A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

Advancements in AI-Enhanced Design Inspiration and Personalization for Web-Based Applications

Retrieving User Interaction and Personalizing Content through Recommender Systems

Leonor Rodrigues Pereira - 48778

Work project carried out under the supervision of:

Rongjiao Ji (*Nova SBE*)

Fransisco Tenente (*3DWays*)

20/12/2023

Abstract (100 words maximum)

Designers and product creators can struggle to express their vision in the early stages of a project, resourcing to abstract ideas. A Machine Learning platform to aid authors in getting visuals and gather user preferences is in demand. The tool presented helps in this process, by giving users graphic representation of their ideas, which are entered via keywords. The app is aware of users' inputs and learns as it cycles through the process. Using reputable models from Open AI, Meta and in-house built recommender systems, a Streamlit application was created to compile a package of useful functionalities to creators.

Keywords: Product Development, Machine learning, Web Application, Image Suggestions, Word Suggestions, Recommender Systems, Mood Board, Text Mining

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Table of Contents

1 Introduction (group part)

1.1 Background

In a panorama where solutions exist for nearly every conceivable issue, Artificial Intelligence (AI) has played a significant role, with algorithms delivering time-saving results or even autonomously executing tasks beyond their initial training scope. However, there is still a big gap in the integrations AI can perform. Even most comprehensive models often fall short in combining multiple functionalities, which may involve different code realms, to form a complete tool tailored specifically to an organization's objectives (Cox 2023). The work produced in this thesis encompasses the creation of a unique solution addressing a creative and inspiration challenge identified by 3DWays while providing their services to customers.

3DWays is a Portuguese start-up specialized in product design, dedicated to assisting investors in industrializing new products and getting them to the market. Their services range from conceptualizing a new product to its final implementation and usage, aiming to be a “One-Stop Hardware Partner” (3DWays 2023a). The company’s customers include other companies working in one of the sectors from product design to manufacture, as well as individual inventors looking to industrialize their creations.

3DWays mission, defined in the company’s LinkedIn page:

“We perform user and market research, eco-design, prototyping, supply chain management and contract manufacturing. Through our partner network, we design business models and pricing strategies, we help get products certified and patented and we help them raise funds and hardware-related legal advisory.” (3DWays 2023b)

Such a complete service package is a big help for other startups in avoiding risks associated with manufacturing hardware, since there is more time to refine the product’s specifications that will differentiate itself from the rest of the market.

Besides operating at the prototyping and manufacturing levels, with the assistance of an advanced 3D printing facility, 3DWays is also specialized in mentoring companies designing products.

1.2 Objectives

The requirements stage is essential in any type of project. Companies like 3DWays should outline major milestones and agree with the client on the most important characteristics of the final product. Numerous challenges encountered during consulting jobs are found to be caused by weak planning in an early stage. To mitigate these issues, companies demand a novice tool to first help clients discover and define their ideas, and secondly help them visualize and express those designs.

For product design, this implies offering a visual survey for insights collection, instead of traditional survey with limited and inaccurate description by words. The web application developed helps users express and define the product concept, just like it fosters the elaboration of new ideas about that concept. A human brain can only create based on what it has seen before (Greisdorf and O'Connor 2002). Therefore, showing additional content related to users' searches, via word and image recommender systems, promotes the generation of new ideas and boosts creativity.

The product of this thesis will play a crucial role during the design requirements sessions phase, marking the initial interaction between 3DWays and any client. The combination of AI driven content generation with recommendation systems allows for practical and thorough understanding of user needs. Moreover, the final output, a mood board compiling user choices, provides a better context to each client's vision and expectations. Building upon the foundational work of Arabella Specker, conducted in a directed research internship at Nova SBE and analysed in the literature review chapter, this research is extended to consider more sophisticated models for each component of the tool (Specker 2022). Such models include pre-trained algorithms from tech giants Open AI and Meta, and purpose-built recommender systems.

1.3 Thesis structure

This report starts with a literature review (chapter 2) where current state of the art in AI-driven product design is studied, alongside the challenges businesses face integrating AI in their workflows, plus an analysis of the work done in the previous thesis.

The following four chapters (chapter 3 to 6) delve into the distinct components developed for the tool, discussing each part's role, design, and implementation in the context of the project's goals. Most in-depth analysis of the models used and how they were adapted for the project’s needs can be found here. A similar structure was used for all individual parts: Introduction, Literature Review, Methodology, Results, and Conclusion. Each chapter focuses on topics from the four main research areas: word recommender systems, image recommender systems, image modification models, and web application integration.

Chapter 7 contains a breakdown of the three main databases considered for the development of this project. Chapter 8 covers insights regarding the generation of the final mood board. Chapter 9 encloses tips and guidelines on how users and future developers should handle the web application.

Lastly, chapters 10, 11 and 12 are the results, discussion (limitations and future recommendations) and conclusion of the project, respectively.

1.4 Project flowchart – User journey

Figure 1 is a summary of the web applications’ capabilities. Each functionality’s inputs and outputs were coordinated to fit the rest of the pipeline. This overview provides a reference for the reader to contextualize the upcoming chapters of the report.

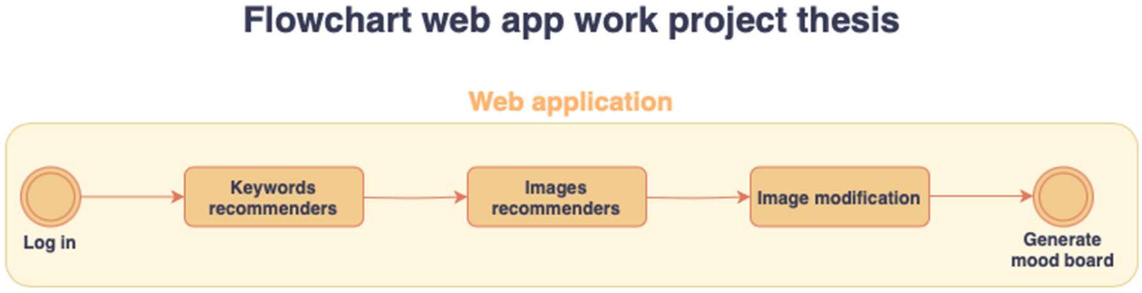

Figure 1 - Web Application flowchart

The chosen tool to merge all components was the library Streamlit, working as a web page generator for our tool. Users' first interaction with it is a login page, where a personal account will be created to store individual preferences and interactions (chapter 6.3.1).

After entering the app, the first prompts required are keywords. These are essential for the next steps involving images or perfecting each user's keywords list. Two word-recommender systems will be used, one using Open AI GPT model (chapter 3), and another designed to check previously used related keywords (chapter 6.3.2).

With the selection of keywords completed, two image-recommender systems operate to present images to the user. The first model matches the inserted words to images from a database (chapter 4), while the second one tries to find patterns with previous images selected by other users (chapter 6.3.3).

The final section of the web application is dedicated to modifying images selected by users, to better align with their vision. This includes options for object segmentation and object removal (chapter 5).

The relevant outputs are finally combined in a real-time generated mood board (chapter 8). This mood board represents the automated solution that 3DWays sought to efficiently capture and understand client requirements.

2 Literature Review and current state of the art in AI-driven product design (group part)

2.1 AI in Conceptual Design and Product Development

In the evolving landscape of digitalization, AI has emerged as a central force in transforming the design and development of new products. The integration of AI in the design process represents a significant shift from traditional methods, preparing the way for more developed, efficient, and innovative design systems (Rosenthal and Niggemann 2022).

Engineers and designers today face a challenging task of creating products that cater to the dynamic needs of millions of consumers. This demands a balance between design, engineering, manufacturing, and craftsmanship, all while following specific guidelines and expectations. AI technology has become instrumental in this regard, offering invaluable support in capturing the right data and guiding product design and development processes. A McKinsey survey from November 2020 highlights that over half of the organizations have adopted AI in at least one function, with many reporting a significant increase in revenue due to AI adoption, especially in manufacturing (McKinsey 2020).

The transformation of AI's role in product development has become apparent in recent years. AI, which was less prevalent in the field, has become a fundamental component. Major technology companies such as Google, IBM, and Amazon have been instrumental in advancing its use in various aspects of product development, including engineering. This shift is largely due to AI's ability to streamline complex processes, speed up the time to market for products, and encourage innovation in product design (Salierno, Leonardi and Giacomo Cabri 2021).

In the realm of conceptual design, especially for products not yet present in the market, AI offers considerable potential. AI-driven tools empower designers to quickly visualize and refine new product concepts, playing a key role in the early phases of product development. These tools seem to improve the efficiency of the design process while also providing new opportunities for creativity and innovation, enhancing the overall design experience.

2.2 Challenges in AI integration for businesses

The integration of AI has become more prevalent within businesses, however remains a challenge due to various reasons. Companies face multiple obstacles, including the need for technical expertise, resistance to organizational change, financial limitations, data management difficulties, and ethical and legal concerns. These issues, ranging from internal training gaps to regulatory compliance, impact every aspect of AI implementation in businesses. The first challenge is the demand for technical expertise. AI technologies are evolving rapidly, requiring a depth of technical knowledge that many businesses find challenging to acquire (Russell and Norvig 2016). The adaptation of generic AI solutions to meet specific business needs often necessitates a bespoke approach, tailoring algorithms, data sets, and models to align with operational difficulties (Fox 2023). This situation has been seen by Munich Re, where the company tackled the technical expertise gap through internal training and strategic external collaborations (Fox 2023).

Another critical challenge is the resistance to change within organizations. The integration of AI often disrupts established processes and roles, bringing out a range of responses from employees, from passive non-compliance to active resistance (Kotter 1996). Overcoming this resistance requires comprehensive change management strategies. These include extensive staff training which focuses on the operational aspects of AI while also focusing on the rationale and benefits of these changes. Requiring a culture that embraces change, backed by leadership commitment and open communication, is crucial in this regard (Lewin 1951; Fox 2023). Financial constraints represent a third challenge, particularly seen in small to medium sized enterprises (SMEs). The costs associated with AI integration especially for the initial setup, ongoing maintenance, necessary data storage and hiring specialized employees can be substantial. Addressing these financial challenges requires exploring various funding options such as government grants or venture capital, alongside strategic planning to ensure a favourable return on investment.

Data availability and quality is another obstacle. Effective AI applications depend heavily on the availability of high-quality data. Many businesses struggle with gathering sufficient data and concurrently ensuring its relevance and cleanliness (Davenport 2018). Addressing data quality issues is critical, as failure to do so can compromise the integrity and effectiveness of AI systems. Lastly, ethical and legal considerations present a complex terrain for businesses implementing AI. Issues such as algorithmic bias, data privacy, and compliance with evolving regulations laws like the GDPR in the European Union are at the forefront of these challenges (Bostrom and Yudkowsky 2014; ESPU 2018). An example is UNESCO's effort to minimize gender bias in AI search engines, highlighting the need for ethical audits and the development of frameworks for responsible AI use (Fox 2023).

2.3 Current state of the art – Building on Specker's work

Last year's paper, by Arabella Specker (2022), made progress with regards to integrating AI into the design process, specifically through the development of a mood board tool. This tool, designed to retrieve images from Pinterest using image similarity recommendations, incorporated web scraping, natural language processing, and the VGG16 convolutional neural network. Specker's approach to evaluating the tool's performance involved both qualitative assessments and quantitative measures, like cosine similarity metrics.

Specker's work underscores the potential of AI in enhancing design processes for industrial designers. The tool effectively provided relevant images for inspiration, allowing designers to create and store mood boards efficiently. However, the study also highlighted certain limitations. The tool's reliance on Pinterest as the sole image source restricted the diversity of available images. Moreover, while the average cosine similarity score indicated a moderate level of image relevancy, the potential need for more diverse and dissimilar images for creative inspiration was noted.

The thesis represented a significant step in using AI for industrial design, particularly in the creation and utilization of mood boards. This tool streamlined the initial stages of the design process by providing designers with a novel way to gather inspiration, thus facilitating a quicker transition from concept to development.

