is adopted to detect objects and acquire a pooled feature vector for each region proposal [13]. For the language model, transformer networks will be used because of its improved results over LSTM models. The Transformer starts applying a masked self attention on words. Then, words are used for queries and outputs of the last encoder layer used as keys and values on a cross-attention operation. Architecture ends with a final feed-forward network as can be seen in Figure 2

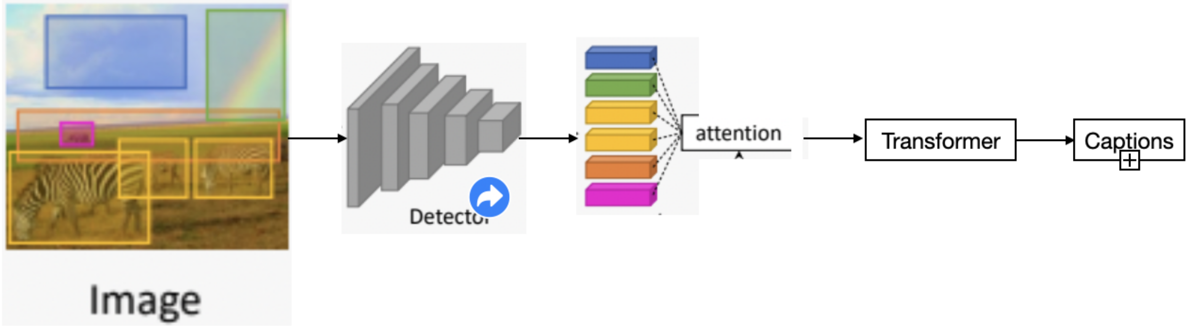


Figure 2: In the figure process of an image captioning with attention over visual regions a Transformer model [13].

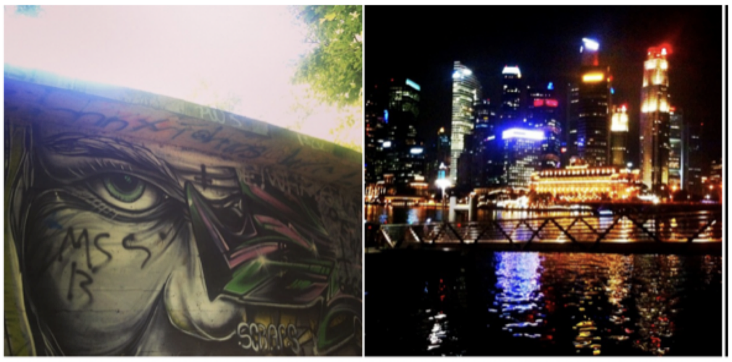
For image captioning there are several state of the art algorithms. These algorithms use different techniques to master the image captioning problem. We will test some of the captioning algorithms to find the best one to create relevant captions from the images. Three state of the art algorithms we have chosen for testing and comparing are Oscar [9], DLCT [10] and X-LAN [11]. The Oscar model uses BERT [15], DLCT uses Transformer networks [14] and X-LAN uses LSTM architecture. These models have proved that they are one of the best algorithms in their own language model. As for the visual encoder, they all use attention over visual regions.

Model	Visual Encoding	Language Model
Oscar [9]	Attention over visual regions	BERT [15]
DLCT [10]	Attention over visual regions	Transformer [14]
X-LAN [11]	Attention over visual regions	LSTM

Table 2: Selected captioning algorithms that will be tested during the project.

D. Captioning Stylization

Image captions hold a very important position in the Pics2Story algorithm. The main reason is they connect the semantic integrity of the story with the given image. So, the story generation algorithm takes a prompt as an input and generates related stories. However generating descriptive and realistic captions may affect the creativity of the generated stories. To solve this, we plan to fine-tune a state of the art captioning algorithm with a creative dataset. For this task we will use the InstaPic dataset from the “Attend to You” [3] paper. The InstaPic dataset was created by collecting instagram pictures and post descriptions from several different users. Since these captions are written in a more personalized language than COCO [8] captions, they will be used to generate sentimental captions. Eventually, these captions will give rise to more variety to the created stories. Figure 3 shows an example of captions created from the descriptive and plain forms of the objects in the images made with the InstaPics database.



