**Data and Project Management Plan**

**Data Management Plan**

In this research study for analyzing the capacities of image captioning with four models - **CNN-LSTM, Show-Attend-Tell, ImageBERT, and CLIP**, the data should be managed in such a way that the overall project and research can be effectively completed. The data management plan for this project involves a systematic approach including processes such as collecting, managing, storing, and using the data. Here a large amount of multimodal data mainly consisting of images and captions has been used to evolve the capacities of the four models. Below are the main steps of our data management plan.

***Data Collection Approaches***

The data collection approach includes the use of three datasets that are the COCO data set, the flicker30k, and the open image data set V7 by Google API. The reason behind choosing the COCO dataset is that it can provide a large scale of data that has been used for object detection, image segmentation, and captioning data sets. Also, this has played a significant role in image captioning as it consists of 300,000 challenging and high-quality visuals for computer vision which are mostly state-of-the-art neural networks. Also, COCO has been used by algorithms to compare the performance of real-time object detection. The required data set also has been collected from the Flickr30k dataset of about 31,000 photographs from a photo-sharing website-Flickr. This dataset includes the matching neural language description with 5 distinct human-generated captions for every image which results in 150,000 captions. Along with these two, the open image data set V7 by Google API may be used as a comprehensive tool for computerization research and development. The USP of this dataset is that it contains a collection of over 9 million photographs in various categories and can be found as a scalable database of pictures. This dataset also provides more ease of access with a better user interface. Along with this, it can be mentioned here that web scraping techniques may also be used to gather images and captions from the web such as other Google APIs and web crawlers (Hossain et al., 2019).

***Management methods***

Because the datasets that are employed are large in size, effective data management becomes essential to obtaining the desired results. Eliminating duplicate data is a crucial component of our data management strategy since it keeps our data consistent and guarantees that our models are trained on clean, non-repetitive data. Furthermore, it is essential to remove noise, outliers, and any other unnecessary data points because these could create errors and degrade the performance of our models as a whole.

We've also underlined how important it is that our data have high-quality driver photos along with captions. This entails filtering the data to make sure that, in line with the particular objectives of picture captioning, only the most pertinent and educational samples are utilized in our research. Our data management plan also includes data augmentation methods. These methods are essential for growing our dataset's size and variety, which can improve our deep learning models' capacity for generalization. We will expose our models to a wider range of circumstances by growing the dataset using methods like picture rotation, cropping, or adding noise. This will make our models more resilient and versatile when used for image captioning in real-world scenarios.

***Storage methods***

It is crucial to make sure that the large image datasets, Flicker30k and MS-COCO, can be accessed safely and easily. In this section, we go into further details about the storage management approach we've used, emphasizing the security and dependability of the data storage as well as the special nature of our approach that was designed to meet the particular needs of our study.

***Secure and Cloud-Based Storage Solutions*:** We have used cloud-based and secure storage technologies to host our datasets. We chose to use Google's Cloud Platform (GCP) as our cloud storage platform. GCP’s robust security features, global accessibility, and simple integration with cutting-edge machine learning technologies-all of which perfectly match the needs of our research, we went ahead with this platform. Not only does it offer high availability, we can also be certain that our datasets will be safeguarded against loss, unauthorized access, and data breaches with GCP.

***Customized Data Security Measures*:** A number of specialized data security measures have been put in place to further improve the security of our stored datasets. These safeguards consist of role-based access control, in-transit and at-rest encryption, and automated activity monitoring for any unusual activity. Since the data we handle is sensitive and valuable, these actions are especially relevant to the particular requirements of our image captioning research.

In conclusion, our storage management methodology for image captioning research is designed to cater to the specific demands of our project, where data security, integrity, and accessibility are paramount. By combining secure cloud-based storage solutions, tailored data security measures, and a robust backup and disaster recovery strategy, we have ensured that our datasets are not only safeguarded but also available for our research endeavors whenever and wherever required.

***Usage mechanisms***

Effective data usage is critical to deep learning research's goal of improving the area of image captioning. This research presents a well-designed mechanism for using data that makes use of rich datasets from Flickr30k and MS-COCO and stores data on the Google Cloud Platform. The first step is gathering and organizing this large dataset such that picture and caption pairings may be accessed in an orderly and smooth manner. Such careful data curation is necessary to support the establishment of the framework for later model training since it establishes the basis for well-informed and organized experiments.

This study emphasizes the significance of preparing the data in particular and is a critical step that optimizes the dataset for image captioning. Depending on the specifics of the Flickr30k and MS-COCO datasets, these preparation steps might involve image resizing, caption tokenization, and possibly domain-specific feature extraction. Furthermore, in line with contemporary methodology, the paper highlights the necessity of careful data processing and model training in order to get the intended effectiveness in picture captioning. This includes training deep learning models iteratively and carefully coordinating data pipelines. By adhering to this thorough data usage strategy, the study aims to provide the model with the necessary understanding and information to provide logical and contextually appropriate image captions, thereby pushing

**Project Development Methodology**

***CRISP DM Methodology***

As for the project development methodology, it can be stated that this project has been carried out by following the CRISP-DM intelligent system development cycle. This cycle has involved some key steps such as Problem Understanding, Data Collection, Data Processing, Model, Model Selection, Model Training, Evaluation, and Deployment. These planned development processes and activities need to be managed effectively. The 6 phases of project development have been stated as follows:

***Planned Project Development Processes and Activities***

***Problem definition*** - Establishing the primary aims and objectives of the project is the initial stage of the procedure. Using multilingual translation to enhance image captioning is the main issue here, especially for those who are blind or visually impaired. This means that selecting the ideal model for captioning images is essential. Another objective will be to develop a translation system that can generate multilingual image descriptions. The project's robust machine learning and cloud computing capabilities makes it an ideal fit for Google Cloud Platform (GCP).