Despite these advancements, Specker's research also highlighted several limitations, which open up avenues for further development. The first notable aspect is that the tool lacks a user-friendly interface, which could be a barrier for designers unfamiliar with the underlying technology, making it less accessible and potentially reducing its effectiveness in the design process. Moreover, the tool's dependence on a limited set of input words, coupled with the inability to include additional descriptive terms like adjectives, limited the spectrum and diversity of accessible design inspirations. This limitation could impact the creativity and relevance of the mood boards produced. There was also an absence of a guidance system within the tool, leaving users without support in selecting the most effective keywords for their search. This could lead to less effective search results and potentially hinder the design inspiration process. Lastly, the tool did not allow users to modify or edit the images it presented. This lack of interactivity could limit designers' ability to tailor the mood board to specific design needs, potentially impacting the overall utility of the tool in the creative process.

Building upon Specker's foundational work, the current research focuses on enhancing the AI-driven design tool by addressing these limitations. Efforts include developing a more user-friendly interface to improve accessibility and ease of use. Expanding the range of input options will allow for more nuanced and tailored image results. Implementing a guidance system will assist users in effectively utilizing the tool for design inspiration. Moreover, incorporating features for image modification and editing will provide users with greater control over their mood boards, enhancing the tool's applicability in the creative design process.

7 Database and dataset overview (group part)

7.1 Open Images dataset V7

The selection of an appropriate dataset for advancing artificial intelligence models is crucial especially in tasks like image generation, captioning, and object selection. Numerous features play a role in the selection process and in this section, we will outline the characteristics of the open image V7.

7.1.1 Characteristics

Size: The Open Images dataset is extensive, incorporating precisely 9 million images across a diverse array of categories and contexts. This substantial size not only ensures an ample supply of data for training machine learning models but also aligns seamlessly with the company's requirement for a sizable dataset.

Accessibility: The Open Images V7 is publicly available and hosted by Google. This availability ensures that researchers, developers, and educators can easily download and integrate into their projects.

Data Structure: Open Images V7 is structured in multiple components catering to varied computer vision challenges (Ultralytics 2020). This structured dataset includes essential components such as:

- Bounding Boxes: With Over 16 million boxes that demarcate objects across 600 categories. These bounding boxes are made-up squares that serve as a guideline point for object recognition and create a collision box for that element (Tasq.ai 2023)
- Segmentation masks: With exact boundaries of 2.8M objects across 350 classes. These segmentation masks define the outline of objects, which characterizes their spatial extent to a much higher level of detail (Krasin et al 2017).

Labels and metadata: Each image in the dataset is annotated with detailed labels, providing valuable information about the objects and scenes depicted. This rich labelling facilitates various computer vision tasks, such as object recognition and segmentation. Additionally, the dataset includes metadata that offers various details about the images, enhancing its usability for diverse applications in machine learning and research.

Applications: As highlighted by Ultralytics (2020), Open-image-dataset 7 is a cornerstone for training and evaluating state-of-the-art models in various computer vision tasks.

Flexibility: The Open Images Dataset offers considerable flexibility in sample selection. This adaptability proves invaluable for testing and refining machine learning models.

To sum up, these characteristics collectively affirm the suitability of the Open Images dataset for our project, which involves the generation of images, creation of captions, and object selection.

7.1.2 Implementation

The integration of FiftyOne in downloading the Open Images dataset offers a comprehensive solution for computer vision developers. FiftyOne, as a python library and web application provide an interactive tool for data exploration, model evaluation, and iterative development. Additionally, the seamless integration of FiftyOne with the Open Images dataset streamlines tasks such as debugging, error analysis, and collaborative exploration, making it an invaluable resource for teams engaged in advanced computer vision projects.

In conclusion, this seamless implementation confirms it as the optimal choice for the development of our web application.

7.1.3 Database structure

After the implementation of the Open Image Dataset, the structure will be explained in more detail below.

Table 2. Open Images dataset structure

Name	Open-image-v7-validation-10
Media Type	Image
Num samples	50
Persistent	False
Tags	[]
Same fields	id: ObjectIdField
	filepath: StringField
	tags: ListField
	metadata: EmbeddedDocumentField (ImageMetadata)
	positive labels: EmbeddedDocumentField (Classifications)
	negative labels: EmbeddedDocumentField (Classifications)
	detections: EmbeddedDocumentField (Detections)
	segmentations: EmbeddedDocumentField (Detections)
	relationships: EmbeddedDocumentField (Detections)

Id (ObjectIdField): is a unique identifier for each sample in the dataset. This ObjectIdField is an auto-generated identifier that makes us differentiate between samples, allowing easy referencing during analysis and manipulation.

Filepath (StringField): contains the path to the image file associated with each sample. This information is important, especially when unable to access or load a special image for further processing.

Tags (List Field (StringField)): The tags field is a list that contains user-defined tags associated with each sample.

Metadata (EmbeddedDocumentField(ImageMetadata)): The metadata field is an embedded document containing metadata associated with the image. Image metadata include information such as Licence, Author, Original size, OriginalLandingURL, and any other relevant details that provide context about the image's origin and characteristics.

Positive_labels (EmbeddedDocumentField(Classifications)): The positive labels field is an embedded document that contains information about labels associated with the image. In the context of the Open Images dataset, it includes Label Name, Confidence, logits, and imageID.

Negative_labels (EmbeddedDocumentField(Classifications)): It is similar to positive_labels, However, the negative_labels field is an embedded document that contains information about negative classifications or labels. Negative labels represent the absence of certain labels in a specific image. It also includes: LabelName, Confidence,logits and imageID.

Detections (EmbeddedDocumentField(Detections)): The detections field is an embedded document that includes information about object detections within the image. In the Open image Dataset, it includes: ImageID, label,bounding_box, mask,index,Confidence, IsOccluded,IsTruncated, IsGroupOf,IsDepiction,IsInside

Segmentations (EmbeddedDocumentField(Detections)): The segmentations field, like detections, is an embedded document that contains information related to image segmentation like Mask Path, ImageID, attributes, label,bounding_box, confidence, index

relationships (EmbeddedDocumentField(Detections)): The relationships field is an embedded document that contains information about spatial or semantic relationships between objects within the image. This includes relationships like Imaged, Label1, Label2, bouding_box, index, confidence, label, mask, attributes, label, and tags.

To sum up, the Open Images Dataset's sample fields offer a wealth of information about every image, including file specifics, tags, metadata, object labels, detections, segmentations, and relationships. These fields can be used by academics and industry professionals to carry out different types of analysis, develop machine learning models, and acquire an understanding of the context and content of the images in the dataset.

7.2 Flickr 8K - Image Captioning dataset

7.2.1 Criteria for Dataset Selection in Image Captioning Research

The selection of an appropriate dataset is a critical step in the advancement of image captioning models within the field of artificial intelligence (Paullada et al. 2021). It forms the backbone for both training and evaluating the performance and robustness of these models. The criteria set forth for this selection are rooted in ensuring that the dataset not only serves the immediate requirements of model training and evaluation but also aligns with the broader research objectives and ethical standards of the field (Fenza et al. 2021).

Size and Diversity: A dataset of adequate size is fundamental to confine a wide range of image contexts and scenarios, facilitating comprehensive model training and testing. Diversity in imagery, including variations in scenes, objects, and activities, is crucial to assess the model's versatility and ability to generalize across different visual contexts.

Quality and Relevance of Captions: The chosen dataset must contain captions that are accurate in describing the image and also being relevant to the application of the model. High-quality captions are essential for using models and evaluating pre-trained models that can generate precise and contextually appropriate descriptions.

Balance and Representation: Ensuring a balanced representation of various subjects and themes within the dataset is vital to avoid biases in model training. A well-balanced dataset contributes to the development of a model that is fair and effective across diverse design contexts.

Accessibility and Ethical Considerations: The dataset's accessibility for academic research and compliance with ethical standards, including the appropriate licensing of images and captions, is a critical aspect of the selection process. This ensures the responsible use of data in line with academic and research integrity.

These criteria were instrumental in guiding the selection of a dataset that not only meets the technical requirements of the image captioning functionality but also aligns with the ethical and practical considerations of this thesis. In evaluating these criteria against available datasets, the Flickr 8K dataset published on Kaggle by Aditya (2020) appeared as a particularly suitable dataset. The following section will provide an overview of the Flickr 8K dataset, detailing its structure and fields, and discuss how it meets the outlined criteria.

7.2.2 Overview

The Flickr 8K dataset is widely recognized in the field of image captioning research for its suitability in developing and evaluating image captioning models, matching well with the established criteria. Flickr 8K is characterized by its manageable size, containing a total of 8,092 images. This moderate scale makes the dataset particularly suited for research scenarios where computational resources are limited, such as training models on low-end laptops or desktops. Additionally, the dataset is freely available, which aligns with the accessibility criterion, making it a popular choice for academic and research purposes.

Labeling and Data Structure: Each image in the Flickr 8K dataset has five distinct captions, providing textual context for each visual scene. This multiple-caption structure enhances the dataset's utility for training and evaluating image captioning models, as it offers diverse interpretations of the same visual content. The dataset is divided into specific subsets, with 6,000 images designated for training, 1,000 for testing, and 1,000 for development purposes. This clear segmentation facilitates the systematic training, testing, and validation of models.

The Flickr 8K dataset comes with the necessary textual files that describe the training and test sets. The Flickr8k.token.txt file, in particular, contains a comprehensive list of the 40,460 captions, offering a pool of data for model training and testing.

7.2.3 Significance and broader application of the dataset

Given its structure and features, the Flickr8K dataset fulfils the criteria set for selecting an image captioning dataset. The dataset's size and diversity make it an ideal choice for comprehensive model training without necessitating extensive computational resources. The quality and relevance of the captions in the dataset are conducive to developing models capable of generating accurate and contextually appropriate descriptions.

The multiple caption per image structure of Flickr 8K allows for a more nuanced evaluation of the model's performance, as the diversity in captions can offer insights into how well the model captures various aspects of an image. Furthermore, the straightforward and well-organized structure of the dataset simplifies the processes of data pre-processing and model training, making it a practical choice for research.

In the broader context of this thesis, the Flickr 8K dataset provides a solid foundation for cross-comparison and collaborative analysis. Its use in different individual components of the project, such as training distinct models or evaluating various image captioning techniques, enables a cohesive and comprehensive research approach. The dataset's versatility and alignment with the project's objectives underscore its significance as a valuable resource within this collective research.

In summary, the Flickr 8K dataset, with its manageable size, well-labelled data, and accessible format, presents itself as an optimal choice for our image captioning research. Its selection is justified not only by its technical and structural features but also by its alignment with the broader goals and ethical considerations of this thesis.

8 Mood board PDF (group part)

Mood boards have long been an essential tool in the creative process, serving as a visual representation of ideas, themes, and inspirations. Their origins can be traced back to the early 20th century, where they were used in fashion and interior design. Initially, these boards were physical collages of images, text, and samples of objects in a composition. Over time, their use has expanded across various fields including graphic design, film production, and even in digital media planning, illustrating their versatility and adaptability to different creative needs.

By their very nature, mood boards are designed to evoke a certain style or mood, providing a tangible way for designers to communicate their concepts and visions. They act as a springboard for creativity, allowing for a free-flowing exchange of ideas and serving as a reference point throughout the design process (Rieuf, Bouchard and Aoussat 2015).

In the context of this project, the mood board takes on a digital form, evolving from its traditional physical manifestation. The digital mood board, particularly in the form of a PDF, becomes a crucial component in the design cycle. It offers a unique way to capture and present the essence of a user's design journey, encapsulating their selections, preferences, and inspirations. For example, in the realm of web design, digital mood boards can be used to visually communicate the look and feel of a website before any code is written. Similarly, in fashion, they can collate various fabric textures, colour palettes, and accessory choices to form a coherent style vision. This digital format not only aligns with the modern, tech-driven approach of the design tool but also offers greater accessibility and ease of sharing, crucial in collaborative design environments (Chipambwa and Chikwanya 2022).

8.1 Requirements

The mood board PDF, as conceptualized within the scope of the project by 3DWays, was designed with specific requirements to ensure both functionality and aesthetic appeal.

Central to these requirements was the stipulation that the mood board be confined to a single page. This single-page constraint is pivotal as it compels the user to refine their creative vision into a concise, yet powerful visual narrative. The challenge is to focus on what is essential and most expressive of their project's theme. This limitation not only enhances the clarity and impact of the mood board but also aligns with the principles of simplicity and succinctness in design.