(GT) the face in the woods

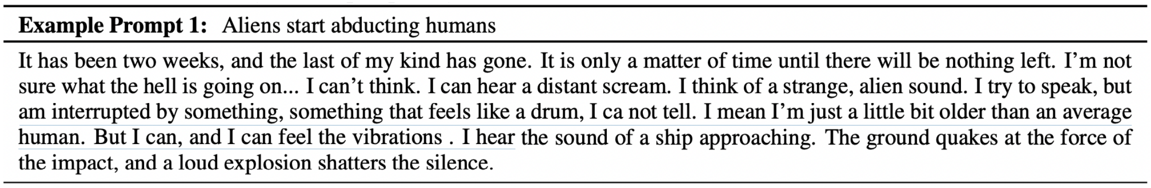
(GT) awesome view of the city

Figure 3: Examples from the personalized image captioning dataset InstaPics [3].

E. Story Generation

For story generation sequence to sequence architecture will be used. Seq2seq architecture consists of an encoder and a decoder unit. The encoder unit transforms each caption into a corresponding hidden vector containing the element and its context. The decoder reverses the process and using the previous output as input context, transforms the vector into output stories. Mainly three models will be tested and the results will be compared to fit the desired perplexity score. These models are Google’s T5 model [4], OpenAI’s GPT-2 [2] and finally FAIR’s Hierarchical Neural Story Generation [1].

Fair's "Hierarchical Neural Story Generation" is the first model that is going to be tested. Models gain further improvements over generic models with a novel form of a model fusion that improves the relevance of the story to the input. Also, adding a new gated multi-scale self-attention mechanism to model long-range context. [1] Alongside with FAIR's algorithm Google's T5 algorithm [4] and OpenAI's GPT-2 [2] algorithms will be fine-tuned to generate stories. Both GPT-2 and Google T5 are strong examples of seq2seq models trained under terabytes of datas for all purpose usage. In this project we plan to test all algorithms and try to hit 38-40 perplexity scores on the “WritingPrompts” dataset. In Figure 4, Example story generated by the hierarchical fusion approach.



Example Prompt 1: Aliens start abducting humans

It has been two weeks, and the last of my kind has gone. It is only a matter of time until there will be nothing left. I’m not sure what the hell is going on... I can’t think. I can hear a distant scream. I think of a strange, alien sound. I try to speak, but am interrupted by something, something that feels like a drum, I ca not tell. I mean I’m just a little bit older than an average human. But I can, and I can feel the vibrations . I hear the sound of a ship approaching. The ground quakes at the force of the impact, and a loud explosion shatters the silence.

Figure 4: Example Story [1].

F. Web Application

Finally, we plan to reach the end user and create a web/mobile system where membership-based photos are uploaded and a story is received as a result. Although these targets are in the early planning stages, the systems we plan to use include software frameworks like Flask and React. In the backend we plan to deploy this code on Heroku so that the trained data will deal with the frontend systems. In the frontend part, we want to offer end users two different options, mobile and web. We want to create a mobile application with React native, use Next Js as a web framework, and finally publish it on a platform basis.

7. Professional Considerations

This section of the PSD document, topics such as engineering standards, Realistic Constraints and Legal Issues related to the project are examined and explained in detail.

A. Methodological considerations/engineering standards

- We will use Github for source control and Google Drive as a document archive along with project planning.
- Depending on the task we may use Keras or Pytorch as main libraries while implementing the project.
- Python will be the main language to implement algorithms and model training.
- For training we will use the MS COCO, InstaPics and WritingPrompts datasets.