***Data collection*** - During this stage, data must be gathered for model evaluation and training. MS-COCO and Flicker30k are the two extensive datasets that have been selected. MS-COCO is a great tool for training image captioning models because it has a large library of photos with thorough captions. Another useful dataset with a wide variety of photos and captions is flickr30k because GCP provides storage options such as Google Cloud Storage, using it for data storage and retrieval guarantees easy access to these datasets.

***Data analysis*** - Data preprocessing such as cleaning, filling, augmentation, and annotation of the gathered data are all essential components of data analysis. The quality and preparedness of the data for model training are guaranteed throughout this step. To expedite the data analysis process, GCP offers a variety of data processing technologies, such as Dataflow and Dataprep. These tools can be used to clean up and get the data ready for training.

***Model development*** - We will work on creating the image captioning models in this phase. Model selection, design, implementation, and training are all included in model development. It will be necessary for us to select suitable machine learning frameworks, like PyTorch or TensorFlow, which are both supported by GCP. Using strong cloud-based GPUs, we can effectively train and fine-tune our models employing GCP's AI Platform. It offers several services related to the project that can aid in multilingual image captioning, such as Google Cloud Translation and Vision.

***Model evaluation*** -To evaluate the effectiveness of the trained models, model evaluation is necessary. The quality of the generated captions can be assessed using metrics like METEOR and BLEU scores. We can also evaluate the models' processing speed and scalability. GCP offers benchmarking and monitoring tools, such as Cloud Profiler and Cloud Monitoring, to assess model performance and pinpoint areas in need of development. GCP offers extensive computational resources that can be used to effectively carry out large-scale evaluations.

***Documentation*** - The documentation includes user instructions, cloud platform configurations, data sources and processing, and comprehensive records of the image captioning model. We enable knowledge transfer and enable upcoming researchers and stakeholders to understand and expand upon the project's work by upholding comprehensive documentation. The documentation is also a great resource for version control, security considerations, troubleshooting, and data privacy. It contributes to the long-term success of our multilingual, color-focused image captioning research using the Flickr30k and MS-COCO datasets, improves project reproducibility, and supports adherence to best practices.

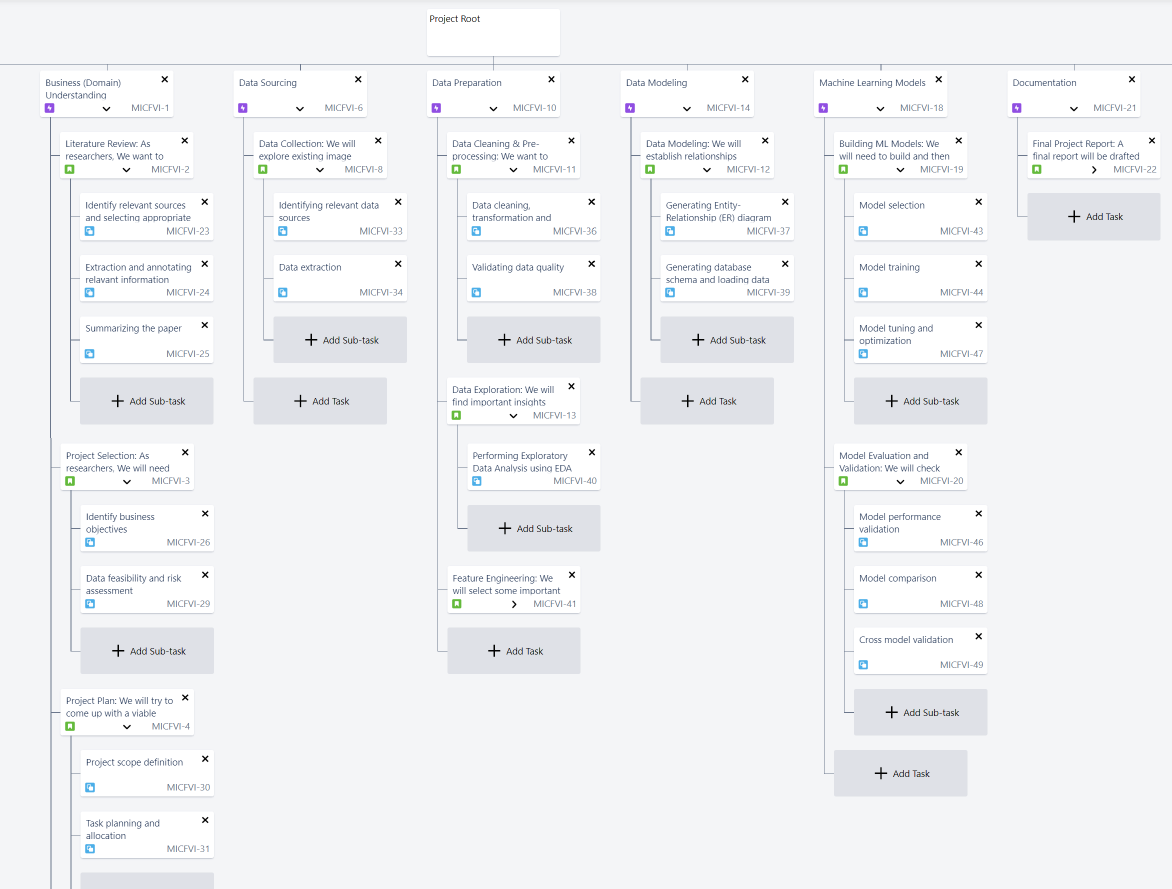
**Project Organization Plan**

The development of this project and its research can be represented with a work breakdown structure that shows the division of all phases, tasks, and deliverables in an incremental and hierarchical format. The work breakdown structure here depends on the CRISP-DM approach which arranges the work into six main phases where each stage contains different undertakings in the request for their execution.

