Tartu University

Faculty of Science and Technology

Institute of Technology

Yagub Hajiyev

Image context analysis for use in social media

Bachelor's thesis (12 EAP)
Science and technology

Supervisor:

PhD Egils Avots

Tartu 2024

Resümee/Abstract

Pildivaate analüüs kasutamiseks sotsiaalmeedias

Meie digitaalses maailmas on osa jagatud pilte valesti mõistetud ja sellised juhtumid võivad negatiivselt mõjutada inimese elu ja vaimset tervist. Seetõttu oleks enne pildi jagamist kasulik teada, kas pilt on kavandatud kasutuse jaoks sobiv. Selle probleemi lahendamiseks keskendub see lõputöö piltide konteksti mõistmisele sotsiaalmeedias. Pildi konteksti kirjeldatakse kahe meetodiga: pildisildistamine ja pildipealkirjastamine. Seejärel kasutatakse suurt keelemudelit, et mõista, kas pilt sobib isiklikuks, sotsiaalseks või äriliseks kasutamiseks. Nii on inimene rohkem teadlik sotsiaalmeedias jagatavatest piltidest.

CERCS: T111 pildindus, pilditöötlus, T120 süsteemitehnika, arvutitehnoloogia

Märksõnad: piltide märgistamine, piltide pealdised, LLM, teksti klassifikatsioon, sotsiaalmeedia

Image context analysis for use in social media

In our digital world some of the shared images have been misunderstood and such events can negatively affect person's life and mental health. Therefore, before a person shares an image, it would be beneficial to know if the image is appropriate for the intended use. To address this problem, this thesis aims at understanding the context of the images in social media. The context of the image is described with two methods: Image tagging and Image captioning. Afterwards, a large language model is used to understand if the image is appropriate for personal, social or business use. And in this way, the person will be more aware of images shared in social media.

CERCS: T111 Imaging, image processing, T120 Systems engineering, computer technology

Keywords: image tagging, image captioning, LLM, text classification, social media

Contents

List of Figures

List of Tables

Abbreviations, constants, generic terms

AI - Artificial Intelligence

API - Application Programming Interface

BERT - Bidirectional Encoder Representations from Transformers

BLIP - Bootstrapping Language-Image Pre-training

CARP - Contextualized Approach for Reasoning Paths

CNN - Convolutional Neural Network

CV - Computer Vision

FOMO - Fear of missing out

LLM - Large Language Models

ML - Machine Learning

NLP - Natural Language Process

OE - Others' Eyes

PaLM - Pathways Language Model

RNN - Recurrent Neural Network

RPN - Region Proposal Network

SEO - Search Engine Optimization

BLIP - Bootstrapping Language-Image Pre-training

1 Introduction

In our digital world, understanding how we're perceived by others is crucial. An image analysis startup called Others' Eyes OÜ seeks to improve online interactions by offering safe and fair comments on our digital content. The goal of this thesis is to provide image analysis tools that support this goal by testing the relevance and correctness of image tagging and interpretation. This thesis contributes to the project's overall goal of reducing misunderstandings in the digital sphere by improving image understanding.

1.1 Problem Statement

Understanding images on social media can be tricky. When we look at an image, there's a lot to consider: what objects are there, what they look like, and how they're arranged. But it's even more important to grasp what the image conveys. Sometimes, what we think an image shows and what it's actually trying to communicate can be quite different. This misalignment can lead to confusion or misunderstandings, affecting how others perceive us and our messages online.

In addition, when people share images, they often want to convey a specific message or feeling [1]. However, if the context of the image isn't clear, viewers may interpret it in unexpected ways. This can influence personal and professional relationships, especially if the intended message is lost or distorted. This project tackles these issues by focusing on how images are analyzed and understood. The goal is to ensure that images on social media are interpreted in the way the sharer intends, reducing errors and enhancing the clarity of digital communication. This will help in building a community that values understanding and empathetic interactions, aligning with the mission of Otherseyes.com to foster genuine connections.

1.2 Goals

This thesis aims to refine the process of tagging and captioning images to understand their contextual relevance. It involves critically assessing existing methods, pinpointing their limitations, and proposing innovative solutions. A significant objective is to create a tool that improves how the context of images is understood on social media, ensuring they convey the user's desired message.

By integrating theoretical research with practical applications, this project not only advances academic knowledge but also improves practical usability in everyday social interactions. Additionally, by examining the impact of misinterpretations on social media on self-esteem, this work highlights the importance of correct image representation for maintaining authentic online and offline connections. The study will conclude with an in-depth evaluation of the proposed methods, assessing their effectiveness and exploring potential future advancements.

2 Social Media and Mental Health

2.1 Social Media Awareness

A website or software that lets its users generate and share content or engage in social networking via a computer or personal electronic device is referred to as social media. Social media websites include LinkedIn, Snapchat, Instagram, X (Twitter), Facebook and others. People can upload private images and personal information on dating websites and blogs, which are open to the public and may appear to anyone searching for them (media platform logos shown in Figure 2.1). [2]

Being mindful of who you are speaking with and what you are posting on social media is an advantageous routine. This information is often publicly accessible whether you publish a comment, status update, image, or like and follow a page on the internet. Once someone else has shared it, it can be difficult to take control of or delete. Employers frequently check job candidates' social media accounts to see what they post, then use what they learn about them to generate opinions about their professionalism, character, and personality.

The largest risk associated with social media use is probably oversharing, and there are several ways in which this might go wrong. Identity thieves are hunting for easy targets, scammers are seeking for susceptible people to con, hackers are searching for information they can use to access personal accounts, and potential burglars are keeping an eye out for people who are going on vacation. There are countless more instances, but the fundamental problem remains the same: too much information is accessible to individuals who shouldn't be.

The increasing amount of social media use has also brought attention to the problem of cyberbullying. People have shown themselves to be highly comfortable saying things on the internet that they would probably never say in person. Frequently, individuals who post without fully understanding the circumstances of a certain situation increase the issue. Being mindful of your online behavior is essential if you want to prevent being a victim or making things worse. [3]

In a time when information is disseminated at previously unseen rate, being aware of social media serves as a defense against the spread of dangerous and false content. People can contribute to a healthy online environment by making intelligent choices by remaining knowledgeable about the digital world. [4]

False information spreads quickly on social media, which can cause confusion and even real-world problems. Recognizing the possibility of misleading information is the first line of protection against it. Knowing who you are "friends" with on social media is also crucial because you will be sharing information with these individuals. You have the option to reject friends or follow requests that you receive from strangers. [1]

Figure 2.1: Popular social media platforms [5]

As previously discussed, once something is hared on social media, it becomes publicly available and there is very little control over what happens to it. It's possible that someone else shared or copied your post or photo, so even after you remove it, people may still see it through a search.

The problems with social media will change as technology does. To stay ahead, one must be willing to use these platforms properly and maintain a constant state of adaptation and education. Keep in mind that your posts are helping to establish your brand. Personal viewpoints can be expressed on social media with great success, but exercise caution. When choosing what to write, it's a good idea to plan ahead and choose your words with the knowledge that the message will eventually be seen by the public. Prospective employers undoubtedly review candidates' social media posts, but existing companies might have a policy on social media and may review as well to keep an eye out for policy violations. [1],[3]

2.2 Social Media Awareness

People are social creatures. To thrive in life, we require the company of other people, and the quality of our relationships greatly affects our happiness and mental health. Keeping up social ties with people can lower stress, anxiety, and depression, boost self-esteem, provide solace and happiness, prevent loneliness, and even lengthen life. On the other hand, your mental and emotional health may suffer greatly if you don't have many intimate social connections. Social networking is by its very nature supportive. By producing dopamine, a "feel-good chemical" associated with enjoyable activities like dating, eating, and social interaction, using it stimulates the reward center of the brain. These platforms are linked to physical illnesses, anxiety, and depression and are made with the intention of being addictive. [6]

For many people living with mental illness, social networking has become an indispensable part of daily life. In a broad sense, social media refers to online and mobile platforms that let people interact with one another in a virtual network and share, co-create, or trade different kinds of digital content, such as messages, images, videos, and information.