- Python Flask will be the main backend framework and Reactjs will be the main frontend framework

B. Realistic Constraints

In this section, we describe our solutions for project design and what might be encountered in the next phases of the project, examining potentially hazardous scenarios and other aspects that may affect project completion. In these sections, economic, environmental, ethical, health, sustainability and social constraints are explained as stated in the subsection titles.

a. Economic

The end-user product that will come out of the project to be done is currently not available in any market place or in open source projects (in Github etc.). However, as research projects, there are some examples in research papers. If we succeed in finishing the project with our current goals, we will have to pay a certain fee for the backend version of the code to run on the server. The most likely of these would be to deploy to the Heroku site. Heroku helps developers store and open their code to the outside world in exchange for certain packages. Most entry-level packages start at $5 per month. If we are going to implement this, we will have to pay 60 dollars annually from 5 dollars a month. In order to turn this expenditure into profit, we can create a mobile application or website as a product for the end user and enable users to upload their photos/images/paintings into the story for a certain fee. In the same way, a certain fee is charged by both Google and Apple for the website or mobile application to be listed in the markets. Since the realization of the project and making a profit from it will bring a personal income, it has no other feature other than bringing dollars to the national economy.

b. Environmental

There is no physical material or tangible electronic component in the project, there is no situation that will affect environmental problems directly. But if we look at the needs that will be required for the project to finish, the backend servers that we get help from outside have certain side effects in electricity consumption and their release/effects. Apart from this, the project works with zero emissions and zero environmental pollution.

c. Ethical

In the project, we are going to create a story from the photos/images by deducing some labels from them, certain ethical problems arise in the part of uploading photos and hosting them. Especially in the product to be published to the end user, serious consequences may occur if users upload photos with adult content or photos with copyright that do not belong to them. The simplest way to prevent this would be to check the photos by an admin. The most complicated way would be to prevent banned photos by working on a photo recognition artificial intelligence just for this job. In addition, if the photos uploaded by users are captured by hackers, another problem will arise. Leak and this situation can result in serious court cases when noticed by users. As the last situation, the photos uploaded by the users can be used to be retrained and unfortunately become available for sale.

d. Health and Safety

There is no outside factor in our project that will cause any health or social problems.

e. Sustainability

The system we will establish in this project consists of new roots, since there are very few examples around. Due to this situation, if the necessary time and economic opportunities can be provided, the project can expand and can evolve very quickly. Since it is highly sustainable, it can make a permanent profit in the long run with a good business plan. Since the product to be presented to the user is not a technological tool, the software has the ability to renew itself and survive for a long time, and this project is an example of this.

f. Social

The end product to be produced in our project is based on the trained data, the more accurate and socially appropriate the data is, the more clean and understandable the output to be given to the end user. There is no element in our project that can cause any physical damage or mental effects. While it is completely universal, it does not vary according to a community or political view. It’s not overlap with social and socio-cultural values.

g. Legal considerations

We will extract certain keywords from the photos/images in the project, the process of uploading photos/images and keeping them on the server has to go through several legal processes. This can be accomplished successfully by showing end-users a consent text before uploading their photos and allowing them to disapprove of that text. In this way, legal problems that may arise in the future will be prevented. Another problem is that the datasets we will use to create the stories come from texts written by others. In this part of the problem, other people's articles can be directly associated with copyright materials.

8. Management Plan

A. Task phases and their durations

- **Phase 1:** Determining the problem for selected project topics and finding possible ways to achieve final work.
- **Phase 2:** Literature survey and searching related works about image labeling and story generation from given input words.
- **Phase 3:** Gathering all the information about our topic and writing a PSD document.
- **Phase 4:** Trying to find the most optimal way to achieve the project and making small tests in different algorithms to see if we can find a better outcome.
- **Phase 5:** Writing of ADD document by combining all the small tests and researches. Preparing for a presentation to finish the first part of the project.
- **Phase 6:** Choosing the best methods for our algorithm according to previous tests and start implementing.
- **Phase 7:** Finalizing the algorithm and testing the final model.
- **Phase 8:** Creating Mobile/Web Application based on algorithm functions.
- **Phase 9:** Writing a Thesis and preparing a presentation.

B. Division of responsibilities

The task distribution table in our project group is as indicated in Figure 5 below.