The first phase of WBS is business (domain) understanding and project planning which is divided into four distinct user stories. Initiating the project by literature review for investigating the potential use cases of image captioning along with finding relevant sources and a deep study about color recognition algorithms and techniques. The primary emphasis is on understanding the objectives of the project, which is to unravel the intrinsic details of the object which is color-focused and can be used as an aid for visually impaired individuals. Not only this but to perform multilingual translation for diverse public understanding. A realistic project plan is incorporated which meets the requirements. Further, the data understanding phase is decomposed into one user story and that user story is divided into multiple distinct tasks. In this phase, applicable data for image captioning is examined by getting data from the MS-COCO data source and Flickr30k data source. During this phase, the initial data analysis is performed to understand the data structure testing assumptions, transformations, or imputations. In the data preparation phase, appropriate pre-processing strategies or techniques are selected to guarantee that the data is ready for modeling. The tasks that are included in this phase are data cleaning and transformation, data exploration, performing feature engineering such as extracting color histograms of the images and finally building a data model. Further setting up a database, data warehouse, and building ETL pipelines on GCP. In the modeling phase, the task is to select, train, and test four different deep-learning models that can effectively generate image captions that are color-focused and have multilingual translations. After successfully testing the models, now in the evaluation phase, the aim is to evaluate the execution of these models on variations of the BLEU and METEOR evaluation matrices and compare them with other technical standards and benchmark scores of the models. In the final documentation phase, a report is created that can frame the procedure and results of our exploration along with model assessment.

**Figure 1**

*Work breakdown structure of the project*



**Project Resource Requirements and Plan**

***Hardware Requirements***

Since machine learning (ML) models typically have heavy computation requirements, any ML project must consider the hardware requirements into the account. These models frequently contain complex algorithms that necessitate managing multiple pieces of data at once, which can be time and resource consuming. If model training times are increased due to a lack of processing power, this could lead to project delays and cost increases.

Thus, the speed and efficiency of this ML-based project may be greatly impacted by having appropriate hardware specifications, such as a powerful CPU, lots of Memory, and a suitable GPU. Additionally, it helps improve project outcomes, speed up model training, and increase model accuracy. The project's hardware requirements are listed in Table 1.

**Table 1**

*Hardware requirements of the project*

Hardware Memory Configuration Purpose

14 core CPU 16 GB memory 6 performance core Computation 8 efficient cores Requirements for ML/DL frameworks and Development Environments

5888 CUDA Core GPU 16 GB dedicated ML and DL models

memory

***Software Requirements***

Technology-focused research usually involves software packages, so in addition to hardware requirements, software requirements also need to be taken into account. To ensure optimal performance and avoid any issues, it is crucial to confirm that the software versions being used are compatible. The project's software requirements are listed in Table 2.

**Table 2**

*Softwares/Libraries Used in the Project.*

Environment/Libraries/Packages Purpose Version

PyTorch Robust framework for designing 2.1.0 deep learning models, like CNN-LSTM and CLIP

TensorFlow Framework for developing deep 2.14.0 learning models, like Show-Attend-Tell

Keras High-level and user-friendly 2.14.0 framework for developing deep learning models, like ImageBERT

spaCy Library for natural language processing 3.5 that delivers diverse functions and tools for processing and manipulating captions

OpenCV Library for computer vision that delivers 4.8.0 diverse functions and tools for processing and manipulating images

NLTK Library for natural language processing 3.7 that delivers diverse functions and tools for processing and manipulating captions

Pandas Library for data handling and manipulation 2.1.1

NumPy Library for matrix operations and 1.19.2 mathematical functions

In addition to hardware and software requirements, databases also play a crucial role in data driven projects. Table 3 depicts this project’s database requirement.

**Table 3**

*Database Used in the Project.*

Database Purpose Version

Google Cloud Databases Globally distributed based NoSQL 2019 Edition database by Google.

**Table 4**

*Tools and Licenses.*

Tools Purpose License

Google Colab Cloud based environment for the execution Proprietary

Of ML/DL frameworks

Jupyter Notebook Web based environment for the execution BSD-3-Clause

Of ML/DL frameworks

GitHub Web based software development Version Various control system

Excel Spreadsheet software Proprietary

Google Docs Web-based word processing, spreadsheet, Proprietary and presentation software

Discord Instant messaging and VOIP platform Proprietary

JIRA Web based tool for Project management and Proprietary

issue tracking

Zoom Phototelegraphy software for conferencing Proprietary

Grammarly Premium Web based tool to check grammatical errors Proprietary and plagiarism

***Project Cost and Justification***

Breakdown of the project's resource costs and rationale is provided in Table 5. This table includes an explanation for each expense and the expenditures associated with the resources required to complete the project.

**Table 5**

*Project Resources Cost and Justification*

Resource Justification Duration Cost

Hardware Cost Required Hardware to perform the 4 $1700 Research

Google Cloud Database used to store the data 4 $150 Databases

Grammarly Premium Tool used to check grammatical 4 $120

Errors and plagiarism

**2.5 Project Schedule**

***Gantt Chart***

It is a project management tool that is used to schedule tasks and determine their current status. It usually represents the timeline of a particular task in the form of a horizontal bar that shows the task’s start and end dates. This tool is used to schedule tasks, monitor the progress of tasks, allocate resources, identify dependencies and relationships between tasks, and identify the priority of tasks. This tool works as a bridge between team members of the project and stakeholders by ensuring that the complete status of the project is transparently communicated to stakeholders. It ensures that the project remains on track by providing the red exclamation sign in front of the task that is not completed on scheduled time, a yellow exclamation sign if the task didn’t start on scheduled time, and a green exclamation sign if the task is going to start in next few days which also aids to keep the budget on track.