In terms of image selection, the guidelines set a minimum of five and a maximum of ten images. This range was carefully chosen to allow for sufficient flexibility in showcasing a variety of elements while maintaining a clear and uncluttered layout. The decision to limit the mood board to six images was made to optimize the balance between visual diversity and coherence. This number ensures that each image can be presented with enough detail to be meaningful, yet not so many that they overwhelm the viewer or dilute the mood board’s overall impact.

Regarding the size of the images, while 3DWays did not specify exact dimensions, the requirement was that they should be clearly visible. To meet this criterion, a decision was made to standardize the images at 50 pixels by 50 pixels. This uniformity in size aids in creating a harmonious and aesthetically pleasing grid layout, ensuring that each image is visible and contributes equally to the mood board’s narrative. The choice of image size and the decision to go with six images were made to ensure that the final mood board is not only visually appealing but also practical in terms of visibility and comprehension.

These carefully considered requirements and decisions reflect a deep understanding of design principles and user experience. By setting these parameters, 3DWays aims to guide users in creating mood boards that are not only visually striking but also effectively communicate the essence of their design projects. The process of adhering to these constraints is an exercise in creativity and critical thinking, pushing users to select images that are most representative of their ideas and that work together to tell a cohesive story.

In summary, the mood board PDF's design, dictated by the requirements set by 3DWays and the subsequent decisions made within the project's scope, is a testament to the careful consideration of how best to visually communicate design concepts. The single-page format, the specific number and size of images, and the overall layout are all elements that come together to create a tool that is both useful and inspiring for users, aiding them in bringing their creative visions to life.

8.2 Implementation

The implementation of the Mood board PDF within the project was a nuanced process, heavily reliant on the capabilities of the PyFPDF library, a Python tool known for its efficacy in PDF generation (PyFPDF 2023). The integration of this library was a critical step in the final stage of the dashboard process. As users navigated through the dashboard, selecting images that resonated with their design vision along with the generated captions, these selections were systematically stored in our database. This data storage was crucial for the creation of the mood board PDF.

A unique aspect of the implementation was the handling of image modifications. When users altered an image, the modified version, rather than the original, was designated for inclusion in the mood board. This approach ensured that the mood board accurately reflected the user's creative input and final vision. Additionally, for these modified images, the keywords associated with the images were utilized instead of traditional captions. This decision was driven by the workflow of the application, as outlined in the Streamlit part of the thesis, where image captions are generated prior to any modifications. Consequently, the modified images lacked specific captions, necessitating the use of keywords as a descriptive alternative.

The challenge of implementing the Mood board PDF was further compounded by PyFPDF's reliance on absolute positioning. This required precise placement of every image and text element at specific coordinates within the PDF file that is being created. To address this, a process was developed to standardize the dimensions of each image to 50 pixels by 50 pixels. This standardization was crucial in maintaining a uniform and aesthetically pleasing layout, accommodating the diverse range of images selected by users.

Given the constraint of fitting at least six images onto a single page, alongside their respective keywords, the layout design required careful planning and precision. A dynamic template was devised that could adjust to the varying images and text, ensuring each element was given adequate space and prominence. This template was not only visually appealing but also enhanced the readability and impact of the keywords, which were pulled directly from the database entries.

In conclusion, the Mood board is a fusion of traditional design principles with contemporary digital functionalities. Defined by the requirements set by 3DWays, it effectively encapsulates a user's design process within a single, coherent page. The tool adheres to specific requirements regarding the number and size of images, ensuring a focused narrative that accurately reflects the user's creative process. The inclusion of modified images and associated keywords adds a layer of personalization to the mood board, making it a true reflection of the user's vision and modifications.

The implementation, utilizing the PyFPDF library, navigated the technical challenges of absolute positioning and image standardization. This approach resulted in a layout that balances functionality with the need to convey a clear narrative. It offers users a distinctive tool to visualize and share their design inspirations, bridging the gap between concept and presentation.

9 User Guidelines (group part)

Effective navigation and utilization of a web application are essential, particularly in the context of design and creative ideation. This section is dedicated to providing users with clear and detailed guidelines for interacting with our web application, a tool specifically crafted for design inspiration. These guidelines serve as a practical framework, enabling users to efficiently leverage the application's capabilities to transform their design ideas into visual representations. The application offers a range of features that assist in assembling a digital mood board, an integral component in the visualization of design concepts. By adhering to these guidelines, users will be able to proficiently use the application, facilitating the creation of a mood board that effectively summarizes their design inspirations and ideas. This approach ensures that users can fully engage with the tool's functionalities, making the most of its potential to aid in the design process.

The initial step in accessing the inspiration tool involves account creation, a prerequisite for using the application. This process requires users to set up a username and a password, with the password needing to be a minimum of five characters for security purposes. It is imperative for users to remember these credentials, as they are essential for subsequent access to the application.

Access to the login page is provided on the left side of the computer screen. Within the *Navigation* section, a *Go To* option is available, accompanied by a dropdown button. This button presents two options: *Sign Up* and *Log In*. Selecting *Log In* directs users to the appropriate page for entering their login details.

Users should be aware that the first entry into the application might involve an extended loading time. This delay is attributed to the initialization phase, where necessary libraries are loaded to ensure the proper functioning of the application. This setup is a one-time process, vital for the activation and readiness of all features and tools within the application.

Upon reaching the main page, users are presented with various interactive buttons. The application requires an input of five keywords to begin. While these buttons can be explored in any order, it is recommended to start with the *Generate Suggestions* button for an efficient experience. Subsequent to this step, users are encouraged to interact with other features, further enhancing their engagement with the application.

In the application's interface, users are prompted to input keywords in a designated text box, located above the *Generate Suggestions* button. The interface instructs users to *'Enter at least 5 keywords (separated by spaces):'*. It is crucial for users to note that the application's programming requires these keywords to be separated by spaces for accurate processing. In cases where users intend to input multi-word phrases, such as *'Minimalist Camera'*, the application interprets each word separately. To ensure the phrase is treated as a single entity, users should employ an underscore, inputting it as *'Minimalist_Camera'*. This notation is essential for maintaining the integrity of compound terms, including colour-specific items like *'black_bottle'*, to ensure the application accurately interprets the user's intent.

Upon pressing the *Generate Suggestions* button, two additional buttons become accessible. These buttons are designed to provide users with expanded suggestions of keywords, fostering creativity and offering new design ideas. Should users wish to incorporate these new keywords into their search, they must enter them into the text box above the *Generate Suggestions* button. Furthermore, the *Meaning of the word* button requires users to input a word into the corresponding text box to retrieve its definition. In instances where the application does not recognize a word, it will display a message: *'No definition found for word'*. Users seeking further clarification may resort to external sources such as Google for definitions. Additionally, the *Generate Images* feature presents each image alongside a *Select/Drop this image* button. This functionality allows users to save images they find appealing for further modification. If a user decides against an image, they can deselect it by pressing the button again. Selected images are displayed on the website below the image gallery. Similar to the image selection process, pressing the *More Images* button will display additional images, following the same logic for selection and deselection.

The operational guidelines for the *Modify images selected* button and its associated sub-buttons are outlined as follows:

Image Selection for Modification: Upon clicking the buttons below, a pop-up window will display the selected image. Users are prompted to select points on the image to delineate the object they wish to isolate. Optionally, they can also define a bounding box to specify the area containing the object.

Defining Object Position: Users can interact with the image using mouse clicks to specify areas of interest. A left mouse click is used to mark positive input points – areas to be included in the object isolation. Conversely, a right mouse click is used to mark negative input points – areas the model should ignore. To exit the pop-up window and finalize selections, users should press the *q* key.

Bounding Box Specification: To define a bounding box around the object, users should select two points with a left mouse click. The first click establishes the top-left corner of the box, while the second click sets the bottom-right corner. The pop-up window will automatically close after the second point is selected.

Post-modification, each edited image is saved under a unique filename. Users have the flexibility to edit an image multiple times, with each version being saved. However, it is important to note that only the last modification of an image is retained in the database. Consequently, the application tracks only the final version of each modified image for each user. Should users desire multiple modifications of the same image, they are advised to repeat the entire modification process from the beginning for each desired version.

Furthermore, it is important to note that modified images are not immediately visible within the application. The modifications made by the user will only be displayed at the conclusion of their interaction with the inspiration tool. Therefore, users are encouraged to carefully follow the given instructions during the image modification process to ensure their desired outcomes are achieved.

In conclusion, adhering to these user guidelines is instrumental in maximizing the effectiveness of the web application. By following the outlined steps and recommendations, users can ensure a seamless and productive experience. The guidelines are designed to empower users, enabling them to fully engage with the application's capabilities and to express their creative visions effectively.

10 Results (group part)

After the integration of all functionalities, users are now capable of generating a mood board. All previous Result sections provide insights for the complete tool. To complement them, this chapter focus on presenting the final output. This personalized creation is a culmination of the choices users make throughout their journey on the web application.

At the end of the design process, users can access the "Generate Mood Board" feature, in order to create a cohesive display of the selected images. These images are indicated with a message showing their respective identification numbers (Figure 13). Additionally, the user is prompted with a button to "*Start Mood Generation*" and a hyperlink to download the PDF locally.

This process starts with a filter applied to the 'data' database, retrieving all saved history of the specific user in the web application. Another selection process is needed to choose the final six images which will be compiled in the PDF file. Finally, each image is plotted, supplemented by the descriptive caption generated by the system previously described, to provide additional context to the mood board. These captions are only available to images coming from the image database, as only them are inserted in the caption model. Images suggested at a later stage by the recommender system based on previous interactions are stored differently in the database. Therefore, only the keyword label is presented in the mood board.

An example of a mood board is available in Figure 15 in the appendix.

11 Discussion (group part)

Solution impact 3Dways

Previously to the solution, the process behind product design was archaic, filled with meetings and complex conversations on the topic, without a clear visualisation or sometimes even an idea of how the product should look. Information also needed to be recovered during development as it travelled from mouth to mouth. With the implementation of the application, the process is expected to be cut short. With the help of computation, ideas will flow naturally to the user with visual examples that incentivise the creative procedure. The aid of visual examples looks to maintain the process as true to the idealisation of the client.

11.1 Limitation

While the application holds promise for various advantages, it is crucial to recognise and rectify specific limitations. Acknowledging and addressing these constraints is imperative, as they can impact the application's overall functionality and user satisfaction.

The first limitation is the inheritance of biases in the training data of the models within the image captioning and keyword suggestion functionalities. These biases in the training data could affect the accuracy and neutrality of their outputs. Moreover, the functionality of the keyword suggestion generator is heavily reliant on the continuous availability of the OpenAI API. In instances where the API experiences downtime or is not accessible, the tool's ability to generate keyword suggestions is significantly hindered. This dependency poses a risk, as any interruption in the API service directly impacts the user experience and the overall reliability of the tool. An alternative perspective highlights that the application includes an independent keyword suggestion feature based on user history, functioning without reliance on third-party services such as OpenAI. The effectiveness of the keyword suggestion generator relies on the initial user input, which may limit its utility if users provide vague keywords that the AI cannot give good keyword suggestions in return.

Besides this, it is often not easy to objectively evaluate the output and relevance of the image captions. Some images can be interpreted differently by different people. An example: Consider an image showcasing a rainy street scene with people carrying umbrellas. While the AI model might generate a caption like "People walking on a wet street with umbrellas," different viewers might interpret the scene differently. One viewer might see it as depicting gloomy, rainy weather affecting city life, while another might interpret it as a cosy, atmospheric scene typical of urban life. This variation in human perception highlights the challenge of ensuring that AI-generated captions align with every user's subjective interpretation and the image's intended emotional or thematic context.

In the development of the CLIP model, due to low computational power, the sample size was restricted in order to manage challenges posed by time and memory constraints. As a consequence of the small sample size, a more restricted collection of images limited the model's ability to generalise results in some cases. The similarity scores are low due to the quality of the dataset. It did not abide the case that CLIP's zero-shot classifiers can be sensitive to wording or phrasing. Another limitation is that the model is unable to count the number of objects in an image or classify their distance, making it unreliable in images with a large number of objects.