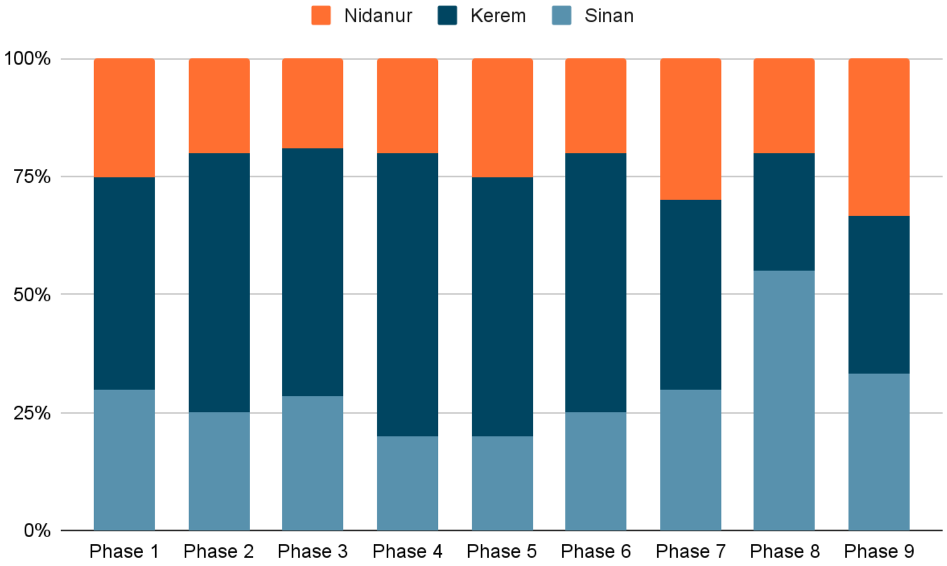


Figure 5: Contributions table

C. Timeline with milestones

The distribution of all phases from the beginning to the end of the project according to the months is as indicated in figure 6.