The Gantt Chart (Wilson, 2003) for this project is created using the WBS Gantt Chart plugin for JIRA software and our Gantt Chart follows the waterfall model that is each phase of the project begins only when the previous one ends and there is no overlapping present between them. Figure 2 provided below depicts the first phase of the project which is business understanding. In this phase, the goal was to determine the domain by performing tasks such as an extensive literature review where several research papers were referenced to identify the relevant papers to our project which are papers with the concept of image captioning, multilingual translation in image captioning, and using color characteristics to specify the color focused captions for the images, then the summary of all these papers were created to get an overview of the methodology followed in each paper. This is an important step to determine if it is sensible to proceed in this domain or not, it also assists in getting a clear image of the current methodologies and what are the flaws in them. It helps to set the base to perform the future tasks planned in this phase like selecting the project which includes identifying the objectives of our project, the feasibility of data, and the novelty that could be achieved; creating a project plan that includes identifying the scope of the project, potential risks, planning of tasks needed to be done and allocation of necessary resources; developing project plan by doing SWOT analysis of team members and drafting proposal. It is easy to see from the figure provided below that each task is assigned to one team member and is dependent on the status of other tasks. It is also ensured that the number of hours allocated to each team member is almost the same so that the workload is distributed equally and no task exceeds the limit of 2 days so that the level of granularity in tasks is properly maintained. Some tasks are running in parallel like “Identifying business objectives” and “Data feasibility” to make sure that no team member is sitting idle for too long and the progress of the project doesn’t fall onto the shoulders of some team members only. All the tasks in this phase were completed according to the provided timeline.

**Figure 2**

*Gantt Chart of Business Understanding Phase*

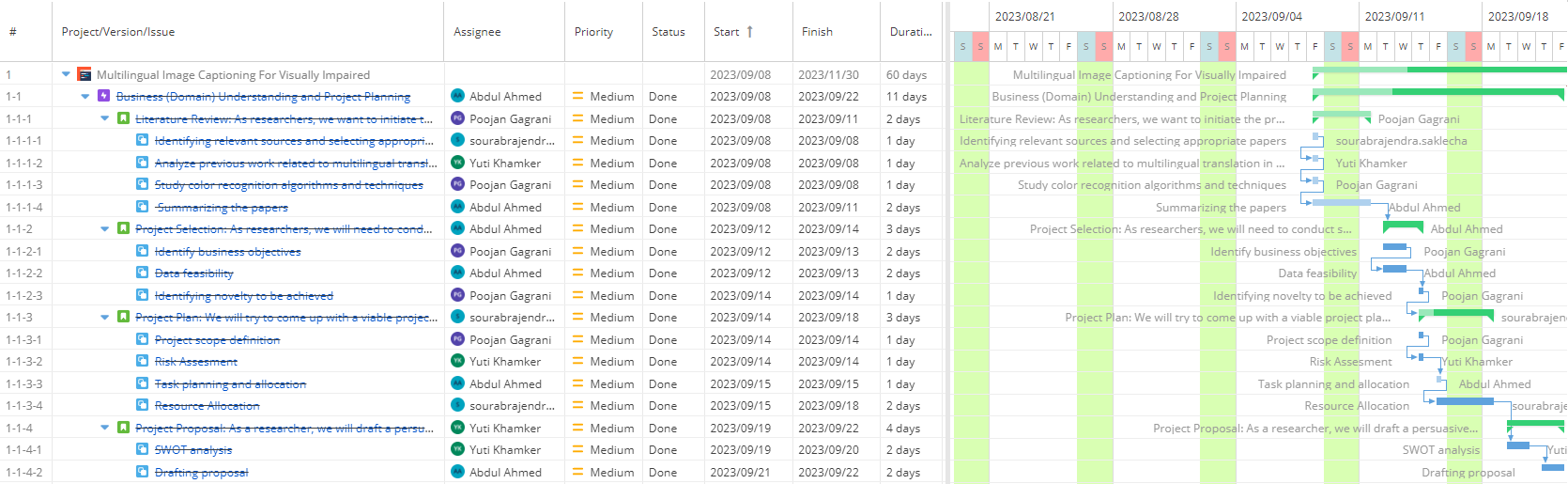


Figure 3 provided below portrays phase 2 of the project which is data understanding. In this phase, the main aim was to get comprehensive information about the data and its quality. This phase involved various tasks like identifying relevant data sources, setting up the framework to perform extraction of data from data sources, carrying out exploratory data analysis (EDA) to get a clear understanding of the quality of data, and performing data labeling and annotations. Identifying the relevant data sources is one of the most important tasks as it helps to determine the quality of data and the accuracy of the analysis performed. Because of this reason, the priority of the task was set too high. After scouring through a lot of data sources, some of the data sources like MS-COCO (Lin et al., 2014), Flickr30k (Young et al., 2014), and Google Open Images v7 were found relevant to the project. EDA is performed on the data extracted from relevant data sources in Google Colab to get a clearer view of the data and to determine if there is something missing data or values that are needed for the project. It is quite evident from the figure below that all the tasks are assigned to different members of the team to ensure equal workload and that all the tasks are completed within the assigned time.