Meanwhile, for the image modification tool, the implication was that the performance depends on the quality of the input images. Lower resolution or poorly detailed images may not yield optimal results. Inserting an image in the image encoder takes around 1.5 to 2 minutes, using the processing power available in the development phase. This could create a considerable wait time for the user but also enable real-time processing capabilities once all images are processed due to the rest of the system being lightweight.

Finally, some bugs were detected within the segmentation process while integrating the image segmentation code with the Windows operating system. These were only found in the final stage, as the development process was performed in MacOS. This is something to consider when deploying the solution to the client. With limited time on this deliverable, the web application needs clear improvement. Users have to follow the specific steps referenced in the guidelines for the application to work; otherwise, it can crash, causing the program to restart. Modifying images is a time-consuming effort, attributed to the prolonged processing time for the set-up code of the segmentation procedure.

11.2 Recommendations for future work

A general advice for future improvements includes a fine tune of the employed AI models to combat biases and overfitting that might occur. With increasing utilisation of the tool, new user information (text and image) will be created, which can be used to better test the models.

A possible solution for the last limitation mentioned is modifying the web application button that feeds selected images into the SAM model to consider only images the user intends to modify, rather than all images selected for the mood board. Such modification could significantly reduce wait times when the images requiring modification are fewer than those included in the mood board.

When developing the SAM project, Meta created a pre-trained image dataset, SA-1B, including more than one billion masks over 11 million images. If it fulfils all the image requirements, this dataset can be considered to accelerate the segmentation process. In this project, the decision was made to prioritise flexibility by enabling the use of any image database and allowing users to create custom segmentations through points and box prompt inputs rather than limiting the scope to predefined parameters.

By implementing an AI-powered tool to analyse project data and generate a mood board based on shared elements and themes, the tool could effectively capture the project's essence and provide a unified reference point for all involved parties. This would eliminate the need for individual users to select images and would ensure that the mood board accurately reflects the collective vision of the project team. Moreover, adopting a project-level approach would foster collaboration and promote alignment among different entities within the project. The improved mood board would facilitate better communication and understanding of the project's overall aesthetic direction by providing a shared visual reference point. Overall, transitioning from a user-level to a project-level mood board would significantly advance the tool's capabilities. It would enhance collaboration, streamline communication, and ensure the mood board accurately reflects the project's unified visual identity.

Regarding the application, the interface could be improved from the customer's perspective, making it more appealing and user friendly. Another addition should take into account multiple users for a project, creating a specific ID that connects every subject in a project.

12 Conclusion (group part)

The final product of this thesis accomplishes the business objectives defined in the introduction section. Still subject to testing from 3DWays, the tool improves the comprehension of clients’ ideas and refines the requirements for a typical 3DWays project. Given the short development period, there were many limitations and recommendations identified. Nonetheless, we are confident that this is a strong foundation to work on and improve to create a final deliverable tool, as evidenced by the results section.

The platform is compatible with different image databases, contributing to the versatility of the web application. Different variations can be created for specific product design types, each featuring different images available and perhaps additional functionalities.

Overall, the potential and scope of a tool such as this one is substantial, providing an innovative approach to the initial stages of project management, a methodology that can be adapted to other businesses as well. The Python code underpinning the tool holds potential beyond its current application. Each implemented functionality can be tailored to serve different purposes, showcasing the tool’s flexibility. Furthermore, with the rapid development and release of AI models, newer, more efficient, and effective versions of the algorithms used can be integrated into this tool to keep it up to date.

6 Retrieving User Interaction and Personalizing Content through Recommender Systems (Leonor R. Pereira)

6.1 Introduction

6.1.1 Background

As requested by 3DWays, the delivery of the website to the user (clients of the company) is a critical aspect of the project, involving the seamless integration of several essential components, facilitated by the utilization of Streamlit. Subsequently, the establishment of a comprehensive database became imperative to effectively record and store all user interactions. This approach ensures a seamless and efficient user experience. There is a need to improve the user experience on the created inspiration tool, enhancing content based on user interactions, and compiling sophisticated word and image recommendation systems.

Streamlit is a Python library that allows users to create web applications easily and without prior experience. It is user-friendly and was originally designed for Machine Learning users. Streamlit offers an efficient way for data scientists to implement models in their work. It is a good choice for quick and efficient data science work (Gopisetti et al. 2023). The databases can be stored locally for individual users or hosted over internet-based services for broader organizational access. This flexibility also allows exporting databases to larger systems such as PostgreSQL or Oracle (Verma and Sharma 2022). As articulated by Venkatesan and Hariharan, "SQLite3 can be integrated with Python using the sqlite3 module, providing an SQL interface compliant with the DB-API 2.0 specification" (Venkatesan and Hariharan 2021).

In the realm of machine learning, the focus extended to sophisticated word and image recommendation systems. On the one hand, word recommendation system delves into historical interactions, predicting and suggesting potential keywords based on user input. On the other hand, image recommendation system informed by prior interactions on the web application, offer users the most analogous images based on the output of another toll, the imaged and keyword. The system specifically utilizes the Visual Geometry Group's 16-layer convolutional neural networks (VGG16 CNN), developed by Simonyan and Zisserman at Oxford University (2015). VGG16's innovation lies in substituting large filters with sequences of smaller 3x3 filters, contributing to improving model performance. Trained for several weeks using NVIDIA Titan Black GPUs, VGG16 plays a significant role in enhancing the inspiration of the design application (Simonyan and Zisserman 2015).

6.1.2 Objectives

The overall objective of this chapter is to explain how the improvement of "Retrieving User Interaction and Personalizing Content through Recommender Systems" was developed.

The first step was to construct a user-friendly web application interface using Streamlit. The focus of this web application is to seamlessly integrate all its components, resulting in a unified platform. This platform aims to empower designers and product creators by allowing them to follow their creative processes. The web application provides users with a versatile environment that allows effortless interaction with various tools and features, leading to an elevated and personalised user experience.

Subsequently, the second objective of this section aims to create a reliable database management system to document and store user interactions accurately.

The third purpose aims to explore the implementation of a Word Recommendation System by designing and developing an advanced system that draws upon historical user interactions to provide recommendations for new inputs. The system leverages user-input keywords as a basis for its predictions to improve content and enhance the experience of this design inspiration tool within the platform.

Fourth, the Image Recommendation System plays a crucial role in achieving the objectives of this study. This sophisticated system that leverages past interactions to suggest images most similar to those previously provided to the user is worthy of consideration. This advanced technology employs an automated analytical process to evaluate the user's prior image selections, identifying resemblances and recommending similar alternatives. This technology can significantly reduce the time and effort required to search for new images independently.

Finally, the primary purpose of this section is to build a user-friendly web interface by providing a reliable database system that documents and stores user interactions. To do so, it explores implementing a Word and Image Recommendation System based on previous interactions to enhance the user experience by providing a tailor-made interface. Overall, this word recommendation system will help users find more relevant information, leading to increased engagement and satisfaction. This image recommendation system will save users time and effort in finding new images independently. Both recommendation systems will improve content personalisation, as users will receive recommendations based on their interests and preferences.

6.1.3 Structure Thesis

The structure of this thesis is organized to be coherent and to flow through three core elements: web application, recommended words, and recommended images.

The paper begins with an introduction to the problem and objectives. This is followed by a chapter on Literature review composed of exhaustive research into these topics and related ways of applying them, establishing the foundation of each topic. The topics in consideration are Streamlit implementation and its connection with creating two databases for both recommendation systems of keywords and images. Connecting these topics with the realm of Human-Computer Interaction (HCI) and Machine Learning (ML) is divided into two parts. The first is Natural Language Processing (NLP) and the second Computer Vision. Furthermore, the next chapter, Methodology, has the explanation and description of the methodologies used for the development of the website and the database, the recommended words and image suggestions. Succeeding this, the Results present the findings and the output of the research. Finally, the Conclusion and Discussion chapter summarizes the most important findings and their applications.

6.2 Literature review

6.2.1 Text Mining

The advent of digital data acquisition technologies has generated a considerable volume of data, a significant portion of which is unstructured or semi-structured information. Analysing text documents within this vast dataset presents a formidable challenge. Text mining (TM) is a process for extracting valuable and non-trivial patterns from an extensive corpus of text documents. It employs various techniques and tools for future prediction and decision-making. The selection of appropriate and effective TM techniques enhances speed and reduces the time and effort required for information extraction. Additionally, it identifies critical issues in the field that impact results, accuracy, and relevance, contributing to a nuanced understanding of TM challenges and opportunities (Talib et al. 2016).

The data overload in this era of Big Data overwhelms users with information. The integration of TM techniques emerges as a valuable strategy for recommender systems, which play a crucial role in alleviating the overload of information. TM can be effectively harnessed for recommender systems by discerning user preferences through user reviews and extracting contextual data from diverse textual sources, including item descriptions and users' writings on platforms such as social networks. This TM involves performing activities like sentiment analysis, opinion mining, and information extraction. Furthermore, it incorporates semantic techniques, particularly ontology-based approaches. Ontologies are a structured vocabulary that helps define types, properties, and interrelations of concepts to support data mining tasks and enhance recommender systems. They establish relationships between users and their preferences in recommendation contexts. In TM, ontology-based recommender systems use controlled vocabularies to categorise topics from natural language text messages. This approach supports information extraction and contributes to a more nuanced, context-aware recommendation process. The existing works in this domain are based on the types of textual data analysed, like user reviews, user-generated texts, or item-related text, and underscore the significance of considering user context and preferences. The exploration of user-generated texts beyond reviews is identified as an underexplored area, suggesting a future integrated approach combining various TM techniques to build a context-aware recommender system (Betancourt and Ilarri 2020; Gruber 1993; Leite Da Silva et al. 2023).

Case Study

TM possesses innumerable applications and employs a plethora of methods, among which the CountVectorizer (CV) function is of particular significance, and it will be extensively examined in this paper. As per example:

In 1982, an early application of text mining was introduced in a paper titled 'Natural Language Access to Structured Text' (Hobbs, Walker and Amsler 1982). This paper addressed the issue of providing natural language access to textual materials by employing logic representation and coherence analysis to retrieve relevant passages. Precisely, the Text Access component of the system translated user requests into logical form and resolved discourse problems, while the Text Structure component analysed paragraphs and organised them hierarchically. Notably, the system featured automatic procedures to generate structured text from the main content, being it an early venture into TM methodologies.

The ever-evolving field of technology has expanded the possibilities of TM, including its potential to detect misleading information on social media platforms like Facebook and Twitter. To achieve this, three distinct techniques were utilised: Term Frequency-Inverse Document Frequency (TF-IDF), CV, and Hashing-Vectorizer (HV), all of which fall within the realm of text mining. In order to determine which classification model is most effective at identifying inauthentic features, a set of twelve classifiers that were previously defined were employed. Specifically, the use of CV involves the transformation of a vector of words constituting sentences or text into a numerical data matrix, a necessary preprocessing step for applying ML to the model (Kaur, Kumar and Kumaraguru 2019).

9.2.2 Image Similarity

Historically, image similarity was calculated using different methods, such as Mean Squared Error (MSE) and Structural Similarity Index. In more detail, the MSE computed the mean value of the square of the differences between the pixels of two images. Having different calculations for the images, mean the location of the pixel intensities and separately doing the same calculation for the colour. However, these pixel-based methods could have been more effective when the input images were taken under different angles or lighting conditions, even though that has a low computational cost. Nowadays, with the advancement of technology permits, the use of ML and deep learning methods such as convolutional layers to learn from the image's features. Although this method requires more computer capacity, the method is more effective and does not have the limitations that the traditional methods have (Ornek, Celik, and Alper 2021).

Moreover, image similarity can be calculated with other sophisticated methods, such as the scene graph method, which converts image similarity into text similarity. Another approach is ratio and distance calculation between features, which provides a normalised similarity value. These methods have improved the accuracy and robustness of image similarity calculation, catering to the increasing complexity and variety of image data (Bohush et al. 2020).

In this paper, for Image Similarity, the model chosen was VGG16 CNN. This model follows a deep-learning architecture in the field of machine learning. Specifically, a convolutional neural network (CNN) is widely recognised and used in a large-scale of image recognition tasks. Furthermore, as part of a broader field of deep learning, CNNs are characterised by their ability to learn hierarchical features from the input data automatically. VGG16 achieves its depth network by stacking multiple convolutional layers, enabling the extraction of intricate patterns and representations from images (Simonyan and Zisserman 2015). It is considered one of the top models from the ILSVRC-2014 competition, with an accuracy of 92.7% in the ImageNet dataset (Simonyan and Zisserman 2015).