Social media platforms like Facebook, Instagram, Snapchat, YouTube, TikTok, and X (formerly Twitter) are used by a large number of people in today's culture to communicate with one another. Though each has benefits, it's important to remember that social media can never completely replace face-to-face communication. The only time your relaxing and mood-enhancing

hormones come into play is when you are physically interacting with other people. [7] Despite popular belief, social media is intended to bring together people. On the other hand, overindulging in its use can exacerbate mental health conditions including depression and anxiety and heightened feelings of loneliness and isolation. Excessive use has been linked to adolescent loneliness, difficulties focusing, and trouble falling asleep. Sleep is also essential for the healthy growth of teenagers. [7]

In an effort to feel more accepted by their social circles and more secure in themselves, people divulge information in the hopes of receiving positive feedback.[8] Individuals regularly ask questions like "Did I receive as many likes as someone else?" when comparing their own social media activity to that of others, and "Why did this person (group) like my post, but other people didn't like my post?". Instead of forming genuine connections with people in real life, they are turning to the internet to find approval.

2.3 Social Media Disadvantages and Its Impact On Health

Fear of missing out (FOMO) might cause you to constantly check social media. FOMO will have you believing that there are more things that can wait or don't require an answer right away, even though there aren't many. Maybe you're concerned that if you don't keep up with the latest stories or news on social media, you won't be included in discussions at work or school. Or maybe you worry that your connections will suffer if you don't immediately like, share, or comment on other people's posts? As an alternative, you can worry that you won't get invited or that other people are having a better time than you. [9]

Social media is often used as a "security blanket" by humans. We reach for our phones and open social media if we're feeling uncomfortable, lonely, or nervous in a social environment. Naturally, using social media to communicate with others simply takes away from the in-person interactions that can reduce anxiety. [7]

It's possible that other underlying issues like stress, despair, or boredom are being covered up by your excessive social media use. If you use social media more often in times of depression, loneliness, or boredom, you may be using it as a coping mechanism or to divert your focus away from uncomfortable feelings. Finding better techniques to control your moods can be facilitated by allowing yourself to feel, even though it can be challenging at first. [7]

3 Image Content Assessment

Digital image context understanding involves tasks such as object detection, image classification, semantic segmentation, scene understanding, image tagging, and captioning. These tasks enable machines to analyze images comprehensively, identifying individual elements and understanding their relationships and the broader context.

3.1 Image Tagging

At its most basic, image tagging is the act of giving photographs meaningful or instructive labels. Consider it as a helpful assistant for managing and organizing your electronic files, similar to how properly labeling each spice on your kitchen rack makes it easier to locate the perfect zest for a recipe at glance. [10]

To make that happen, apply image tagging, which allows images to be categorized using labels and tags (see Figure 3.1). This makes it possible to properly categorize visuals in databases and to search and identify photographs quickly. Understanding the contents of their visual content is crucial for both individuals and enterprises. This is the method by which individuals and organizations can filter through the huge number of photographs that are continuously produced and shared online and use them appropriately.

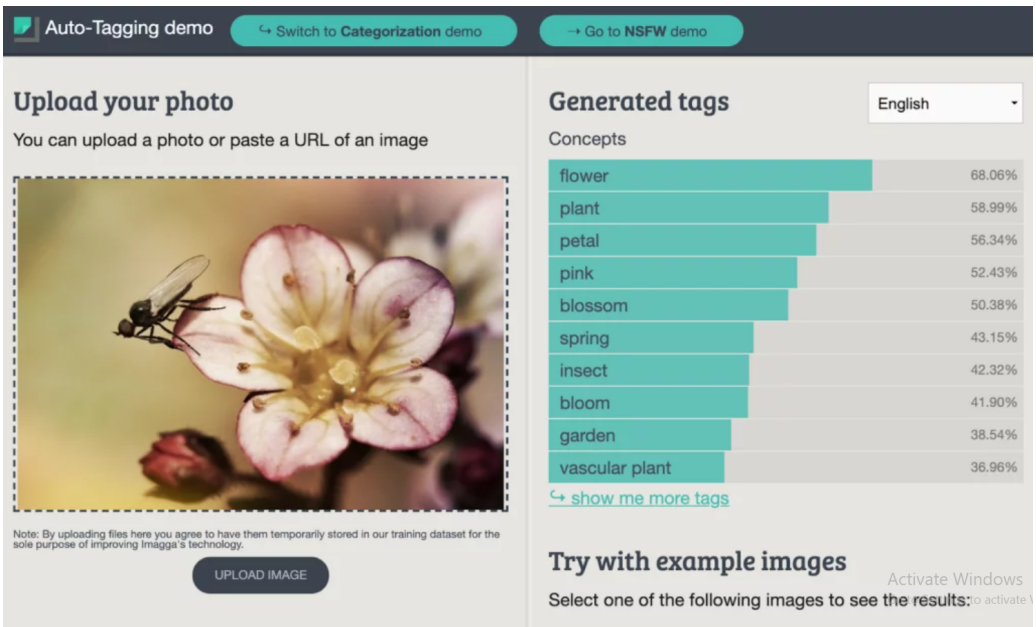

Figure 3.1: Image tagging example [11]

In the new areas of machine learning (ML) and artificial intelligence (AI), it attracts attention. These algorithms use pictures with tags as their main ”learning material” to identify patterns and to understand the characteristics of the images. This useful application advances facial and object recognition research and is even crucial to the development of autonomous car technology. [10] Adopting image tagging has several advantages for people and companies equally. Image tagging is incredibly helpful whether you're organizing your digital photo collection or building an image-based machine learning model.

3.1.1 Image Tagging Applications

Today, photo tagging is crucial for many different types of digital enterprises. To manage their visual assets, e-commerce, stock photo databases, reservation and travel platforms, traditional and social media, and a host of other businesses require efficient image sorting systems.

People can also benefit from image tagging. Without user-friendly image classification and keyword discoverability, organizing and searching through personal photo archives is time-consuming, if not impossible. [12] Example of common tagging applications:

- Improved Accessibility and Organization: Similar to finding a book in a well-cataloged library, finding an image in a large digital library is made much easier with the right tags.
- Improved Search Engine Results: Appropriate image tags can aid in enhancing a web page's Search Engine Optimization (SEO), which will increase the material's discoverability via the internet for marketers and content producers. [10]
- Optimization of Machine Learning: Tagged photos can be used as learning datasets by AI and ML developers to improve image recognition algorithms. [13]
- Increased User Engagement: Image tags can improve user experience for digital platforms by making it easier for users to find similar information.