**Figure 3**

*Gantt Chart for Data Understanding Phase*

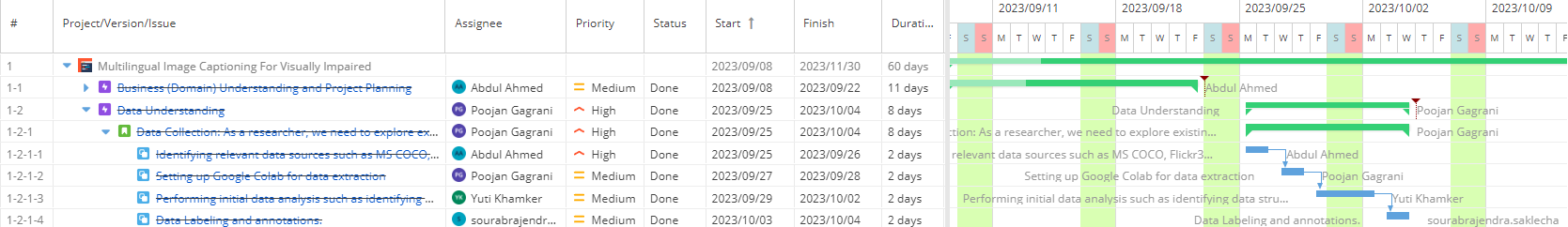
**

Figure 4 shows phase 3 of the project which is data preparation. This phase focuses on preparing raw data for further analysis by carrying out tasks like data cleaning, data pre-processing, data exploration, and feature engineering. Phase 3 began with performing extensive data cleaning on the already loaded data on Google Colab. This is done by using the capabilities of pandas and other libraries in Python. Data cleaning is given high priority because only after the data is cleaned and transformed it is ready to perform EDA and other tasks in the data preparation phase. After handling data quality issues like misleading captions, ambiguous descriptions in captions, inconsistency between captions and images, and lack of contextual information the data is validated to check if any issues are still left to be resolved so that there would be no problem in the future during modeling and rest of the phases. Pre-processing of images is also done in this phase to intensify the quality by removing noise and preparing it to perform further analysis. Followed by the extraction of color features from the images to get training data where the captions contains text with color specifications mentioned in it.Data cleaning aids in providing the data free from all the quality issues, on which further analysis can be easily carried out. EDA is again performed on the cleaned data to extract meaningful information and get more insights about the data. This proved obliging in determining which features or fields needed to be selected for modeling, generated if there is any field that is needed and can be derived from the combination of one or more fields, and transformed if its data type doesn’t prove useful in analysis but changing it could help to get valuable insights. EDA also assisted in extracting important features of the images such as color histogram, detection of objects, etc. The selection of important features during feature engineering was beneficial during data modeling as the final features that would be used in modeling were already known. Some tasks were performed on the same day like generating the ER model, doing database setup, assessing the feasibility of the data warehouse, and setting up the data warehouse so that the project with all the phases could be completed before the assigned deadline for final report submission and no team member would be idle for a longer time. All the tasks were completed by all the assignees within the assigned time in this phase.

**Figure 4**

*Gantt Chart for Data Preparation Phase*

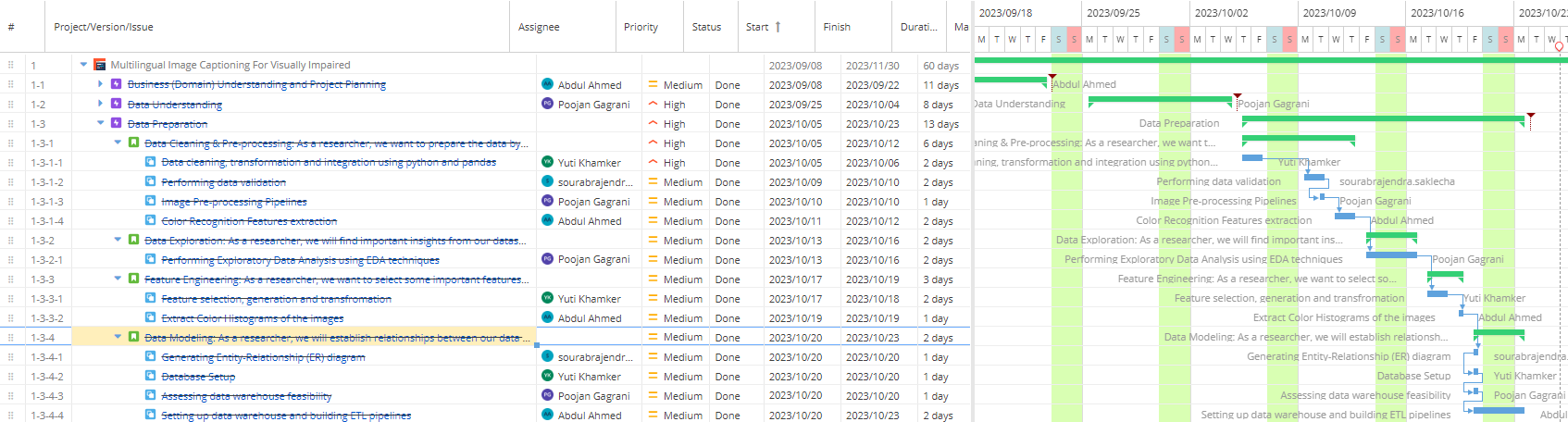
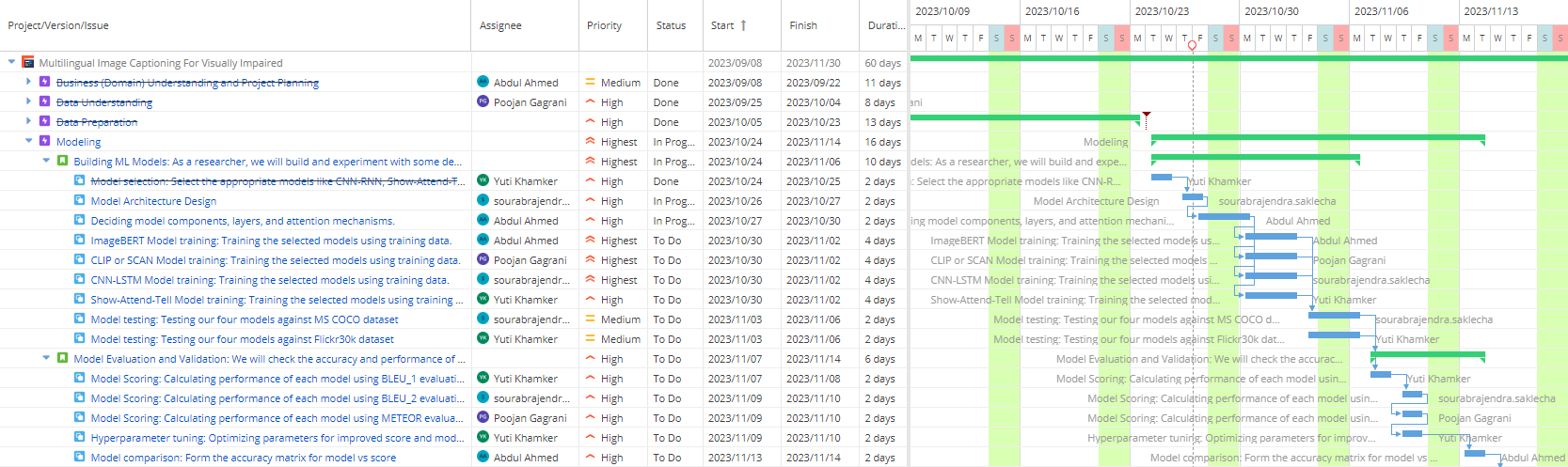
**