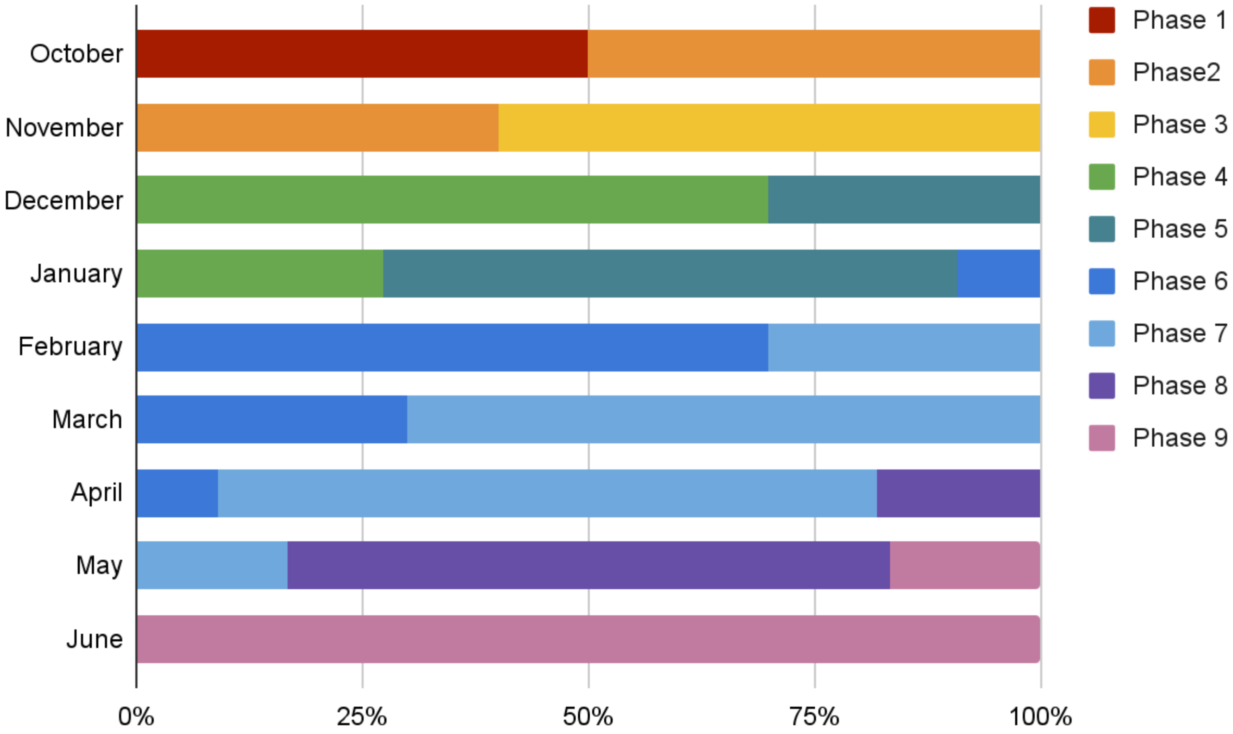


Figure 6: Monthly planning Table

9. Success Factors and Risk Management

A. Measurability/Measuring Success

a. Implementing the Captioning Algorithm

- The generated caption should describe the image with utmost details.
- The captions shouldn’t include irrelevant stuff.
- The algorithm should portray the images with some emotion.
- We will use the BLEU score to evaluate the captioning outputs in the project. BLEU stands for Bilingual Assessment Sub-study. Its main purpose is a score used to compare the candidate translation of the text with one or more reference translations. We aim to get a 38-40 BLEU score for the captioning algorithm we will use in our project.[16]

b. Design and Implementing the Story Generation Algorithm

- Generated stories shouldn’t have major grammar mistakes.
- There should be no repeated words in the stories produced.
- The perplexity score is one of the most common ways to evaluate language models. Score is used to evaluate how similar a dataset is to the text distribution when trained on a particular model. In this project we expect to get between 37-40 perplexity scores to generate stories.

c. Creating the Multimodal Network

- Image captioning and Story generation algorithms should work simultaneously.
- Generated stories from the given captions should have semantic integrity.

d. Web/Mobile Application

- The Web/Mobile application should run without any bugs.
- The story generation algorithm should create the story within the utmost 1 minute.

B. Risk Management

a. Captioning Algorithm

1. **Risk 1:** We may not be able to find pre-trained versions of the captioning algorithms that will be used in this project, and if not, training these algorithms is nearly impossible in the limited time and resources we have.
    - ○ **Solution:** To solve this problem we have chosen multiple captioning algorithm options. In case we run into this problem, the algorithm with a pre-trained model will be selected.

2. **Risk 2:** There may be an error in the implementation of the algorithm that we have chosen. The main reason for this is that the code of the algorithm is written by other contributors, not the article writers themselves.
    - ○ **Solution:** The solution to this problem is that we can choose the algorithms that are written by the author of the articles.

3. **Risk 3:** The Image Captioning algorithm that we chose is not able to create creative words from the given picture.
    - ○ **Solution:** The solution to this problem can be solved by paraphrasing the captions to add personalization.

b. The Algorithm Cannot create Meaningful Stories

Firstly, we will fine-tune the T5/GPT-2 [2][4] algorithm using the “WritingPrompts” dataset. However, the story we produced from the algorithm may not reach the 39-41 perplexity score we aim for. The main reason for this is that there may be distortions in the sentences of the story or shifts in the integrity of the meaning. To solve this problem we can try other methods such as the Hierarchical neural story generation algorithm [1] which is created by Facebook for story generation purposes.

10. Benefits and Impact of the Project

A. Benefits/Implications

Photos are details that we can encounter everywhere in our daily life. There are basically two steps used to tell the story, which is using text or images. Thanks to photographs, we can clarify many situations and explain things better. Also with this situation, the person in front of us can take a more dominant position in the events.

With the development of technology, the way of telling stories changes. The benefit of using photos/images to better tell stories cannot be underestimated. Of course, in this case, how we create the story is also important. It is important that the events are human-touching in order for the reader to be affected by the event and to better understand the event. The effect of this method increases even more, especially in the transfer of emotional content. It is possible to strengthen this effect with artificial intelligence. It is envisaged to make many contributions to humanity by producing stories from photographs.

B. Social Impact

The first of these covers visually impaired individuals. Have you ever seen a visually impaired go to an art exhibition? With the development of this technology, these individuals will now be able to participate in the exhibitions. We can think that the presented pictures/photos are transformed into a story with artificial intelligence, then printed in Braille alphabet and exhibited. Thus, the visual impairment is kind of eliminated and the artificial intelligence that is wanted to be described in the picture is transferred to them.