3.1.2 Automatic vs Manual Image Tagging

Conventional image labeling needs input from humans. This was possible when working with fewer photos, despite the fact that it took longer. [12] The photos might be ”tagged” by hand, which would entail a person going through the system one by one, looking at each picture, and giving it a set of tags. Although the method is effective with small-picture datasets, it immediately breaks down when applied to large collections. The process of manually labeling photos takes a lot of time and work. You have to look at each image separately in order to manually tag it. They then have to manually enter the relevant keywords, which are often selected from a list of pre-approved concepts. If more keywords are needed, they may usually be added. [11]

Making sure there is consistency and avoiding errors in the tagging process would be the second, more serious problem. Based on their linguistic and conceptual knowledge of the image, anyone may participate in its tagging. Despite being entirely true, they might not align with operational requirements. Consequently, the business would have to invest a great deal of time and energy in training a sizable workforce to use a consistent tagging language. Additionally, human error in spelling could occur, rendering tagging useless as you wouldn't be able to find an image using that tag. Image tagging has become possible thanks to automated algorithms. Compared

to human picture tagging, automated picture tagging is faster and more precise. It also does a fantastic job of classifying data and conducting cross-category searches for particular pieces of information. Instead of using a human editor to process photos, an automatic system tags images. The machine vision skills will automatically assign relevant keywords and tags based on the results. [13]

Artificial intelligence (AI) image tagging can be finished in minutes as opposed to hours or days thanks to software programs that can tag hundreds of photos simultaneously. To further reduce human inconsistencies, the same AI algorithm would reliably return the same tags for the same image. It is possible to create more efficient corporate procedures in less time by hiring and onboarding new staff. Real-time and consistent modifications can be made to photos with AI-powered tagging. [13]

3.1.3 Image Tagging Using CNNs and RNNs

An automatic visual system becomes necessary in order to automatically describe images using natural language. It has been widely accepted as a key topic of research to create precise computer systems approaches to automatically expect a textual description of the visual semantics of a given image. Understanding the relationships between labels becomes essential as image tagging works with several labels. The development of practically applicable visual recognition systems has advanced significantly with the introduction of deep learning (deep neural networks). Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are two popular methods for this. [14]

For regular grid-based data, like images, CNNs have demonstrated a compelling ability to learn powerful features; yet, RNNs are found to be more effective in learning combinatorial inter-label relationships in textual data while maintaining a realistic computational complexity. Consequently, combining CNN with RNN offers a viable way to discover a sophisticated computational model for image annotation.

The tagging uses a two-step process to take advantage of the inherent sequence-learning and sequence-prediction capabilities of RNNs. Given an image representation computed using a CNN, the first step allows the RNN to predict a sequence of confidence scores corresponding to all the labels, and the second step uses these confidence scores for each label across the sequence and assigns the maximum score (see Figure 3.2). [14]

3.2 Image Captioning

One of the main areas of artificial intelligence is called ”deep learning”. Just as the human brain uses a million neurons to accomplish various sensing methods, we attempt to replicate ”human brains” in our machines through the use of deep learning. As the name implies, image captioning is just that-image captioning. It entails giving an object’s description and receiving back the text captions for the same things (see Figure 3.3). [15]

Image captioning is the process of composing a written description for an image. It has shown to be an important and vital work in the deep learning sector. There are thousands of applications for image captioning. NVIDIA is developing an application to assist those with impaired vision utilizing image captioning capabilities. [17]

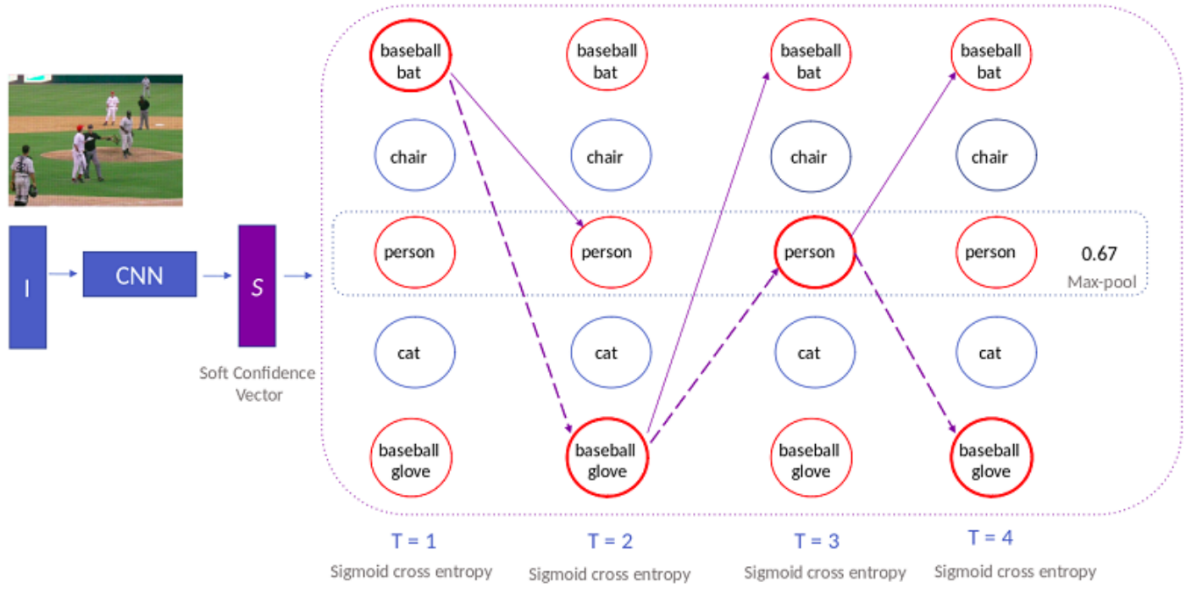

Figure 3.2: Image tagging and learning of multiple inter-label relationships [14]

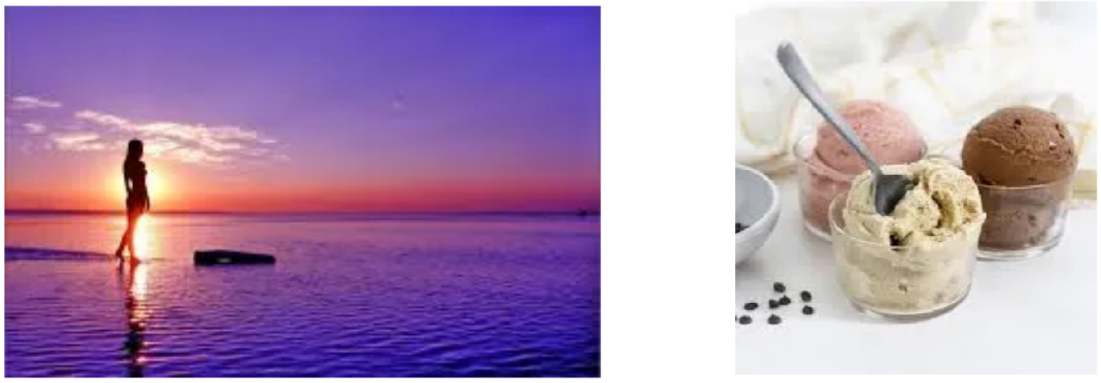

Figure 3.3: Image captioning and its example with COCO dataset [16]

The process for generating captions involves computer vision and certain language processing methods. A dataset of photos and the matching output text captions are first needed in order to obtain the necessary captions. This model uses a recurrent neural network (RNN) for sentence synthesis and a convolutional neural network (CNN) for feature extraction. [15]

While language processing has led to more accurate caption production, computer vision has helped to improve image captioning systems. Applications for image captioning include social media, image indexing, editing software, virtual assistants, assistive technology for the blind, and many more natural language processing uses.