Case Study

The model can be used in an extensive range of fields, from health to security, and location. The application on health areas more directly in the area of imageology, such as Magnetic Resonance (MR) and X-ray, helps to identify diseases such as cancer, tumours (Sowrirajan, Balasubramanian and Raj 2022), and pneumonia (Jiang et al. 2021).

In recent years, the evolution of these technologies has increased immensely, together with the increase of research in the domain of brain image classification, leading to a deep learning approach like CNN. The utilisation of VGG-16 Deep Convolution Neural Network (DCNN) in combination with transfer learning is crucial for preventing overfitting on a small dataset while classifying abnormal or normal MR. This is achieved by extracting features from raw images while learning and evolving from the same set of images. The previously described method is necessary to complement the classical ML methods, which require handcraft features to perform classifications. The model VGG16, where just the last layers are trained to accommodate new image categories, achieves a high recognition rate, outperforming classical methods. Despite a prolonged training time, the study highlights the potential of pre-trained DCNNs for efficient MR brain image classification, emphasising the superiority of end-to-end categorisation on raw images without intricate feature extraction (Bhanothu, Kamalakannan and Rajamanickam 2020; Kaur and Gandhi 2019).

The VGG16 CNN model has practical security applications, particularly facial recognition, as the paper by Lou and Shi (Lou and Shi 2020) explores the use of the same. The authors highlight the importance of image recognition in the current era of social media and the rapid growth of image data. The paper emphasises the critical role of CNNs in image recognition, focusing on the VGG16 model. The model has been applied to various datasets, such as the ORL Face Database, BioID Face Database, and CASIA Face Image Database, to validate its recognition accuracy.

6.2.3 Web Application

The field of Human-Computer Interaction (HCI) explores the design and utilisation of computer technology, emphasising the interfaces between users and computers. Its origins are from the late 1970s with the advent of personal computing. The intersection of cognitive science, incorporating psychology, AI, linguistics, and philosophy, with personal computing needs gave rise to HCI as an early example of cognitive engineering. HCI encompasses diverse disciplines, such as human factor engineering, documentation development, and various design areas. This discipline ensures user satisfaction and draws from a multidisciplinary pool, including computer graphics, operating systems, linguistics, psychology, and engineering. Envisioning the future, HCI is poised to embrace ubiquitous communication, high-functionality systems, mass availability of computer graphics, mixed media integration, high-bandwidth interaction, large and thin displays, and information utilities. Rooted in historically fortuitous developments, HCI has continued to evolve, reflecting a synthesis of diverse perspectives and paradigms since its significant emergence around 1980 (Andurkar 2015). In recent years, technologies have advanced exponentially. Nowadays, everyone desires speedy responses, leading to a never-ending cycle where fast interactions result in technological advancements, leading to even faster interactions, with the fields of HCI and Computer Vision constantly evolving.

The tool created is delivered to the user through the Streamlit interface, which is used in the field of HCI. Streamlit is designed explicitly for delivering ML tools, which means the potential applications are the same as ML.

6.3 Methodology

6.3.1 Streamlit & Database (sqlite3)

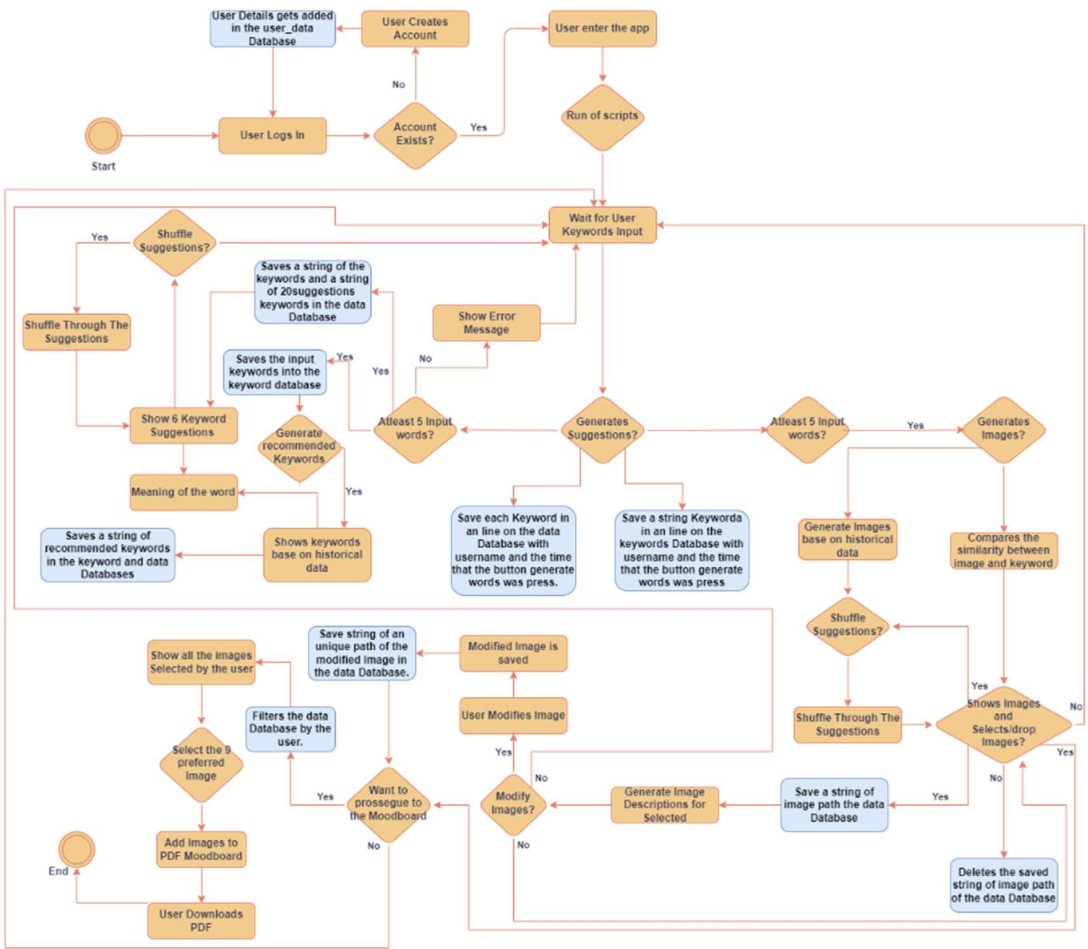

Figure 3. Flowchart of the web application plus the datasets

Streamlit, as referenced before, is used to create the web application's front end. To better understand the user perspective, the flowchart above illustrates the connection between the user perspective and the interlinked databases.

The flowchart uses two colours, orange and blue, which represent the app and the databases created, respectively. Sequentially, it is divided into shaped lozenges and rectangles. The lozenges describe the actions and decisions the user can make, while the rectangles represent the tools and the consequences of those actions.

The web application initiates with a sign-in page, enabling users to register with a unique username and a password of at least five characters. A warning is displayed if the chosen username is already in use, as the username serves as a primary key in the 'user_data' database. In addition, a dropdown button offering the option to log in is situated on the left side of the interface. Opting for a login directs the user to the corresponding login page. Here, users input their usernames and passwords. A warning message appears when the password is incorrect. Upon successfully entering the credentials, the user is directed to the main application page. In the background, the username and password are saved if the user is not in the 'user_data' database.

Resource-intensive programs and scripts are executed during this initial phase to ensure optimal application performance. Upon completing these processes, the user gains access to the application with each page and database linked to the respective username.

Entering the main page, the user encounters a title accompanied by a box to input keywords. In the background, two databases are created and subsequently connected. The first, named 'data' includes columns such as *'username'*, *'keywords'*, *'sug_keywords'*, *'current_datetime'*, *'images_paths'*, *'selected_images'*, *'recommended_words'*, *'recommended_images'*, *'images_description'*, and *'inpaint_save'*. The second database, named “*keywords*”, comprises columns like *'username'*, *'keywords'*, *'current_datetime'*, and *'recommended_words'*. While both databases serve distinct purposes, 'data' is considered the primary database, while *'keywords'* supports the word recommendation system. This database was created not only to enhance the user experience but also to support the app and connect outputs and inputs of all used tools in the website to make it unique.

Returning to the user's perspective, the user inputs five keywords into the text box, prompting a warning if the count is lower than five. Three simultaneous actions occur after pressing the "Generate Suggestions" button or the enter key, connecting to a tool. In the background, the databases are updated with the username, current date-time, and keywords. The insertion format differs for each dataset, with 'data' accepting one keyword per line and 'keywords' receiving a string containing all keywords. Additionally, a string of twenty suggested words generated by the tool is inserted into the primary database in the 'sug_keywords' column. After pressing the referred buttons, two more appear: 'Shuffle' and 'See previously used Keywords'.

A global variable was created to store suggested words that presented a crucial obstacle. The 'Shuffle' feature randomly chooses six words from this variable without compromising the integrity of the suggestion system, thus streamlining operations and reducing costs for the company, as explained previously. At the same time, 'See previously used Keywords' provides recommended words based on historical data from the 'keywords' database, saving on both databases on the column 'recommended_words' a string with the results. Furthermore, users have the option of enquiring about the meaning of a word. If the user is unfamiliar with a word, the user can write it in the text box above the meaning and press the 'Meaning of the word' button, which reveal the word's definition below.

By pressing the 'Generate Images' button, the similarity between keywords and images is calculated, and the results are stored in a dictionary. This setup is essential for the 'Select/Drop' button, allowing users to choose or remove images. This selection process is a vital part of the user experience. When a user selects or drops an image, the image path is added or removed in the 'data' database under the 'selected_images' column. Additionally, this button displays all the images without needing to rerun this tool. This enhances the application's efficiency, as it avoids repetitive calculations and provides a smoother user experience. This button has other characteristics, such as connecting to a tool that generates captions for the selected images and saving or deleting these captions from the primary database in the 'images_description' column.

Moreover, the 'Generate more images' button was introduced to calculate the similarity of historical images with the ones generated by the 'Generate Images' button, providing the most similar images to each one given by the previous button. A string containing each recommended image path is saved on the primary database in the 'recommended_images' column.

Pressing the 'Shuffle Images' button triggers a shuffle that randomly displays an image per keyword from the three images calculated by the keyword. These images are displayed in the same way as the ones above, however, the 'Select/Drop' button for these images works slightly differently. The button creates one more line in a data database, with the values of the columns filled except the caption values, resembling an image generated by the button 'Generate Images' above it. Another characteristic of the button is the option of unselecting the images when pressed again, deleting the corresponding line from the database previously created upon the selected image. Using the selected images, the user can proceed to the next step.

The 'Modify images selected' button connects with the editing tool, enabling users to modify images and save them. The edited images are saved in the same folder as the originals ones, each with a unique name to track multiple modifications. These names are created using a combination of the current time and a unique identifier, ensuring that every edit is distinct. The edited images are then stored in the 'inpaint_images' column of the 'data' database. This approach efficiently handles the storage and retrieval of multiple image edits.

Finally, the user presses the button 'Mood Board Generation?' and another window is opened connecting to the database to retrieve all the user's historical data.

6.3.2 Words Recommendation System

Text mining, also known as knowledge discovery from text (KDT), involves a machine-supported analysis of textual data and merging techniques from information retrieval, information extraction, and natural language processing (NLP). This interdisciplinary approach integrates algorithms and methods from knowledge discovery in databases (KDD), data mining, machine learning, and statistics. Focused primarily on unstructured textual documents, text mining requires innovative approaches to data analysis. Its definition varies, ranging from equivalence to information extraction, the application of machine learning and statistical algorithms to texts (akin to data mining), and integration into the broader knowledge-discovery process. Text mining is primarily perceived as text data mining, emphasizing the extraction of meaningful patterns from textual data to categorize, structure, and unveil valuable insights within extensive text collections (Hotho Nürnberger and Paaß 2005).

The word recommender system can be used as a content-based recommender system. Having content-based recommenders rely on the inherent characteristics or features of the items being recommended to provide personalized suggestions to users.