When we look at our daily life, we are faced with a new news headline almost every minute. Producing these news and entering the headlines should be fast and impressive so that the viewing and click-through rates are high. With the algorithm of this project, newsletters will work more easily. With the uploading of the images coming to the agency to the system, the remarkable narration of the event will now provide artificial intelligence.

In social media, we write explanations under the pictures. We share a lot of pictures describing childhood memories, family environment, experiences. Every painting has a story, and artificial intelligence is at work to tell those stories.

When the topic becomes social media, it is necessary to mention digital content production, which is one of the popular professions of today. Digital content includes many types such as images, photos, texts, maps. In such content, it can be used to create subtitles and content information by storytelling.

C. Scientific Impact

If we look at its scientific impact outside of daily life, a new algorithm will be developed in the fields of Computer Vision and Natural Language Processing (NLP) in computer science.

D. Potential Impact on New Projects

It is a project that will meet the basic requirements of many future projects. For example, by interpreting the results of the patient such as X-ray and MRI, diseases are diagnosed more quickly. Thus, the treatment process begins quickly.
So this project will have a pioneering effect for future projects.

Finally, we would like to point out that the project has no impact on national security.

References

[1] A. Fan, M. Lewis, and Y. Dauphin, “Hierarchical neural story generation,” *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018.

[2] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language Models are Unsupervised Multitask Learners,” *OpenAI blog*, vol. 1, no. 8, 2019.

[3] C. C. Park, B. Kim, and G. Kim, “Attend to you: Personalized image captioning with context sequence memory networks,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[4] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. Liu, “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer,” *Journal of Machine Learning Research*, vol. 21, 2020.

[5] J. Kiros, “Ryankiros/neural-storyteller: A recurrent neural network for generating little stories about images,” *GitHub*, 2015. [Online]. Available: https://github.com/ryankiros/neural-storyteller. [Accessed: 27-Nov-2022].

[6] “Seq2seq,” *Wikipedia*, 08-Sep-2022. [Online]. Available: https://en.wikipedia.org/wiki/Seq2seq. [Accessed: 27-Nov-2022].

[7] T.-H. (K. Huang, F. Ferraro, N. Mostafazadeh, I. Misra, A. Agrawal, J. Devlin, R. Girshick, X. He, P. Kohli, D. Batra, C. L. Zitnick, D. Parikh, L. Vanderwende, M. Galley, and M. Mitchell, “Visual storytelling,” *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016.

[8] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft Coco: Common Objects in Context,” *Computer Vision – ECCV 2014*, pp. 740–755, 2014.

[9] X. Li, X. Yin, C. Li, P. Zhang, X. Hu, L. Zhang, L. Wang, H. Hu, L. Dong, F. Wei, Y. Choi, and J. Gao, “Oscar: Object-semantics aligned pre-training for vision-language tasks,” *Computer Vision – ECCV 2020*, pp. 121–137, 2020.

[10] Y. Luo, J. Ji, X. Sun, L. Cao, Y. Wu, F. Huang, C.-W. Lin, and R. Ji, “Dual-level collaborative transformer for image captioning,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 3, pp. 2286–2293, 2021.

[11]Y. Pan, T. Yao, Y. Li, and T. Mei, “X-linear attention networks for Image captioning,” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[12] Z. Bai, Y. Nakashima, and N. Garcia, “Explain me the painting: Multi-topic knowledgeable art description generation,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[13] M. Stefanini, M. Cornia, L. Baraldi, S. Cascianelli, G. Fiameni, and R. Cucchiara, “From show to tell: A survey on deep learning-based image captioning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2022.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in NeurIPS, 2017.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pretraining of deep bidirectional transformers for language understanding,” NAACL, 2018.

[16] J. Brownlee, “A gentle introduction to calculating the BLEU score for text in Python,” *MachineLearningMastery.com*, 18-Dec-2019. [Online]. Available: https://machinelearningmastery.com/calculate-bleu-score-for-text-python/. [Accessed: 30-Nov-2022].