Figure 5 shows the tasks and the dependencies of phase 4 which is modeling. The objective of this phase is to create and evaluate the machine learning (ML) models using the capabilities of deep learning to perform the identification of objects in images and capture captions using vector support models. This phase started by doing substantial research on different kinds of models that could be used to make predictions.CNN-LSTM (Nakamura et al., 2021) , Show-Attend-Tell (Chen, 2023), ImageBERT (Qi et al., 2020), and CLIP (Fang et al., 2022) were some of the models that were finalized to go ahead with. Then, the creation of model architecture design is done where the configuration and structure of the neural network are determined along with the possible algorithms that could be used, the type and the number of layers that will be there in the neural network, and the activation mechanism that may be useful, etc. Some of the tasks that are planned to be done in the coming weeks are splitting the dataset into training and testing data for both the datasets MS-COCO and Flickr30k; taking training data created from different datasets like MS-COCO, Flickr30k, etc to train the above-mentioned models; and testing the accuracy of models for both the datasets by evaluating them against the test data of the datasets. Training of each of the selected models will be done by each member of the team simultaneously to ensure the faster completion of the phase and efficient usage of the resources. After the creation of the model, the next tasks are evaluating the performance of each model against different evaluation metrics like BLEU\_1, BLEU\_2, BLEU\_3, BLEU\_4 (Papineni et al., 2002), METEOR (Banerjee & Lavie, 2005); carrying out hyperparameter tuning to improve the model’s performance and obtaining maximum precision by optimizing parameters; and performing cross-model validation by creating an accuracy matrix where comparison of models concerning the evaluation scores obtained for different metrics. The priority assigned to most of the tasks mentioned in this phase is either high or highest because the model creation is a very important step to classify the target level which is the color-focused multilingual captions for the images in our case and obtaining the best prediction model is the ultimate goal of the project. This phase is currently in progress and the tasks are completed by the assignees as per the scheduled timeline.

**Figure 5**

*Gantt Chart for Modeling Phase*

**

Phase 6 of the project shifts the focus toward the evaluation of the project. Figure 5 shows the tasks that are planned to be completed and their corresponding dependencies in this phase. The main purpose of this phase is to evaluate to determine if the project meets the intended expectations or not. It would be beneficial to determine if the project is going in the right direction and if it is not then what are the actions that need to be taken so that intended expectations can be fulfilled successfully? Tasks like reviewing the project objectives, comparing the project outcome with the intended goal, identifying the shortcomings of the project and devising a plan to overcome them, and determining the relevance of the project concerning real-world scenarios are really important to achieve the proposed goal of the phase and to understand the ground reality of the project. “Review Project Objective” and “Relevance to real word” are two tasks that are assigned priority as high in this phase because they serve as the foundational step to begin the evaluation or assessment of the project. This phase is planned to be completed in the same sprint as phase 4 because the number of days required to perform tasks in it is less. All the members are assigned tasks in this phase to ensure that the workload is evenly distributed. Since the modeling phase is still in the in-progress stage, the evaluation will come after that and hence it will be in the to-do stage.

**Figure 6**

*Gantt Chart for Evaluation Phase*

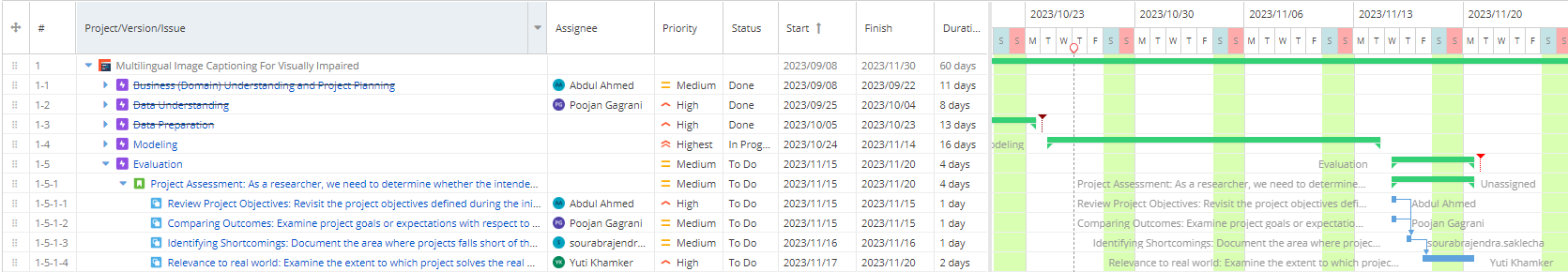
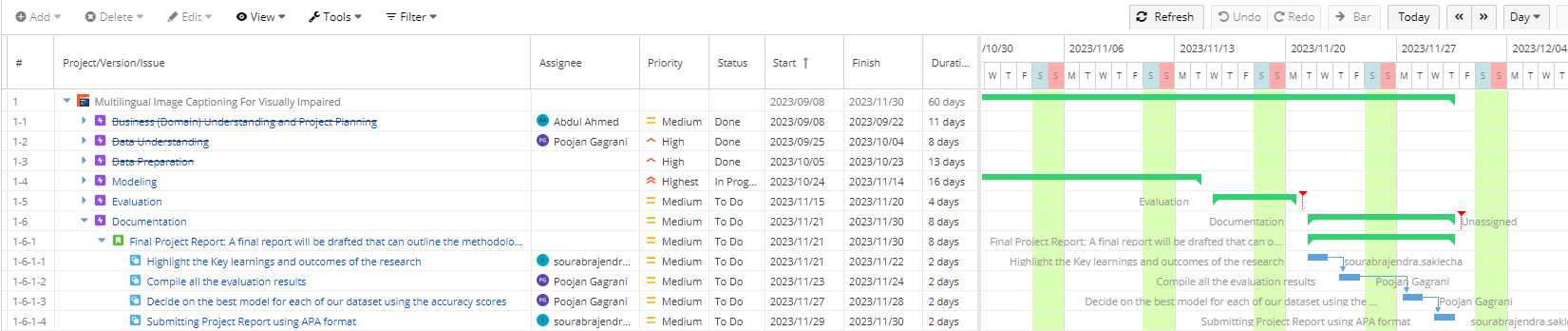