More precisely, the tool used in the text mining process on this paper is CV. CV is library that is utilized to convert text data into a form that can be comprehended by machine learning algorithms. Specifically, it converts the text into a matrix of token (word) counts. Each row in the matrix corresponds to a vector of words and each column represents a unique word in the text vector. The value in each cell is the number of times a word appears in a document (Kaur, Kumar and Kumaraguru 2019). Specifically, when utilizing CV and tokenizing the words, the resulting matrix is structured such that each row represents an entry of keywords, while each column corresponds to a unique word in the vocabulary, effectively capturing the frequency distribution of words across documents.

This transformation changes the vectors of strings to vectors of zeros and ones, allowing the use of the metric cosine similarity to calculate the similarity between each pair of vectors resulting in a similarities matrix. This matrix provided a measure of how similar each set of keywords is to one another. Cosine similarity serves as a metric for quantifying the similarity between two vectors within a multi-dimensional space. The computation of cosine similarity involves obtaining the dot product of the vectors and subsequently dividing it by the product of their magnitudes, also known as Euclidean norms. Mathematically expressed, the cosine similarity between vectors A and B is denoted as:

$$cosine\_similarity(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

In this representation, $A \cdot B$ signifies the dot product of vectors A and B, while $\|A\|$ and $\|B\|$ denote the magnitudes (Euclidean norms) of vectors A and B, respectively. The resultant cosine similarity score spans from -1 (indicating complete dissimilarity) to 1 (representing absolute similarity), representing the between the vectors. In the provided code context, this calculation is employed to assess the similarity between keyword vectors, facilitating the identification of commonalities between input keywords based on their keyword usage.

The flowchart below provides an overview of the process used in the recommended words for the inspiration tool.

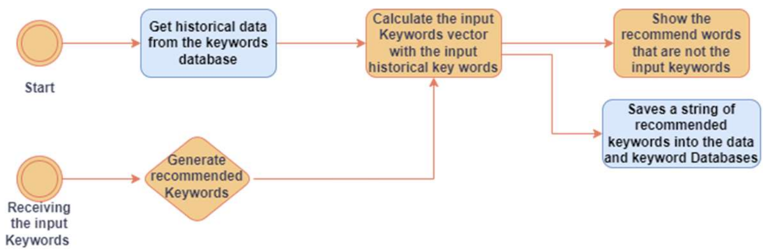

Figure 4. Flowchart of the Words Recommendation Systems

Undertaking the calculation of the similarities between two vectors, one vector that comprises all the input words of the user with each vector that is a line in the keywords dataframe, retaining the top three most similar vectors. Subsequently, the shared words and any repetitions are removed within the top three vectors, thus yielding a unique keyword vector that is subsequently transformed into a string. Returning some probable words based on historical data, the user can rely on the recommended words for more inspiration or ideas.

6.3.3 Image Recommendation System

The Image Recommendation System uses CNN and deep learning. CNNs are a class of deep neural networks commonly applied to analysing visual imagery. The CNN and DCNN are known for their deep architecture, which consists of multiple layers that can capture a hierarchy of features from low to high levels. Despite the success in various tasks, CNNs are often considered 'black boxes' due to complex and opaque decision-making processes. Visualisation techniques have been developed to interpret the activations and features within CNNs, providing insights into CNN's internal workings. These techniques involve transforming the network into a graph-based representation, clustering neurons, and employing algorithms to reduce visual clutter. Visualising the activation patterns and feature hierarchies makes diagnosing and refining the network possible, enhancing the model's performance and interpretability. This approach is crucial for advancing the understanding of CNNs and developing more transparent ML models (Zeiler and Fergus 2014; Liu et al. 2017).

The features of the images used are extracted using the model VGG16 CNN, with the similarity calculation between these features relying on the cosine similarity explained above. Combining the model and the cosine similarity makes up the recommendation systems. In order to use the VGG16 CNN, it was necessary to download the pre-trained model, which is a prize-winning model and a popular one in computer vision. This uses the ImageNet dataset, which is capable of recognising objects in images. The main advantage of using this model is the capability to extract high-quality features from the images (Simonyan and Zisserman 2015).

The architecture of VGG-16, as illustrated in Figure 23, exhibits a sequential processing approach. The input image, with dimensions 224 x 224, is initially passed through the first convolutional layer, conv1, which applies a series of convolutions with a 3 × 3 receptive field. The convolutional stride is set at 1 pixel. Following specific convolutional layers, spatial pooling is achieved through five max-pooling layers, each with a stride of 2 pixels and a 2 × 2-pixel window. This results in the down sampling of the spatial dimensions. The architecture concludes with three fully connected (fc) layers, each with a channel size of 4096, 4096, and 1000 neurons, respectively. In these layers, each neuron receives input from the activations of neurons in the preceding layer (Kaur and Gandhi 2019).

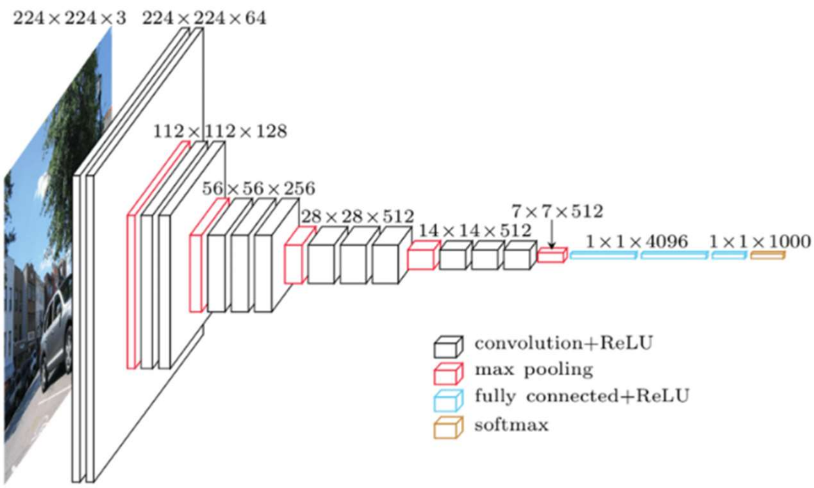

Figure 5. Visualizing Complexity: A Snapshot of a Convolutional Neural Network in Action (Simonyan and Zisserman 2015)

The flowchart below gives a general view on the role of the image recommender system within the tool.

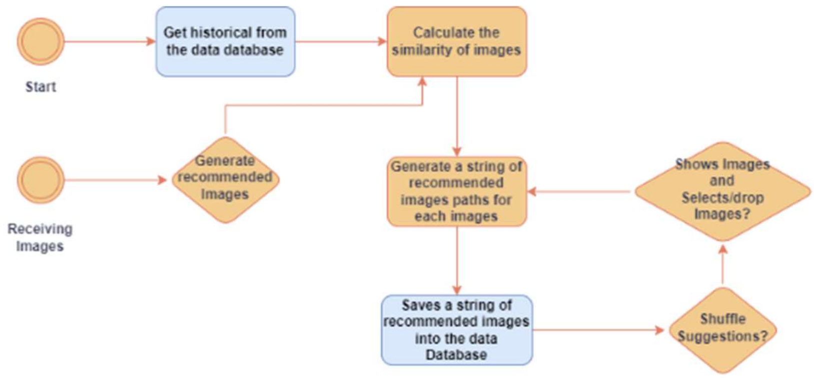

Figure 6. Flowchart of the Images Recommendation Systems

The application uses the VGG16 CNN model to make the image recommender system. Upon launching, the code initiates by establishing a connection to the sqlite3 database, fetching relevant historical data, meticulously handling missing values in the dataframe, dropping rows with incomplete information and prioritising data integrity. The model is used to extract the features of each historical image and save them in a list as the respective path to the images. Subsequently, the user's inputs on the app select a subset of image paths, and features are extracted for these images using the model. Furthermore, cosine similarity is used to calculate the historical images most similar to the new images, using the feature vectors previously saved in a list. Then, images corresponding to each keyword are saved in a dictionary with respective paths. This comprehensive workflow showcases the seamless integration of historical image data and the utilisation of pre-trained deep-learning models to provide meaningful and relevant image recommendations within the application.

This method enhances the quality of recommendations and contributes to the system's transparency and interpretability, which are crucial for user trust. The system personalises recommendations by integrating these historical data and improving user experience. It compares new image inputs with historically saved data, identifying similarities and offering users a selection of images that closely align with their preferences. This seamless integration of advanced image processing and user-focused design culminates in an intuitive and responsive user interface, where personalised recommendations enhance user interaction and engagement with the application.

6.4 Results

Integrating all components into the web application using Streamlit and sqlite3 has created a robust foundation for the creativity tool, significantly enhancing user experience. This application features a comprehensive warning system and intuitive navigation, making it highly user-friendly. The databases - 'user_data', 'data', and 'keywords' - are pivotal in storing and utilising historical user interactions, thus forming the backbone of personalised recommendation systems. Detailed explanations of these databases and their structures are available in appendix Table 2, also with some results.

Crucially, all variables are stored in text format for simplicity and efficiency. These databases are instrumental in chronicling user interactions, capturing data from the oldest to the most recent activities. They play a vital role in the development and functionality of the website, ensuring a seamless and personalised user experience. The recommendation systems developed using these databases are tailored to individual users, making the web application unique for each visitor. Detailed results and structures of the tools used in the development process, as outlined in chapter 6.3.1 - Streamlit & Database (sqlite3), are illustrated in Figures 7 to 12 in the appendix.

The Word Recommendation System, powered by CountVectorizer, exemplifies how text mining and machine learning can be employed to personalise the website. It considers not only recent interactions but also historical data, thus offering truly bespoke recommendations. For instance, the system can discern user keyword patterns and preferences, such as 'minimalist_camera', 'DSLR', and 'lens', and offer tailored suggestions like 'camera_gear' and 'Waterproof'. This adaptive recommendation approach enhances user engagement by providing unique and relevant content.

Similarly, the Image Recommendation System, driven by the VGG16 CNN model, is vital in enriching the user experience. It recognises a wide array of objects by leveraging its pre-trained capabilities, allowing for high-quality feature extraction essential in computing image similarity. This system offers users visually analogous images based on their interactions and ensures these recommendations are relevant and engaging. This dynamic technology and user interaction data underscores the effectiveness of the Image Recommendation System in delivering a visually rich and personalised user experience.

Overall, the results from integrating these systems within the application align with the project's objectives of enhancing user interaction and personalisation. The application facilitates creative expression and ensures a seamless, intuitive, and personalised experience, thereby significantly improving user engagement and satisfaction.

6.5 Conclusion & Discussion

In conclusion, this paper "Retrieving User Interaction and Personalizing Content through Recommender Systems" has addressed the critical aspects of creating a dynamic and personalized web application. By seamlessly integrating Streamlit, sqlite3, and sophisticated recommendation systems for words and images, the tool offers users an intuitive and engaging experience.

The utilization of Streamlit, known for its characteristics of easy use, has facilitated the development of a user-friendly interface. The integration with sqlite3 ensures the efficient recording and storage of user interactions, paving the way for a robust foundation. The Word Recommendation System, powered by CountVectorizer, leverages text mining techniques to enhance content based on historical user interactions, providing personalized word suggestions. Simultaneously, the Image Recommendation System, driven by the VGG16 CNN model, excels in suggesting visually similar images, contributing to a visually engaging user experience. The tool's architecture, outlined in the methodology section, demonstrates a coherent and interconnected flow of actions. From user registration to keyword and image recommendations, the application strives to offer a versatile environment for users to interact with multiple features seamlessly. Results indicate that the integration of all components has effectively delivered a creative and personalized web application. The Word Recommendation System, with its ability to predict and suggest keywords based on user input, contributes to the improvement of content relevance. The Image Recommendation System, utilizing the capabilities of the VGG16 CNN model, enriches the user's visual experience by providing relevant and appealing image suggestions. While the project has achieved its objectives, it is important to acknowledge certain limitations, such as potential biases in recommendations and the impact of resource-intensive processes on performance. Recommendations for the future include exploring collaborative filtering techniques, addressing biases, and optimizing performance for larger datasets.

In summary, "Retrieving User Interaction and Personalizing Content through Recommender Systems" contributes to the intersection of technology and creativity while at the same time provides a robust and versatile platform for users to enhance the user creative processes. The project's success lies in its commitment to user-centric design, personalization, and the seamless integration of diverse components.

