Using graphics to assess wellbeing in a conversational user interface

Andreas Achilleos
Supervisors: Garrett Allen, Ujwal Gadiraju, Derek Lomas, Willem van der Maden
EEMCS, Delft University of Technology, The Netherlands

June 19, 2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Abstract

With the COVID-19 pandemic testing humanity worldwide in unforeseen ways, wellbeing assessment has stepped to the foreground of individual health status. Conversational User Interfaces (CUIs) prove promising as an assessment tool, but lacking the means to retain users engaged during the process. This research aims to explore a solution introducing a graphical modality to a CUI, that improves user experience during wellbeing assessment. A psychological projective testing technique was researched, the H-T-P test, which exposes one's personality, mood and feelings through drawing. A system was designed that utilizes a simplified version of the test and image classification to provide a new means of wellbeing assessment. The system's performance and usability show that such an approach is indeed feasible. This paper describes the design process of the system, the resulting prototype and possible improvements for future implementations.

1 Introduction

The COVID-19 pandemic has challenged humanity in unprecedented ways, and its repercussions reduced human wellbeing around the world; which made staff and member wellbeing more significant for organizations [1]. "My Wellness Check" [2] is an example of a product-service, assisting universities to assess student and staff wellbeing through a survey, and take action to improve that communal status.

Assessing wellbeing at this point in time is a lengthy and labour intensive process, due to the required manual analysis of the responses from the user being assessed. For an organization, such as TU Delft, this would be a difficult task to execute at large scale. If the assessment and analysis of wellbeing is a cumbersome procedure then it cannot be carried out for every student, staff, and faculty member. This would mean that TU Delft (or any large organization) would have to selectively carry out an assessment for its community, contradicting the fact that wellbeing is of higher significance.

Solutions have been explored to automate this procedure. One instance was by using smart home technologies to assess cognitive health [3]. An alternative example was by using natural language processing (NLP) to create an AI-based application that helps with the user's mental state assessment [4].

Aside from researchers, the process of assessment is (currently) also taxing for participants as a consequence of the survey's length and the monotonous action of answering questions. This leads to participants ending the survey before completing it and consequently, reducing the quality of the data that is gathered. An additional way to enhance the quality of this data is to enhance the questionnaire experience itself [5]. Improving the experience for the participants will ultimately lead to better wellbeing assessment and thus, better suggestions on how to take action.

1.1 Main contributions

The aim of this thesis and the respective research question is to explore alternative graphical modalities of answering questions related to wellbeing, in the context of a CUI/survey to better user-retainment and experience. Some possible graphical solutions that can be used in this context will be researched, along with what additional benefit (if any) their graphical modalities introduce. Furthermore, it will be tested how these additions can be interpreted and linked to wellbeing assessment in a prototype application that incorporates the aforementioned solutions. This exploration will indicate ways and methods that are suitable to assess wellbeing. The results of this research will be valuable to indicate the feasibility, effectiveness, and advantages of such an approach, so that it can be used in future work more extensively.

Enhancing the assessment of wellbeing in such a way, would move researchers a step closer towards better questionnaire experiences. The end result would be a user-tailored and seamless evaluation of wellbeing. In this way, the assessment could be, for example, incorporated in popular applications and social media platforms to increase accessibility and eventually make more people aware of the importance of their wellbeing.

The remainder of the paper is structured as follows: section 2 provides necessary background information and relative research from literature. The methodology along with the research hypothesis and an overview of the material is found in section 3. Section 4 presents the results of this study, and information regarding responsible research is located in section 5. A discussion about the findings, limitations and future improvements is found in section 6. Finally section 7 concludes this research paper.

2 Background and Related work

This section contains background information that further formulates the problem in question. Various methods of wellbeing assessment are presented through related literature; these methods led to the investigation of projective testing used in psychology. Finally, examples of projective testing are defined, that are suitable to be used in the context of wellbeing assessment in a CUI.

Over the last few years, non-intrusive ways of wellbeing assessment have been explored, evaluated and developed into a prototype. Facial expressions are trivial visible signs of the psychological state of a person which were exploited in previous research [6] to monitor wellbeing. Another example of an unobtrusive assessment of wellbeing is EMOTHAW [7], where one's handwriting and drawings are related to emotional states. Drawing was also related to learning in [8], where students were asked to provide sketches based on the material they are learning to assess their understanding of the topic.