Figure 7 provided below manifests phase 6 of the project which is documentation. The goal of this phase is to focus on the documentation part of the project which is to register all the efforts done by the team in the final report. This phase includes tasks such as highlighting key learning and outcomes of the project, a compilation of all the evaluation scores obtained from different models on different metrics concerning previously mentioned datasets, making decisions about the model to be selected as the best model based on the accuracy score, and finally preparing the report that contains all the information starting from the first phase till the last phase in APA format. It will outline all the different actions performed and what are the outcomes obtained from them. This report will aid in communicating the results obtained in the modeling and evaluation phases and it is often paired with the presentation that summarizes the important events, insights, and findings of the project. With the submission of the project report and completion of the presentation, the last phase of the project will come to an end, and with this, the project will also be completed. If the compilation is done for the number of hours each team member is working then it comes out almost the same for all the members, this is done as mentioned above to ensure equal distribution of workload and to ensure effective usage of the resources at hand. This phase is still in the to-do stage as the current phase in which the project is in the modeling phase.

**Figure 7**

*Gantt Chart for Documentation Phase*



***PERT Chart***

Project Evaluation and Review Technique (PERT) chart (Bagshaw, 2021) is a project management tool that is designed to outline and manage projects with high complexity. It is a pictorial representation of different tasks that are sequentially involved in the project, the dependencies between different tasks, their start and end date, and number of days required to finish that task. It represents the tasks as the nodes and the directed arrows from one task to another will be their dependencies, this provides the complete image of the workflow of the project and the sequence maintained while doing so. It is used to determine the critical path of the project and if some tasks could be avoided to reach the goal faster.

In this project, the PERT chart is created using the task-based approach to understand the tasks and identify the critical path using Critical Path Analysis (CPA). The critical path can be defined as the longest series of dependent activities or tasks that need to be completed within the minimum timeframe to complete the project. Tasks that are present on the critical path cannot have any float, that is if these tasks get delayed then it would lead to the extension of the duration of the project and would result in the shift of the expected deadline of providing deliverables. Tasks present along the critical path don’t have flexibility in their start and end dates but the non-critical path’s task does. They have a chance to get right back on track even if they miss some of the assigned deadlines.

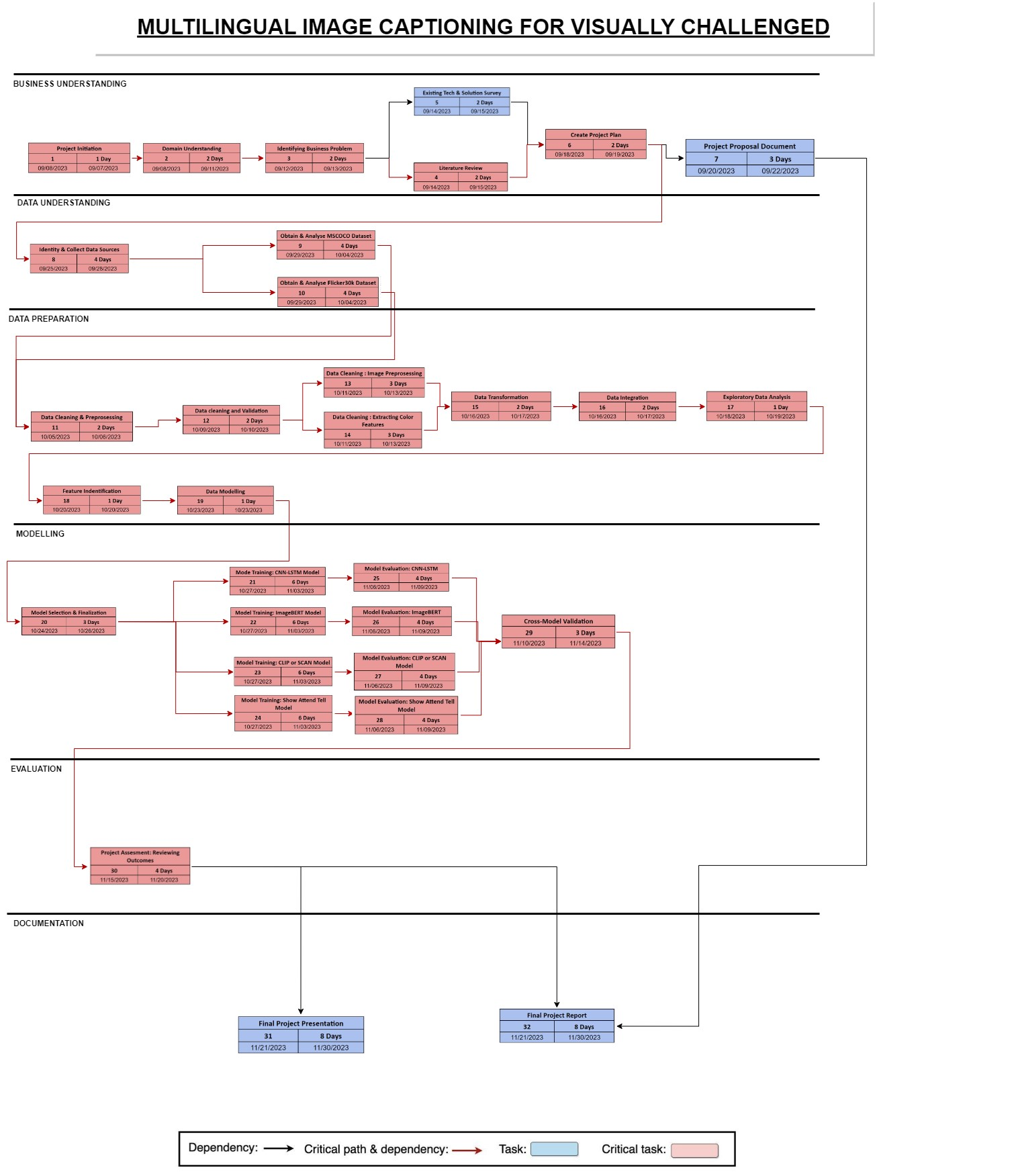
Figure 7 illustrates the PERT chart that was created for this project which includes the tasks that were performed or going to be performed, their dependencies with other tasks, and the critical path of the project. Tasks are represented in the form of nodes which are rectangular boxes that contain information about tasks like task name, task number, number of days required to accomplish that task, and the start and end date of the task. The dependencies between tasks are shown with the help of unidirectional arrows which show the next assigned task cannot start until the previous task is completed. Arrows are shown in two colors-black and red. The one with the red color shows that the path is part of the critical path and the one with black color shows that the path is not part of the critical path. The tasks that are along the black arrow path and are colored in blue are non-critical tasks, and tasks that are along the red arrow and are colored in pink are critical. Red-colored arrows that connect one critical task with another are used to depict the flow of the critical path in the project and the dependencies present in them whereas the black-colored arrows are just used to indicate the dependencies between non-critical tasks.