13 References and appendix

References

3DWays. (2023a). *LinkedIn*. Linkedin.com. https://www.linkedin.com/company/3dways/about/

3DWays. (2023b). *Website*. 3dways.pt. https://3dways.pt/

Aditya, J. N. (2020). *Flickr8K Dataset*. Kaggle. https://www.kaggle.com/datasets/adityajn105/flickr8k/code

Betancourt, Y., & Ilarri, S. (2020). Use of Text Mining Techniques for Recommender Systems. Proceedings of the 22nd International Conference on Enterprise Information Systems. https://doi.org/10.5220/0009576507800787

Bhanothu, Y., Kamalakannan, A., & Rajamanickam, G. (2020). Detection and Classification of Brain Tumor in MRI Images using Deep Convolutional Network. *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. https://doi.org/10.1109/icaccs48705.2020.9074375

Bostrom, N., & Eliezer Yudkowsky. (2018). The Ethics of Artificial Intelligence. *Chapman and Hall/CRC EBooks*, 57–69. https://doi.org/10.1201/9781351251389-4

P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., & Hesse, C. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html?utm_medium=email&utm_source=transaction

Chi, L., Jiang, B., & Mu, Y. (2020). Fast Fourier Convolution. *Advances in Neural Information Processing Systems*, *33*, 4479–4488. https://proceedings.neurips.cc/paper/2020/hash/2fd5d41ec6cfab47e32164d5624269b1-Abstract.html

Chipambwa, W., & Chikwanya, E. (2022). *Design communication: Fashion design student's perspectives on digital vs physical mood boards*. https://www.grid.uns.ac.rs/symposium/download/2022/76.pdf

Davenport, T. H. (2018). *The AI advantage : how to put the artificial intelligence revolution to work*. The Mit Press.

EPSU. (2018). *The General Data Protection Regulation (GDPR) AN EPSU BRIEFING*. https://www.epsu.org/sites/default/files/article/files/GDPR_FINAL_EPSU.pdf

Fenza, G., Gallo, M., Loia, V., Orciuoli, F., & Herrera-Viedma, E. (2021). Data set quality in Machine Learning: Consistency measure based on Group Decision Making. *Applied Soft Computing*, *106*, 107366–107366. https://doi.org/10.1016/j.asoc.2021.107366

Fox, F. (2023, September 19). *5 Challenges Businesses Face When Integrating AI Solutions*. Ratherlabs.com. https://www.ratherlabs.com/blog/5-challenges-businesses-face-when-integrating-ai-solutions

Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications A Translation Approach to Portable Ontology Specifications. Appeared in Knowledge Acquisition, 5(2), 199–220. https://tomgruber.org/writing/ontolingua-kaj-1993.pdf

Hobbs, J. R., Walker, D. E., & Amsler, R. A. (1982). *Natural language access to structured text. In Coling 1982: Proceedings of the Ninth International Conference on Computational Linguistics.*

Hotho, A., Nürnberger, A., & Paaß, G. (2005). A Brief Survey of Text Mining. *Journal for Language Technology and Computational Linguistics*, *20*(1), 19–62. https://doi.org/10.21248/jlcl.20.2005.68

Jiang, Z.-P., Liu, Y.-Y., Shao, Z.-E., & Huang, K.-W. (2021). An Improved VGG16 Model for Pneumonia Image Classification. *Applied Sciences*, *11*(23), 11185. https://doi.org/10.3390/app112311185

Karabiber, F. (2023). Cosine Similarity. LearnDataSci. https://www.learndatasci.com/glossary/cosine-similarity/?fbclid=IwAR2sTQUDdMFbxA4AQFdDZDOtz7nBRDn7-nNzHxQVkeVvle8eQSS1m206-iw

Kaur, S., Kumar, P., & Kumaraguru, P. (2019). Automating fake news detection system using multi-level voting model. *Soft Computing*. https://doi.org/10.1007/s00500-019-04436-y

Kaur, T., & Tapan Kumar Gandhi. (2019). *Automated Brain Image Classification Based on VGG-16 and Transfer Learning*. https://doi.org/10.1109/icit48102.2019.00023

Kotter, J. P. (1996). *Leadership change. Harvard Business School Press: Boston, MA, USA.*

Krasin I., Duerig T., Alldrin N., Ferrari V., Abu-El-Haija S., Kuznetsova A., Rom H., Uijlings J., Popov S., Veit A., Belongie S., Gomes V., Gupta A., Sun C., Chechik G., Cai D., Feng Z., Narayanan D., Murphy K. OpenImages: A public dataset for large-scale multi-label and multi-class image classification, 2017. Available from https://github.com/openimages

Leita da Silva, F. L., Slodkowski, B. K., da Silva, K. K. A., & Cazella, S. C. (2022). A systematic literature review on educational recommender systems for teaching and learning: research trends, limitations and opportunities. Education and Information Technologies. https://doi.org/10.1007/s10639-022-11341-9

Lewin, K. (1951). *Field theory in social science: selected theoretical papers* . https://psycnet.apa.org/record/1951-06769-000

Liu, M., Shi, J., Li, Z., Li, C., Zhu, J., & Liu, S. (2017). Towards Better Analysis of Deep Convolutional Neural Networks. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 91–100. https://doi.org/10.1109/tvcg.2016.2598831

Lou, G., & Shi, H. (2020). Face image recognition based on convolutional neural network. *China Communications*, *17*(2), 117–124. https://doi.org/10.23919/jcc.2020.02.010

McKinsey. (2020). *Global survey: The state of AI in 2020*. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Global%20survey%20The%20state%20of%20AI%20in%202020/Global-survey-The-state-of-AI-in-2020.pdf

Nagalapuram G.D., Roopashree S., Varshashree D , Dheeraj D , & Nazareth D.J. (2021). Controlling Media Player with Hand Gestures using Convolutional Neural Network. *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*. https://doi.org/10.1109/mysurucon52639.2021.9641567

Örnek A.H, Çelik M, & Alper O.C. (2021). Comparing Image Similarity Methods for Face Images. 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET). https://doi.org/10.1109/icecet52533.2021.9698252

Paullada, Inioluwa Deborah Raji, Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns*, *2*(11), 100336–100336. https://doi.org/10.1016/j.patter.2021.100336

PyFPDF. (2023). Readthedocs.io. https://pyfpdf.readthedocs.io/en/latest/ReferenceManual/index.html

Rieuf, V., Bouchard, C., & Aoussat, A. (2015). Immersive moodboards, a comparative study of industrial design inspiration material. *J. Of Design Research*, *13*(1), 78. https://doi.org/10.1504/jdr.2015.067233

Rosenthal, P., & Niggemann, O. (2022). *Problem examination for AI methods in product design*. ArXiv.org. https://arxiv.org/abs/2201.07642

Russel, S., & Norvig, P. (2020). *Artificial intelligence: a Modern approach.* (4th ed.). Prentice Hall.

Rykhard Bohush, Sergey Ablameyko, Adamovskiy, E. R., & D. Savca. (2020). Image Similarity Estimation Based on Ratio and Distance Calculation between Features. Pattern Recognition and Image Analysis, 30(2), 147–159. https://doi.org/10.1134/s1054661820020030

Salierno, G., Leonardi, L., & Giacomo Cabri. (2021). The Future of Factories: Different Trends. *Applied Sciences*, *11*(21), 9980–9980. https://doi.org/10.3390/app11219980

Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S., Crowson, K., Schmidt, L., Kaczmarczyk, R., & Jitsev, J. (2022). LAION-5B: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, *35*, 25278–25294. https://proceedings.neurips.cc/paper_files/paper/2022/hash/a1859debfb3b59d094f3504d5ebb6c25-Abstract-Datasets_and_Benchmarks.html

Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image*

Recognition. ArXiv.org. https://arxiv.org/abs/1409.1556

Sowrirajan, S. R., Balasubramanian, S., & Raj, R. S. P. (2023). MRI Brain Tumor Classification Using a Hybrid VGG16-NADE Model. *Brazilian Archives of Biology and Technology*, *66*. https://doi.org/10.1590/1678-4324-2023220071

Specker, A. A. S. (2022). *Accelerating Digitalization within 3dways: Implementation of a Machine Learning Tool for Design Inspiration and Process Optimization*.

Tasq.ai. 2023. "What Is Bounding Box | Tasq.ai." Tasq.ai. February 22, 2023. https://www.tasq.ai/glossary/bounding-box/#:~:text=What%20is%20the%20bounding%20box,of%20interest%20inside%20each%20image.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. *Lecture Notes in Computer Science*, 818–833. https://doi.org/10.1007/978-3-319-10590-1_53

<u>Appendix</u>

Databases	Variables	Type	Description	Example
user_data	username	TEXT	Represents the username of the user, and it is a primary key.	Lisa
	password	TEXT	Stores the individual keywords entered by the user.	12345
data	username	TEXT	Represents the username of the user.	Lisa
	keywords	TEXT	Stores the individual keywords entered by the user	Red_car
	sug_keywords	TEXT	Contains the generated keyword suggestions.	Canine Hound Puppy Mutt Feline Kitten Pussycat Tomcat Human Individual Being Entity Rodent Cursor Pointing device Trackball Scarlet auto Cherry motor Crimson vehicle Rubicund car
	current_datetime	TEXT	Records the timestamp when the data is stored.	06/12/2023 20:31
	images_paths	TEXT	Stores the paths of images associated with each keyword.	C:/Users/lmrod/fiftyone/open-images-v6/validation/data/4bbd710d9e8483b0.jpg
	selected_images	TEXT	Indicates which images are selected by the user.	C:/Users/lmrod/fiftyone/open-images-v6/validation/data/4bbd710d9e8483b0.jpg
	recomended_words	TEXT	Stores the recommended words for each keyword.	horse,birds
	recomended_images	TEXT	Contains the paths of recommended images for each keyword.	C:/Users/lmrod/fiftyone/open-images-v6/validation/data/1700351849_a1a11e24-3511-4730-98ce-4d886a7ed64c_remove_inpaint.png,C:/Users/lmrod/fiftyone/open-images-v6/validation/data/1700352753_aa972123-fd5e-4b9d-880b-fc35bd3f8376_remove_inpaint.png,C:/Users/lmrod/fiftyone/open-images-v6/validation/data/1700351849_a1a11e24-3511-4730-98ce-4d886a7ed64c_remove_inpaint.png
	images_descripition	TEXT	Holds the descriptions associated with each image.	a red car parked in front of a store
	inpaint_save	TEXT	Stores the paths of images after modification using inpainting.	C:/Users/lmrod/fiftyone/open-images-v6/validation/data/1701895364_380300c3-101f-4b81-bfb3-621ad598099a_sam.png
keywords	username	TEXT	Represents the username of the user.	Lisa
	keywords	TEXT	Stores the individual keywords entered by the user.	dog,cat,person,mouse,red_car
	current_datetime	TEXT	Records the timestamp when the data is stored.	06/12/2023 20:31
	recomended_words	TEXT	Stores the recommended words for each set of keywords.	horse,birds

Table1. Results of the Databases

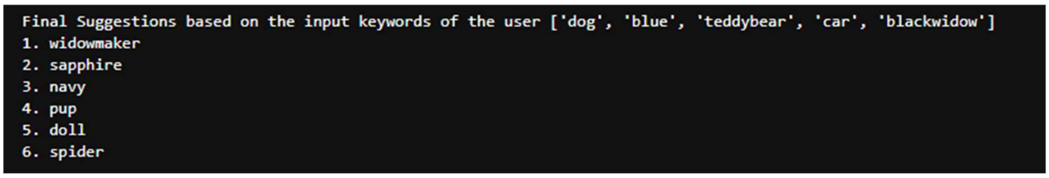

Figure 2. Showcasing final keyword suggestions based on user input.