The act of automatically producing descriptive text from an image's visual data is known as image captioning. It is an approach that blends aspects of computer vision (CV), natural language processing (NLP), and artificial intelligence. In the past few years, a great deal of research has been conducted with impressive outcomes to produce image captions. [15]

Research on deep learning-based image captioning is categorized into three areas for a comprehensive overview: training approaches, optimisation, and the network architecture. After compiling the evaluation metrics, the state-of-the-art methods are compared using the MS COCO dataset or other public datasets. [16]

3.2.1 Image Captioning with CNNs

Standard image captioning architecture encodes the input into a fixed form and decodes it word by word into a sequence. The input image with three color channels is encoded by the encoder into a smaller image with ”learned” channels. A condensed depiction of everything meaningful in the original image can be found in this smaller encoded image. You can use any CNN architecture for encoding. Additionally, transfer learning is an option for the encoder portion. The encoded image is viewed by the decoder, which creates a caption word by word. Next, the next word is generated using each of the predicted words. The connection between image captioning tasks and object recognition is established by using faster R-CNN. The Region proposal model makes use of cross-domain knowledge and is pre-trained on object detection datasets. Furthermore, both models employ one-pass attention with the Up-Down mechanism, in contrast to certain other attention mechanisms. Image feature extraction is done with a faster R-CNN. The Faster R-CNN object detection model locates items using bounding boxes and recognizes things that fall into specific classifications (see Figure 3.4). [18]

R-CNN is faster and detects objects in two steps. The first step anticipates object proposals and is referred to as a Region Proposal Network (RPN). The top box proposals are chosen as input for the second stage using greedy non-maximum suppression with an intersection-over-union (IoU) threshold. In the next stage, each box proposal is given a small feature map (such as 14 × 14) that is extracted using region of interest (RoI) pooling. After that, these feature maps are combined and fed into the last few layers of the CNN. Consequently, each box proposal’s class-specific bounding box revisions and a softmax distribution over class labels comprise the final model output. The official poster serves as the source of the scheme. [18]

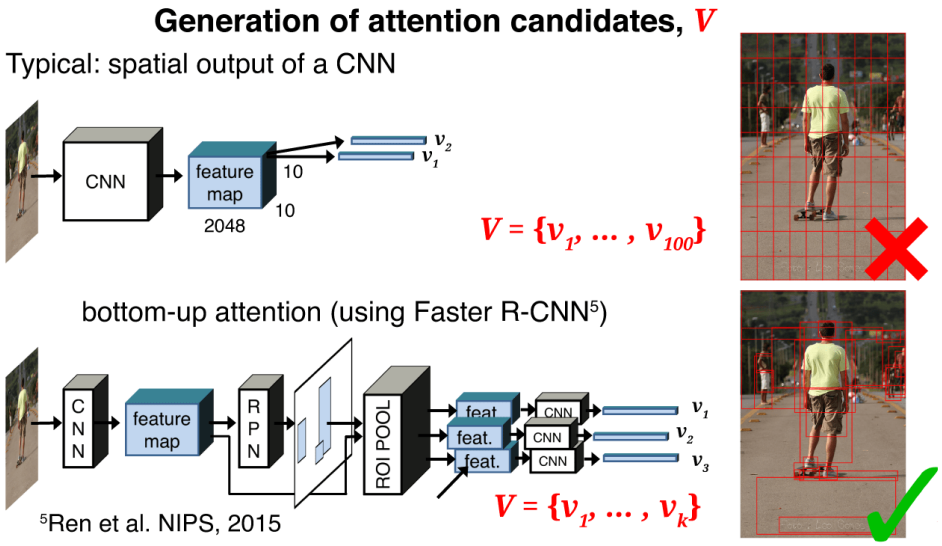

Figure 3.4: Example of encoder model [19]

4 Text Classification

Fundamentally, text classification aims to comprehend and classify the huge amount of text data that is produced on a daily basis. Text classification assists in the large-scale categorization and understanding of textual content, whether it is used to classify emails as spam or not-spam, determine the tone of a product review, or arrange news articles according to their subject. [20]

A machine learning technique called text classification is used to automatically classify unstructured text into many predefined groups. Text classifiers are capable of organizing, structuring, and classifying nearly any kind of text, including online material as well as text from publications, scientific studies, and customer tickets. As subsets of artificial intelligence (AI), machine learning (ML) and natural language processing (NLP) are two of the most interesting new technologies to emerge in recent years. These technologies are capable of text classification, which is the intelligent grouping of text according to sentiment. [21]

One important NLP task that helps with many different business difficulties is text classification. Emails, messages, help requests, and other data management-related issues are at the heart of many of these worries. Businesses can also obtain insightful information that aids in their decision-making. One can categorize new articles based on topics, support tickets based on urgency, chat conversations based on language, brand mentions based on sentiment, and so on. Text classification is one of the primary issues in natural language processing.

4.1 Introduction to Large-Language Models

A large language model (LLM) is a kind of artificial intelligence (AI) computer that can recognize and produce text. Large refers to the fact that massive data sets are used for LLM training LLMs are based on machine learning, namely on a type of neural network called a transformer model. [22]

Alternatively stated, an LLM is a computer program fed enough examples to recognize and understand complex material, such as human language. Many LLMs are trained using hundreds or even millions of megabytes of material from the Internet. However, as the quality of the samples influences how well the LLMs learn real language, their programmers may decide to utilize a more carefully chosen data set.

Deep learning is a subset of machine learning that LLMs use to learn how words, phrases, and characters work together. Through the probabilistic examination of unstructured data, deep learning eventually allows the model to identify differences between content without the need for human interaction. After that, LLMs receive additional training through tuning, which include prompt-tuning or fine-tuning to the specific activity that the programmer wants them to perform, such as translating text or interpreting questions and producing answers. [22]

A deep learning system most likely cannot learn anything from a single line of text. But after going over trillions of sentences, it might learn enough to be able to construct sentences on its own or even predict how to logically end a sentence that isn't quite finished.

The main LLM model applications are:

- Text generation: the ability to create language in response to cues by producing emails, blog posts, or other mid-to-long form content that can be polished and refined. One such example is retrieval-augmented generation, or RAG. [23]
- Content summarization: condense lengthy articles, news items, research studies, company records, and even client biographies into comprehensive texts with lengths appropriate for the output format.
- AI assistants: AI assistants are chatbots that work as part of an integrated, self-serve customer care system to respond to consumer requests, handle backend duties, and deliver comprehensive information in natural language. [23]
- Code generation: this tool assists programmers in writing code, finds faults in the code, and even "translates" between other programming languages to find security holes.
- Sentiment analysis: examine text to ascertain the tone of the consumer in order to comprehend large-scale customer feedback and support brand reputation management.
- Language translation: with accurate translations and multilingual capabilities, language translation gives enterprises greater reach beyond national and regional boundaries. [23]

4.2 Popular Large Language Models

Large language models that are popular have swept the globe. Many have been embraced by individuals in various industries. I'm sure you've heard about ChatGPT, a type of chatbot powered by generative AI (see Figure 4.1). The most well known models are [24]:

- PaLM: Google's Pathways Language Model (PaLM) [25] is a transformer language model that can perform basic and mathematical reasoning, translate, write code, and explain jokes.
- BERT: The Bidirectional Encoder Representations from Transformers (BERT) [26] language model was also created by Google. It is a transformer-based model with question-answering capabilities that can comprehend simple sentences. [27]
- XLNet: Unlike BERT, XLNet [28] is a permutation language model that produces output predictions in a random order. Rather than predicting tokens in a sequential sequence, it predicts tokens in a random order after evaluating the encoded pattern.
- GPT: The most well-known big language models are presumably generative pre-trained transformers. The widely used GPT fundamental model [29], created by OpenAI, has improved with each of its numbered versions (GPT-3, GPT-4, etc.). It can be adjusted to carry out particular functions later on. For example, two instances of this include Bloomberg's BloombergGPT for finance and Salesforce's EinsteinGPT for customer relationship management.