Based on these correlations, it is evident that information can be inferred about one's cognition without constraining the assessment to questions. A projective technique [9] is a test used in psychology which presents the participant with purposefully unstructured and ambiguous stimuli. The goal is to give insight regarding personality traits, true feelings, and motives that stem from the participant's unconsciousness and relate it to their wellbeing status. After researching relative literature, two tests prove to be the most promising when adapted for usage in the context of a CUI/survey:

- **House-Tree-Person Test (H-T-P Test)**

  The H-T-P test was created by J. Buck in 1948 [10]. It was designed to aid clinicians obtain information about the participant's personality traits. By providing a stimulus (drawing something familiar), the participant project parts of their personality and mental state through elements that they incorporate in their drawing on a piece of paper (see Figure 1).

- **Thematic Apperception Test (TAT)**

  The TAT was created by C. Morgan and H. Murray in 1935 [11]. The test was administrated by asking participants to describe a story in 5 minutes, based on a black and white picture (see Figure 2) they were presented with. The story the participant forms provides information about emotions, anxieties, needs, family dynamics, and general cognition. The process is repeated with a series of images designed to expose different aspects from the participant's subconsciousness.

The administration of the H-T-P test marks it as a better candidate to be used as graphical medium for wellbeing assessment. The drawing aspect of the test provides a better incentive for the participant and would prove more user-engaging when compared to other projective tests. An adaptation of the TAT would involve the participant passively clicking or typing through questions about some displayed images. This could lead once again to issues with incomplete user responses, much like a typical survey.

Figure 1: Sample drawing of the H-T-P test.

Source: https://www.semanticscholar.org/paper/\The-H-T-P-test.-Buck/15cfbdca27c1f8c8dd53212348b57e1c60023bc4

Figure 2: Images shown during the TAT.

Source: https://www.psychestudy.com/general/personality/thematic-apperception-test

3 Methodology

This section describes the methods used to investigate the research questions mentioned in section 1. An overview of the research project is specified, along with the hypothesis that supports the research question. Lastly, the material for this study is presented.

3.1 Research hypothesis

The research question posed in section 1 is formulated under the assumption that a graphical modality, such as a drawing exercise, is deemed more interesting than a traditional question-based survey (further backed up by [12]). This research is focused on the feasibility and implementation of the usage of the H-T-P test as a means to assess wellbeing. Any adaptation of the projective tests listed in section 2 needs to be simplified; the goal is to exploit the graphical mediums of the aforementioned tests to assess one's wellbeing, not to provide a psychological diagnosis. Projective tests over time, have been considered controversial and not a representative assessment of the participant's mental state [13], but for the purpose of this study a translated version would suffice. The outcome of the proposed adaptation should not be taken literally as it is merely an exploration of such an approach.

3.2 Approach overview

Following the statements made regarding the characteristics of the H-T-P test in section 2, it provides the best foundation for the adaptation of a graphical solution. The steps listed below comprise the approach to answering the main research question using the H-T-P test:

Adapting the test's interpretation: The H-T-P test is interpreted based on an extensively detailed manual by its creator. The manual explains which aspects of the participant's drawing correspond to what feelings, characteristics, fears etc. Based on simplified versions of the interpretations [14, 15, 16] the drawings can be grouped and the absence/presence of features in the drawing can point to more generalized emotions or mental states.

Adapting the test's administration: The H-T-P test is administrated normally via drawing with a pencil on a piece of paper. The test's duration is approximately 150 minutes. The adapted version of the test must be shorter, and it has to make use of a digital input method such as a touchscreen or a mouse and keyboard.

Drawing classification: After simplifying the aspects of the drawing that have to be recognized, a multi-class image classifier is needed to deduce what the user's input actually means with regards to the generalized interpretation of the H-T-P test. A model must be developed and trained, that can consistently classify the images appropriately.