For the multilingual and color-focused image captioning project, determining the critical path is a very crucial step because it helps to keep proper track of the tasks needed to be completed within the allocated time frame in any sense also known as critical tasks to achieve the intended goal efficiently. Critical tasks began with initiating the project followed by the understanding of the domain of image captioning, identifying the relevant problem in this domain that needed to be resolved and carrying out an extensive literature review of existing papers on different methodologies to capture captions of the image, extraction of color features and conversion of English captions to multiple languages; creating project plan; determining and collecting data from relevant data sources like MS-COCO, Flickr30k etc.

Performing data cleaning, pre-processing, and transformation to obtain the relevant data, data validation to check if the data is accurate or not, exploratory data analysis to extract meaningful insights from the various visualization that were created, and performing feature engineering to determine the final set of features that will be used for modeling; selecting appropriate models, training them on data extracted from two above mentioned datasets and evaluating it against various evaluation metrics to determine which model performs best and meet the needs of the projects are some of the tasks determined by carrying out critical path analysis of the project. Some of the critical tasks like model training of CNN-LSTM, Show-Attend-Tell, ImageBERT, and CLIP models, and evaluation of models are mentioned concurrently in the PERT chart, which states that these tasks can be performed parallelly to increase the efficiency of the project. Each team member can take the task of training a particular selected model with the training datasets and evaluating it against various metrics and this set of tasks can be done concurrently by the team members.

**Figure 7**

*PERT Chart of Multilingual & Color-Focused Image Captioning For Visually Impaired Using Deep Learning Techniques*

****

**References**

Bagshaw, K. B. (2021). NEW PERT and CPM in Project Management with Practical Examples. *American Journal of Operations Research*, *11*(04), 215–226. <https://doi.org/10.4236/ajor.2021.114013>

Chen, J. (2023). Transform, Contrast and Tell: Coherent Entity-Aware Multi-Image Captioning. *In Proceedings of the 2023 International Conference on Computer Vision and Pattern Recognition*.<https://doi.org/10.48550/arXiv.2302.02124>

Fang, A., Ilharco, G., Wortsman, M., Wan, Y., Shankar, V., Dave, A., & Schmidt, L. (2022). Data determines distributional robustness in contrastive language image pre-training (clip). *In Proceedings of the 2022 International Conference on Machine Learning*, 6216-6234. PMLR. <https://doi.org/10.48550/arXiv.2205.01397>

Hossain, M. Z., Sohel, F., Shiratuddin, M. F., & Laga, H. (2019). A comprehensive survey of deep learning for image captioning. *ACM Computing Surveys* (CsUR), *51*(6), 1-36. <https://doi.org/10.1145/3295748>

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *In Proceedings of 2014 European Conference on Computer Vision*, 740–755. Springer. <https://doi.org/10.48550/arXiv.1405.0312>

Nakamura, T., Fukami, K., Hasegawa, K., Nabae, Y., & Fukagata, K. (2021). Convolutional neural network and long short-term memory based reduced order surrogate for minimal turbulent channel flow. *Physics of Fluids*, *33*(2). <https://doi.org/10.1063/5.0039845>

Papineni, K., Roukos, S., Ward, T., & Zhu, W.J. (2002). Bleu: A method for automatic evaluation of machine translation. *In Proceedings of the 40th annual meeting on Association for Computational Linguistics*, 311–318. Association for Computational Linguistics.<https://doi.org/10.3115/1073083.1073135>

Qi, D., Su, L., Song, J., Cui, E., Bharti, T., & Sacheti, A. (2020). Imagebert: Cross-modal pre-training with large-scale weak-supervised image-text data. *In Proceedings of 2020 International Conference on Computer Vision and Pattern Recognition.*<https://doi.org/10.48550/arXiv.2001.07966>

Wilson, J. (2003), Gantt charts: a centenary appreciation. *European Journal of Operational Research*, *149*(2), 430-437. <https://doi.org/10.1016/S0377-2217(02)00769-5>

Young, P., Lai, A., Hodosh, M., & Hockenmaier, J. (2014). From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, *2*(1), 67–78. <https://doi.org/10.1162/tacl_a_00166>