Other less popular examples (but used internationally) are Cohere, Galactica and Lambda.

Figure 4.1: Parameters of transformer-based language models [27]

4.3 Transformers

Transformer models are the name given to the particular kind of neural networks employed in LLMs. The ability of transformer models to learn context is crucial since human language is heavily context-dependent. Transformer models identify subtle relationships between items in a sequence by applying a mathematical term known as self-attention. Compared to other forms of machine learning, they are therefore more adept at comprehending context. [22]

Transformers can handle input data in parallel, in contrast to earlier deep learning models for natural language processing, like RNNs and Long short-term memory (LSTMs), which process data sequentially. This enhances the model's capacity to comprehend language context in addition to increasing its efficiency. [30] The most important concepts in this architecture are:

- Embedding Layer: This process turns words, or input tokens, into vectors - numerical representations of the words plus the text's context and semantic meaning.
- Positional Encoding: In this stage, the embedding of each token in the sequence is enhanced with information about its position. This is done to make up for the Transformer's lack of sequential processing.
- Encoder and Decoder: The encoder parses text input and extracts context, while the decoder predicts words in a sequence to produce coherent replies.
- Self-Attention Mechanism: This mechanism enables the model to assess the relative weights of various words within the input sequence. Because a word's meaning can vary depending on its context inside a phrase in natural language, this allows it to comprehend context and relationships between words throughout the whole input sequence. [30]
- Feedforward neural networks: The feedforward neural networks in the encoder and decoder are in charge of incorporating extra changes into the data that the self-attention mechanism processes.

In addition, the model blocks employ layer normalization and residual connections to improve training stability, prevent the vanishing gradient issue, and aid in the development of deeper neural networks (see Figure 4.2).

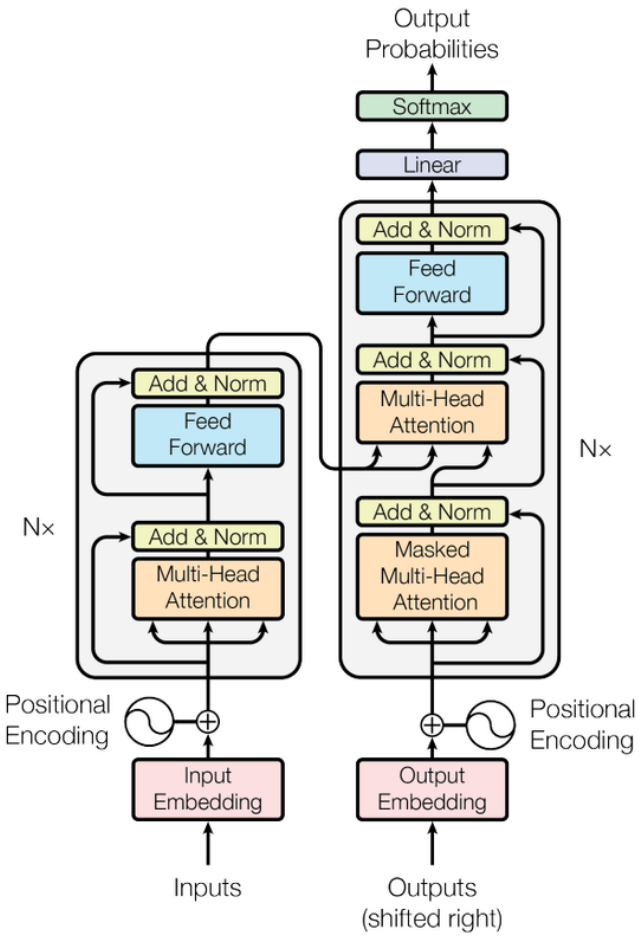

Figure 4.2: Transformer model architecture [30]

4.4 Advantages of ML Based Text Classification

For example, the phrase ”The user interface is very simple to use and clear.” can be fed into a text classifier, which will automatically identify relevant tags like UI and Easy to Use after analyzing the text. For automated text processing systems the following advantages emerge:

- The scalability: Manual analysis and organisation is extremely precise and takes much longer. Millions of surveys, comments, emails, and other types of data may be automatically analyzed using machine learning at a fraction of the cost, frequently in a matter of minutes. Tools for text classification can be scaled to meet any size business need. [31]

- Real-time analysis: Businesses sometimes need to recognize dangerous situations quickly and take appropriate action (e.g., PR disasters on social media). With machine learning classification of texts, you can track mentions of your brand continuously and in real time, allowing to quickly detect important information and take appropriate action. [31]

- Consistent criteria: Due to human subjectivity and distractions, tiredness, and boredom, human annotators make mistakes when classifying text data, leading to inconsistent standards. In contrast, machine learning employs uniform standards and criteria for all inputs and outputs. A text categorization model operates with unparalleled precision once it has been properly trained. [31]

4.5 Text Classification with Large Language Models

There are two methods for classifying texts: automatically and manually. Manual text classification involves the assessment of the text's content by a human observer who then assigns the appropriate category. Although this method can yield extremely good results, it is very expensive and time-consuming. In order to classify text more rapidly, effectively, and precisely, automatic text categorization uses machine learning, natural language processing, and other AI-based techniques.

There are important advantages to integrating LLMs into classification of text operations.

- Adaptability: Whether by fine-tuning using task-specific data or by varying prompts in zero-shot and few-shot learning, LLMs can adapt to a broad range of classification tasks with little effort.
- Efficiency with Less Data: Especially in few-shot settings, LLMs can achieve impressive accuracy with substantially less data, unlike standard models that need large labeled datasets to function well.
- Advanced Reasoning: Newer methods that include explicit reasoning phases in the classification process, such as CARP (Contextualized Approach for Reasoning Paths), may improve interpretability and resilience. [32]

The most widely used methods for categorizing texts are Zero-Shot and Few-Shot:

- Zero-Shot Classification: asking a model directly, without providing any examples, what the label means. It's the easiest way, doesn't require any data, but performance is rather low, and you might get a result that isn't in your fixed class list (hallucination).
  - Pre-LLMs: Making use of Task-Aware Representation of Sentences (TARS) and other open-source models.
  - Post-LLMs: Sending a completed structure and asking LLMs to create a label directly. Compared to pre-LLMs, this method is slower but far more accurate. [32]
- Few-Shot Classification: only a little quantity of annotated data is needed and only a few examples are supplied per class.
  - Pre-LLMs: Making use of TARS and other open-source models.
  - Post-LLMs: Applying LLMs by passing the examples from each class into the prompt's context. Will be more precise than the earlier method. [32]

As LLMs develop further, they have the potential to completely transform the classification of texts by providing solutions that are not only more flexible and data-efficient but also able to include reasoning in previously unfeasible ways. However, in order to fully realize their potential, difficulties with cost, bias, and interpretability need to be addressed. By doing this, it is possible to unleash new possibilities across a variety of NLP applications by utilizing the capacity of LLMs to navigate the complexity of language. [33]

5 Methodology

The aim of this project is to test if pre-trained LLMs (ChatGPT) can be used for imaging context categorisation based on image captions and tags. The role of LLM is to decide if the image belongs to one of the following categories: personal, social, professional.