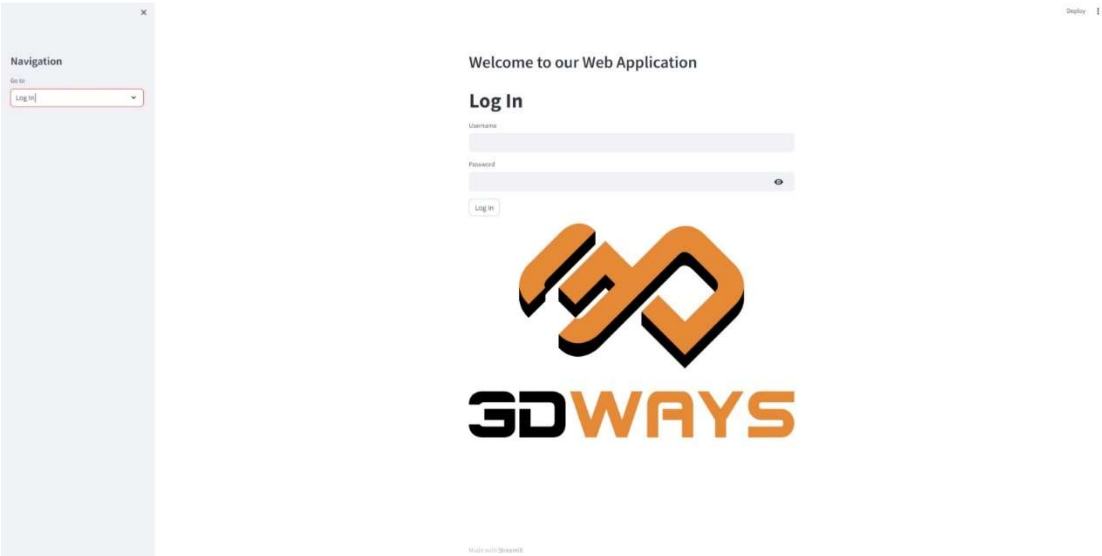

Figure 7. Web application Log in page

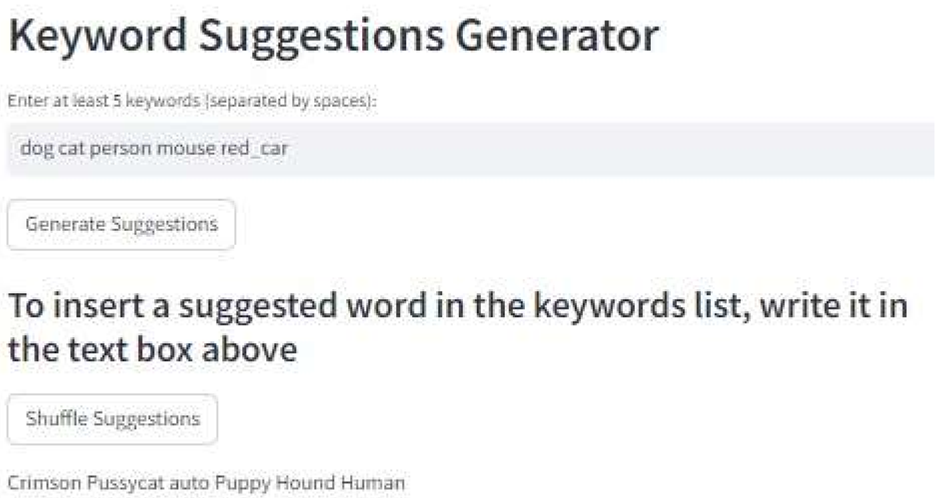

Figure 8. Key word recommender system

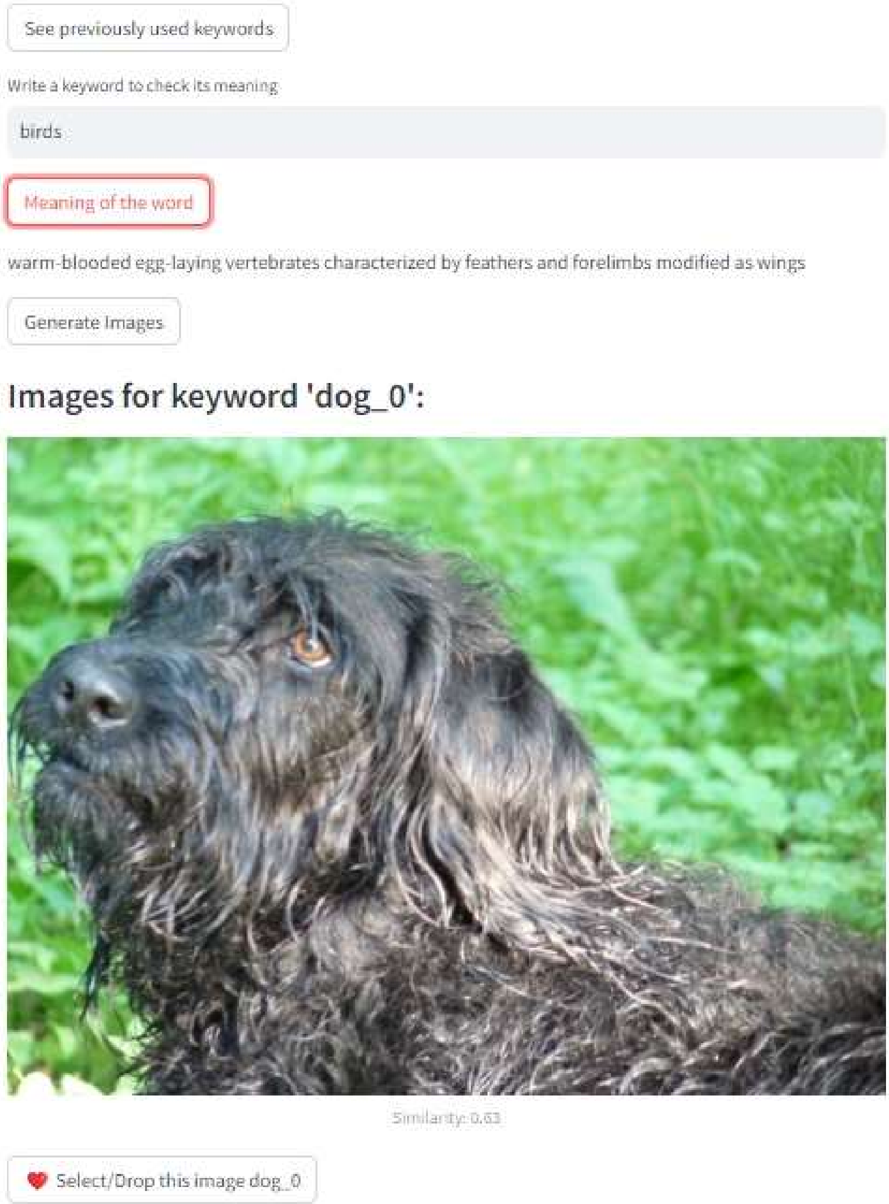

Figure 9. Suggested images based on the key words inserted

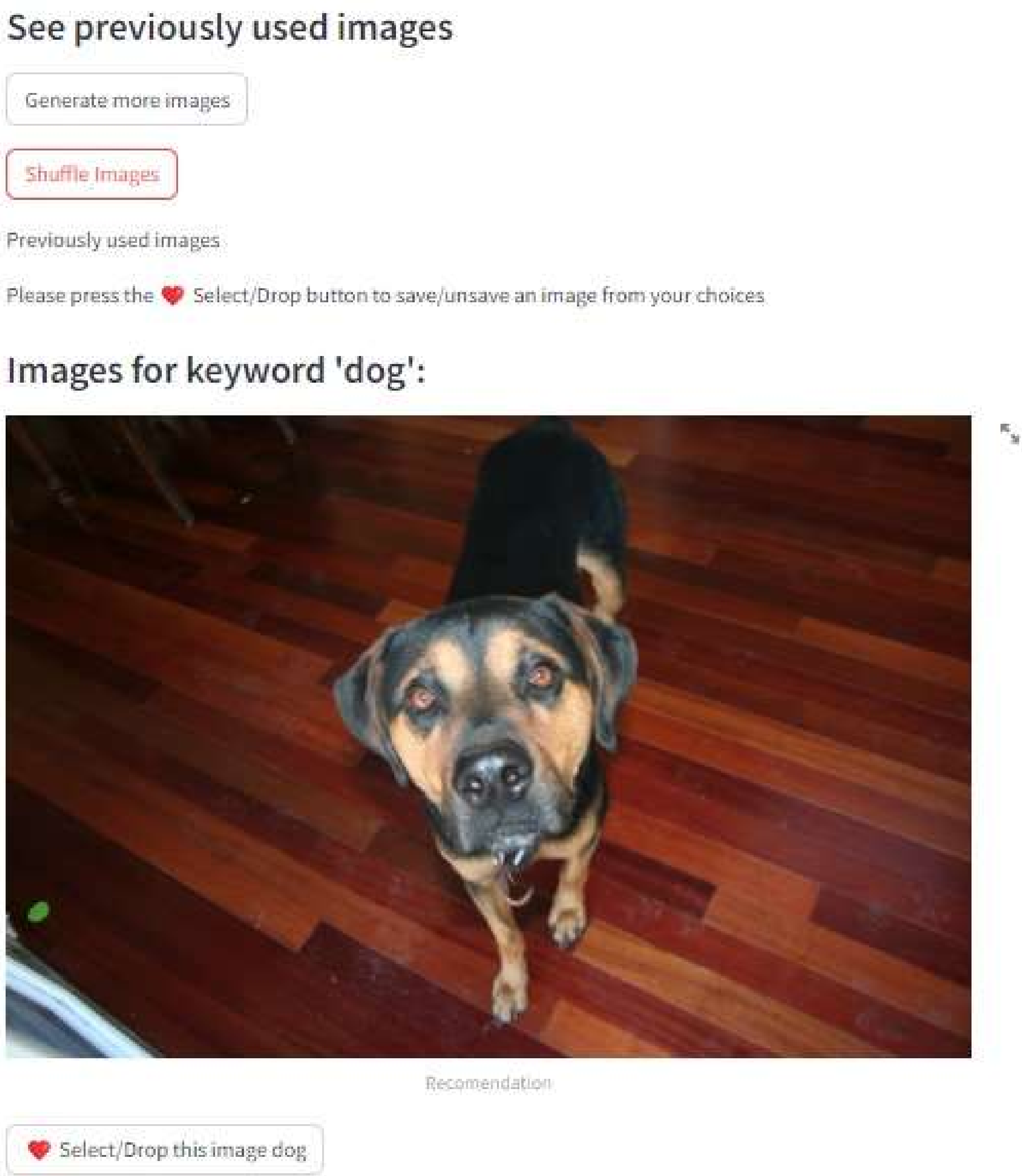

Figure 10. Suggested images based on previous interactions.

Figure 11. Image modifying section.

Figure 12. Pop up with image selected to gather segmentation coordinates.

Figure 13. Mood board generation section

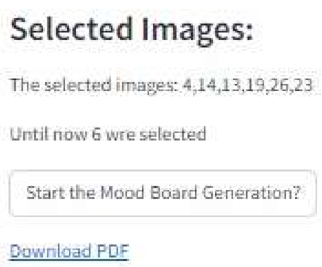

Figure 14. Mood board generation section

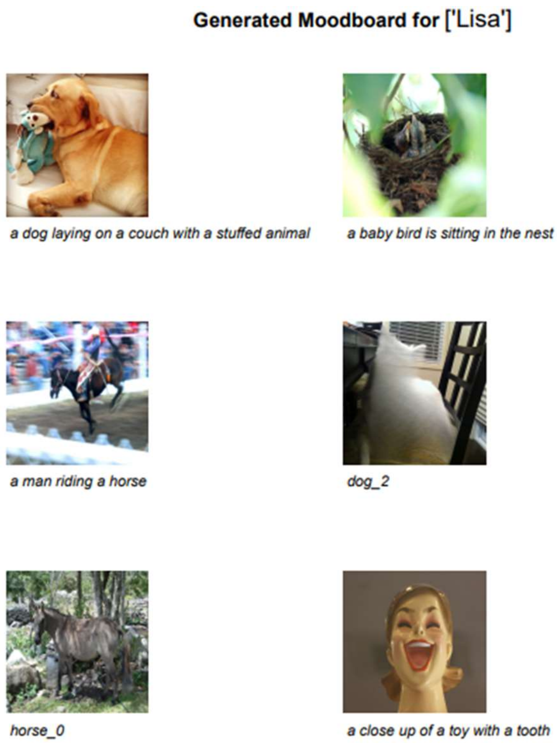

Figure 15. Mood board example