The code for this thesis can be found in the following repository and in the Appendix:
`https://github.com/yagub03/YagubThesis/tree/main`

5.1 Content categories

The main objective of the context analysis tool is to understand in which category an image belongs to. It is assumed, that image has only one category that is the most appropriate. Categories are defined as follows:

1. Personal: Images that are primarily intended for private enjoyment, personal memories, or intimate moments. These photos are often shared with close friends and family or kept for personal reflection. Examples:
   (a) Family gatherings and events
   (b) Vacations and travel photos
   (c) Selfies taken at home or in private settings
   (d) Photos of pets or personal hobbies
   (e) Childhood photos and personal milestones
   (f) Private moments and experiences
2. Social: Images that are meant for sharing on social media platforms or with a wider group of friends and acquaintances. These photos often capture social events, gatherings, or moments meant to be shared publicly or within a community. Examples:
   (a) Casual outings and hangouts
   (b) Parties, concerts, and social events
   (c) Group photos with friends
   (d) Images from community events or public gatherings
   (e) Photos with friends or in social settings
   (f) Trendy or popular locations and activities
   (g) Highlights of your daily life and adventures
   (h) Humorous or entertaining images

3. Professional: Images that are related to your career, professional activities, or intended for use in a professional context. These photos often have a formal tone and are used for work-related purposes. Examples:

   (a) Professional headshots or profile pictures

   (b) Photos taken at conferences, seminars, or work events

   (c) Images of work projects or achievements

   (d) Pictures used in marketing materials or professional portfolios

   (e) Photos taken for company websites or LinkedIn profiles

5.2 Dataset

According to Others Eyes app development requirements, the user has the option to select one of three categories for image analysis. To understand better, if the image is appropriate for the selected category, a dataset was created and labeled by the author. The Others Eyes app currently has 3 categories: personal, social and professional. Therefore a balanced dataset with 100 images in each category was created, with total of 300 images. Few example images from the dataset are show in Figure 5.1.

(a) Personal (b) Social (c) Professional

Figure 5.1: Examples from the dataset

5.3 Research Tools

For my thesis topic, I mostly used use my personal laptop (MSI, CPU i7-11800H 2.30GHz, RAM 16.0 GB) but is some cases, a more powerful PCs with GPU was used. In the developed code we used Python (opencv-python == 4.9.0.80), CUDA 12.2 and cuDNN 8.x. The code also uses BLIP captioning and WD14-tagging, where the textual image description is given to GPT-3.5 turbo and GPT-4 for image context analysis using OpenAI API.

BLIP (Bootstrapping Language-Image Pre-training) is a cutting-edge model in image captioning, combining computer vision and natural language processing. It uses a transformer-based

architecture pre-trained on large datasets of image-caption pairs, allowing it to generate accurate and contextually relevant descriptions for images. [34]

WD14-Tagger is a sophisticated tool designed for automated image tagging, leveraging advanced machine learning techniques to assign relevant tags to images. Built on a deep learning framework, WD14-Tagger is trained on extensive datasets, enabling it to recognize and label a wide variety of objects, scenes, and concepts within images. [35]

5.4 Image context classification

To process all the images for one category the total runtime is around 250-300 seconds, it means one image takes nearly 2.5-3 seconds runtime. After the category is determined, all context results are going to the specified output path as .txt file.

Figure 5.2: Example image with professional context

Image context is classified according to the following steps for image show in Figure 5.2:

1. Select and load input image
2. Obtain three image captions with BLIP:
   (a) A room filled with people and sitting at desks.
   (b) Some students are at their computers in a very big room.
   (c) There are many people sitting at desks using computers.
3. Obtain image tags with WD14-tagger: indoors, multiple boys, 6+boys, multiple girls, desk, chair, camera, computer, sitting, speaker, clock, monitor, window, office.
4. Classify image based using LLM prompt using OpenAI API: ”Classify the following image description into one of these categories: Personal, Social, Professional. Choose the most appropriate category strictly from these three options. Category: {description} Category (Personal, Social, Professional)”
5. Obtain LLM output using OpenAI API:
   - GPT-3.5-turbo: Category: Professional
   - GPT-4 : Category: Professional

6 The Results and Discussion

As previously noted, classification results were obtained using GPT-3.5-turbo and GPT-4. After running the code with these models, it was determined that GPT-4 is more accurate overall. While GPT-4 generally produces reliable outcomes, GPT-3.5-turbo tends to have more errors, particularly in social contexts. Specifically, GPT-3.5-turbo is more accurate with "Personal" categories, whereas GPT-4 performs better with "Professional" categories. According to Table 6.1 and 6.2, it was determined that ChatGPT-3.5 has 59% and ChatGPT-4 has 82% accuracy.

According to Table 6.1, many of the images were misclassified as 'Personal', even tough such

Table 6.1: Classification accuracy results using ChatGPT-3.5

		Predicted class			
		Personal	Social	Profess.	
Actual class	Personal	74	19	7	74.00
	Social	41	46	13	46.00
	Profess.	34	9	57	57.00
		49.66	62.16	74.03	**59.00**

situation is not desirable. From a perspective of app usage and safe advice, it is more advantageous for professional image to be misclassified as personal than the other way around.

The results from Table 6.2 show that a more advanced LLM performs better, even though the

Table 6.2: Classification accuracy results using ChatGPT-4o

		Predicted class			
		Personal	Social	Profess.	
Actual class	Personal	79	19	2	79.00
	Social	4	83	13	83.00
	Profess.	1	14	85	85.00
		94.05	71.55	85.00	**82.33**

captions, tags, and prompts are the same. It can also be seen that misclassifying the personal category as social has the most frequent errors, with only a few instances of misclassifying the personal category as processional. Similarly, the misclassification of the professional category as social occurs more frequently than that of the personal category, with rare instances of misclassifying personal images as professional and vice versa.

The accuracy could be slightly improved with more fine tuned prompt, but the biggest limitation is the input information. While describing the image with captions and tags, valuable information such as emotions and social cues are lost. Currently, combination of various ML models that describe the image context and can be used as a alternative to a multi-modal system with image analysis capabilities. The demonstrated approach can also be used as an alternative when analysing images that do not comply with the multi-modal LLM service provide polices.

Further research will focus on expanding the image dataset to include more challenging images and more categories, such as dating. It would also be beneficial to add other image classification models to context analysis, like age estimation, gender recognition, emotion recognition, and object segmentation, to provide the used LLM with more information.

7 Conclusion

As part of the Othrers' Eyes project, the thesis focuses on using image captioning and tagging methods to interpret the provided image, followed by the classification of the image into personal, social, or professional contexts using pre-trained LLMs. The used approach achieved 59% accuracy with ChatGPT-3.5 and 82% accuracy with ChatGPT-4. These results demonstrate that this approach has the potential to be used as a feature in the Othrers' Eyes application. Such context category suggestions, particularly on social media, would help the users reduce misinterpretations and unfavorable impressions.

The Others Eye project will continue, with future work focusing on comparing text-based classification to multi-modal models with vision capabilities. The main focus will be on the cost of processing an image and exploring the limitations of LLM provider policies regarding image context analysis. In addition to the currently used context categories, the number of categories will increase in the future.

Acknowledgements

I would like to extend my heartfelt thanks to the University of Tartu's Science and Technology program for providing me with an excellent educational environment and numerous opportunities for growth.

I am especially grateful to Assoc. Prof. Ilona Faustova, the Vice Director of the Institute of Bioengineering, for her support and guidance throughout my academic journey. Additionally, I wish to express my gratitude to my thesis supervisor, Egils Avots, for his insights, expertise, and encouragement.

Lastly, I am profoundly thankful to my family and friends for their unwavering support, understanding, and encouragement during my studies. Their belief in me has been a constant source of motivation.

Thank you all for your contributions to my academic journey.

Bibliography

[1] S. U. of New York at Oswego. "Social media awareness." Accessed: 2024-05-22. (2024), [Online]. Available: https://www.oswego.edu/cts/social-media-awareness.

[2] J. E. Trust. "Social media awareness." Accessed: 2024-05-22. (2024), [Online]. Available: https://www.jettraining.co.je/training/it-training/social-media-awareness/.

[3] A. Mohamed. "Social media awareness: Navigating the digital landscape." Accessed: 2024-05-22. (2023), [Online]. Available: https://www.aimtechnologies.co/social-media-awareness-navigating-the-digital-landscape/.

[4] K. M. G. of Schools. "A comprehensive guide to social media awareness for students." Accessed: 2024-05-22. (2023), [Online]. Available: https://www.krmangalambahadurgarh.com/blogs/a-comprehensive-guide-to-social-media-awareness-for-students/.

[5] P. Media. "How is professional social media management done?" Accessed: 2024-05-22. (2023), [Online]. Available: https://pointgorsel.com/en/2023/01/12/how-is-professional-social-media-management-done/.

[6] K. Katella. "How social media affects your teen's mental health: A parent's guide." Accessed: 2024-05-22. (2024), [Online]. Available: https://www.yalemedicine.org/news/social-media-teen-mental-health-a-parents-guide.

[7] M. Hospital. "The social dilemma: Social media and your mental health." Accessed: 2024-05-22. (2024), [Online]. Available: https://www.mcleanhospital.org/essential/it-or-not-social-medias-affecting-your-mental-health.

[8] L. Robinson and M. Melinda Smith. "Social media and mental health are you addicted to social media?" Accessed: 2024-05-22. (2024), [Online]. Available: https://www.helpguide.org/articles/mental-health/social-media-and-mental-health.htm.

[9] H. Bashir and S. A. Bhat, "Effects of social media on mental health: A review," *International Journal of Indian Psychology*, vol. 4, no. 3, pp. 125–131, 2017.

[10] Cloudinary. "Image tagging." Accessed: 2024-05-22. (2023), [Online]. Available: https://cloudinary.com/glossary/image-tagging.

[11] H. Yilmaz. "Image tagging: What is it & how it works." Accessed: 2024-05-22. (2024), [Online]. Available: https://imagga.com/blog/image-tagging/.

[12] H. Yilmaz. "Image tagging: What is it & how it works." Accessed: 2024-05-22. (2024), [Online]. Available: https://www.plugger.ai/blog/image-tagging-what-is-it-how-it-works.

[13] isahit. "What is image tagging?" Accessed: 2024-05-22. (2022), [Online]. Available: https://www.isahit.com/blog/what-is-image-tagging.

[14] Y. Verma. "Neural networks for automated image tagging." Accessed: 2024-05-22. (2020), [Online]. Available: https://iitj.ac.in/techscape/vol01/issue01/issue01/neural-networks-for-automated-image-tagging/Neural%20networks%20for%20automated%20image%20tagging.pdf.

[15] P. Gera. "What is image captioning in computer vision?" Accessed: 2024-05-22. (2022), [Online]. Available: https://codedamn.com/news/machine-learning/what-is-image-captioning.

[16] H. Yilmaz. "Image captioning." Accessed: 2024-05-22. (2024), [Online]. Available: https://medium.com/@pmegne/image-captioning-5162e22ef2ac.

[17] A. Roy. "A guide to image captioning." Accessed: 2024-05-22. (2020), [Online]. Available: https://towardsdatascience.com/a-guide-to-image-captioning-e9fd5517f350.

[18] M. Bhikadiya. "Automatic image captioning using deep learning." Accessed: 2024-05-22. (2020), [Online]. Available: https://medium.com/swlh/automatic-image-captioning-using-deep-learning-5e899c127387.

[19] MobiDev. "Deep learning image captioning technology for business applications." Accessed: 2024-05-22. (2021), [Online]. Available: https://www.iotforall.com/deep-learning-image-captioning-technology-for-business-applications.

[20] P. Maia. "Classifying text using llms." Accessed: 2024-05-22. (2023), [Online]. Available: https://nilg.ai/202308/classifying-text-using-llms/.

[21] X. Sun, X. Li, J. Li, *et al.*, "Text classification via large language models," *arXiv preprint arXiv:2305.08377*, 2023.

[22] Cloudflare. "What is a large language model (llm)?" Accessed: 2024-05-22. (2024), [Online]. Available: https://www.cloudflare.com/learning/ai/what-is-large-language-model/.

[23] Elastic. "What is a large language model (llm)?" Accessed: 2024-05-22. (2023), [Online]. Available: https://www.elastic.co/what-is/large-language-models.

[24] G. Yenduri, G. Srivastava, P. K. R. Maddikunta, *et al.*, "Generative pre-trained transformer: A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions," *arXiv preprint arXiv:2305.10435*, 2023.

[25] S. Narang and A. Chowdhery, "Pathways language model (palm): Scaling to 540 billion parameters for breakthrough performance," *Google AI Blog*, 2022.

[26] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[27] S. M. Kerner. "What are large language models." Accessed: 2024-05-22. (2023), [Online]. Available: https://www.techtarget.com/whatis/definition/large-language-model-LLM.

[28] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *Advances in neural information processing systems*, vol. 32, 2019.

[29] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[30] P. Foy. "Understanding transformers & the architecture of llms." Accessed: 2024-05-22. (2024), [Online]. Available: https://www.mlq.ai/llm-transformer-architecture/.

[31] MonkeyLearn. "Text classification." Accessed: 2024-05-22. (2024), [Online]. Available: https://monkeylearn.com/text-classification/.

[32] S. Kumar. "Exploring text classification with large language models: A deeper dive." Accessed: 2024-05-22. (2024), [Online]. Available: https://www.linkedin.com/pulse/exploring-text-classification-large-language-models-dive-mba-ms-phd-lxr4c/.

[33] P. Orza. "Text classification: What it is & how to get started." Accessed: 2024-05-22. (2022), [Online]. Available: https://levity.ai/blog/text-classification.

[34] J. Li, D. Li, C. Xiong, and S. Hoi, "Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation," in *International conference on machine learning*, PMLR, 2022, pp. 12 888–12 900.

[35] "Stable diffusion webui wd14 tagger." Accessed: 2024-05-22. (2022), [Online]. Available: https://github.com/toriato/stable-diffusion-webui-wd14-tagger.

Appendices

Appendix 1. Requirements

```
pandas==2.2.2
huggingface-hub==0.22.2
Pillow==10.0.1
deepdanbooru==1.0.2
onnxruntime==1.17.3
opencv-python==4.9.0.80
```

Appendix 2. Functions

```
from typing import Generator, Iterable
from tagger.interrogator import Interrogator
from PIL import Image
from pathlib import Path
import argparse

from tagger.interrogators import interrogators

parser = argparse.ArgumentParser()

group = parser.add_mutually_exclusive_group(required=True)
group.add_argument('--dir', help='Predictions for all images in the directory')
group.add_argument('--file', help='Predictions for one file')

parser.add_argument(
    '--threshold',
    type=float,
    default=0.35,
    help='Prediction threshold (default is 0.35)')
parser.add_argument(
    '--ext',
    default='.txt',
    help='Extension to add to caption file in case of dir option (default is .txt)')
parser.add_argument(
    '--overwrite',
    action='store_true',
    help='Overwrite caption file if it exists')
parser.add_argument(
    '--cpu',
    action='store_true',
    help='Use CPU only')
parser.add_argument(
    '--rawtag',
    action='store_true',
    help='Use the raw output of the model')
parser.add_argument(
    '--recursive',
    action='store_true',
    help='Enable recursive file search')
parser.add_argument(
    '--exclude-tag',
    dest='exclude_tags',
    action='append',
    metavar='t1,t2,t3',
    help='Specify tags to exclude (Need comma-separated list)')
parser.add_argument(
    '--model',
    default='wd14-convnextv2.v1',
    choices=list(interrogators.keys()),
    help='modelname to use for prediction (default is wd14-convnextv2.v1)')
args = parser.parse_args()

# get interrogator configs
```

```
interrogator = interrogators[args.model]

if args.cpu:
    interrogator.use_cpu()

def parse_exclude_tags() -> set[str]:
    if args.exclude_tags is None:
        return set()

    tags = []
    for str in args.exclude_tags:
        for tag in str.split(','):
            tags.append(tag.strip())

    # reverse escape (nai tag to danbooru tag)
    reverse_escaped_tags = []
    for tag in tags:
        tag = tag.replace(' ', '_').replace('\(', '(').replace('\)', ')')
        reverse_escaped_tags.append(tag)
    return set([*tags, *reverse_escaped_tags])  # reduce duplicates

def image_interrogate(image_path: Path, tag_escape: bool, exclude_tags: Iterable[str]) -> dict[str, float]:
    """
    Predictions from a image path
    """
    im = Image.open(image_path)
    result = interrogator.interrogate(im)

    return Interrogator.postprocess_tags(
        result[1],
        threshold=args.threshold,
        escape_tag=tag_escape,
        replace_underscore=tag_escape,
        exclude_tags=exclude_tags)

def explore_image_files(folder_path: Path) -> Generator[Path, None, None]:
    """
    Explore files by folder path
    """
    for path in folder_path.iterdir():
        if path.is_file() and path.suffix in ['.png', '.jpg', '.jpeg', '.webp']:
            yield path
        elif args.recursive and path.is_dir():
            yield from explore_image_files(path)

if args.dir:
    root_path = Path(args.dir)
    all_tags = []
    for image_path in explore_image_files(root_path):
        caption_path = image_path.parent / f'{image_path.stem}{args.ext}'

        if caption_path.is_file() and not args.overwrite:
            # skip if caption exists
            print('skip:', image_path)
            continue

        print('processing:', image_path)
        tags = image_interrogate(image_path, not args.rawtag, parse_exclude_tags())

        tags_str = ', '.join(tags.keys())
        all_tags.append(f"{image_path.name}: {tags_str}")

    # Write all tags to a single file
    with open(root_path / 'all_tags.txt', 'w') as fp:
        fp.write("\n".join(all_tags))

if args.file:
    tags = image_interrogate(Path(args.file), not args.rawtag, parse_exclude_tags())
    tags_str = ', '.join(tags.keys())
    print(tags_str)
```

Appendix 3. Caption generation

```
import os
from PIL import Image
import torch
from lavis.models import load_model_and_preprocess

# Setup device to use
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

# Load the model and preprocessors
model, vis_processors, _ = load_model_and_preprocess(
    name="blip_caption", model_type="large_coco", is_eval=True, device=device
)
# Path to the folder containing images
image_folder = "---"

# List all image files in the folder
image_files = [f for f in os.listdir(image_folder) if f.endswith(('jpg', 'jpeg', 'png'))]

# Path for the output text file
output_file = os.path.join(image_folder, "all_captions.txt")
```

```
# Open the output file in write mode
with open(output_file, "w") as f:
    # Loop through each image file
    for image_file in image_files:
        image_path = os.path.join(image_folder, image_file)
        raw_image = Image.open(image_path).convert("RGB")

        # Process the image
        image = vis_processors["eval"](raw_image).unsqueeze(0).to(device)

        # Generate captions
        captions = model.generate({"image": image}, use_nucleus_sampling=True, num_captions=3)

        # Write captions to the output file
        f.write(f"{image_file}:\n")
        for idx, caption in enumerate(captions):
            f.write(f"  {idx + 1}: {caption}\n")
        f.write("\n")

print(f"All captions saved to {output_file}")
```

Appendix 4. Classification using LLM

```
import openai
import os

# Set your OpenAI API key
openai.api_key = "api-key"

# Function to read tags and captions from text files
def read_file(file_path):
    with open(file_path, 'r') as file:
        return file.readlines()

# Read tags and captions
tags = read_file('---')
captions = read_file('---')

# Combine tags and captions into a dictionary
image_data = {}

# Parse tags
for tag_line in tags:
    if ": " in tag_line:
        img_name, img_tags = tag_line.split(": ", 1)
        img_tags = img_tags.strip()
        image_data[img_name] = {"tags": img_tags}

# Parse captions
current_image = None
for caption_line in captions:
    caption_line = caption_line.strip()
    if caption_line.endswith(".jpg:"):
        current_image = caption_line.split(":")[0]
        if current_image in image_data:
            image_data[current_image]["captions"] = []
        else:
            image_data[current_image] = {"captions": []}
    elif current_image and (caption_line.startswith("1:") or caption_line.startswith("2:") or caption_line.startswith("3:")):
        img_caption = caption_line.split(": ", 1)[1]
        image_data[current_image]["captions"].append(img_caption)

# Function to classify context using OpenAI's GPT-4
def classify_context(description):
    prompt = f"""
    Classify the following image description into one of these categories: Personal, Social, Professional.
    Choose the most appropriate category strictly from these three options.

    Description: {description}
    Category (Personal, Social, Professional):
    """

    for _ in range(3):  # Retry up to 3 times if the response is not valid
        response = openai.ChatCompletion.create(
            model="gpt-3.5-turbo",  # Specify GPT-4 model
            messages=[
                {"role": "system", "content": "You are a helpful assistant."},
                {"role": "user", "content": prompt}
            ],
            max_tokens=10,
            temperature=0
        )
        category = response.choices[0].message['content'].strip()
        if category in ["Personal", "Social", "Professional"]:
            return category

    # If all retries fail, return a default category or handle the failure
    return "Personal"

# Analyze and classify each image
results = {}
for img_name, data in image_data.items():
```

```
    if "captions" in data and data["captions"]:
        description = f"Tags: {data['tags']}\nCaptions: {' '.join(data['captions'])}"
        category = classify_context(description)
        results[img_name] = category
    else:
        # If no captions are present, use tags for classification
        description = f"Tags: {data['tags']}"
        category = classify_context(description)
        results[img_name] = category

# Write results to a text file
output_file_path = '---'
with open(output_file_path, 'w') as output_file:
    for img_name, category in results.items():
        output_file.write(f"Image: {img_name}, Category: {category}\n")

print(f"Results have been written to {output_file_path}")
```

Non-exclusive licence to reproduce thesis and make thesis public

I, Yagub Hajiyev

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   "Image context analysis for use in social media"

   supervised by Egils Avots

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Yagub Hajiyev
22.05.2024