Obtaining a dataset: Depending on the number of classes that the classifier must be able to recognize, the dataset needed for training grows larger. A sufficiently large dataset of drawings needs to be obtained for each element of the image: House, Tree, and Person. The dataset must have a balanced distribution of class examples and a variety of drawing examples for each element.

Developing an application: After training a model for image classification, an application will be developed. The app serves as a bridge between the algorithm and the participant. The app needs to instruct the participant through the procedure, and present the results of the test after its completion.

3.3 Material

The following subsections contain information regarding the material used to answer the research question for this thesis. The details of translating the concept of the H-T-P test from psychology, to a GUI of a graphical solution are given. The design choices are explained and linked to implemented features. The process was based on the overview previously mentioned in subsection 3.2.

3.3.1 H-T-P test adaptation

As it was previously mentioned, the solution is based on the H-T-P test by J. N. Buck. The interpretation of the test was greatly simplified by separating the three main elements of the test, namely the house, the tree, and the person drawings. For each of the three elements there is a further subdivision into categories of mental states that can be inferred from that respective element as seen on each row in Table 1. The absence or presence of features in the drawing of each of the elements dictates which category the drawing belongs to. Refer to Table 3 for an analysis of which features correspond to which outcomes. The drawings themselves are also simplified to a doodle-like image. Instead of 150 minutes, the participant has only 45 seconds to draw each one of the elements separately using black color on a white canvas. The time limit was chosen such that the participants would not spend too long on each element, but still have enough time to make a proper drawing. The participants can draw on the screen by clicking and dragging their cursor over the canvas.

	Potential outcomes		
 House	Stress Anxiety	Low self-esteem Introverted Withdrawn	High self-esteem Extroverted Fantasizing

	Potential outcomes		
 Tree	Depression Low energy	Introverted Low ego strength	Extroverted Ambitious High ego strength

	Potential outcomes		
 Person	Depression Low energy	Withdrawal Lack of motivation Boredom	Anxiety Obsession

Table 1: Outcome subdivision for each element of the H-T-P test
Icon source: `https://www.flaticon.com`

3.3.2 Image interpretation

To determine which category (or class) each drawing belongs to, an image classifier is necessary. PyTorch is a python library that provides tools to build deep learning models using neural networks, and has an established community for support for multi-class [17], or multi-label [18] image classification.

Using PyTorch, and a pre-trained model (ResNet-34), an image classifier was built. The first iteration of the classifier was a multi-label classifier that was initially trained on selected drawings from Google's "Quick, Draw" public dataset [19]. The aim was to use multiple labels to encode which features were present or absent in the drawing. However, the performance of the model was poor, with only 50% - 75% of the labels predicted correctly on unseen data.

The next iteration of the model was a multi-class classifier that was trained on manually drawn doodles. The doodles were drawn and labeled using a simple helper tool written in Python. This ensured that the class distribution was balanced and that the model would not overfit. The new model outputs a single class label that the drawing belongs to, rather than an array of labels encoding features. This approach is much more robust and accurate, albeit more simplified than the multi-label classifier. The model accepts black and white images with dimensions of 224×224 pixels.

3.3.3 Python Application

Having a trained classifier is not enough for a complete solution of wellbeing assessment. Software is needed to bridge the gap between the user and the classifier. A python application with a basic user interface was made to serve as an input and output for the classifier. The participant is prompted to draw separately a house, a tree and a person in a total time of 135 seconds (45 seconds for each element). The algorithm in its totality will classify the individual images and assign each one a label. The result is then shown to the participant in the form of sentences that indicate how they could be feeling based on what they drew. More details regarding the resulting application in the following section 4

4 Results

In this section further details about the resulting system are presented, and some values that indicate the overall performance of the model. Furthermore, some insight into the final functionality of the application is given, alongside an explanation regarding the creation of the dataset.

4.1 Model performance

As it was previously stated in subsection 3.3.2, the classifier was made using PyTorch and a pre-trained model, ResNet-34. To improve the performance of the algorithm overall, three separate models were trained for the house images, the tree images, and the person images respectively. The dataset was split randomly for all of the models using 80% for training and 20% for validation. The learning rate ($lr = 0.0001$), batch size ($batch\_size = 4$), and loss function ($CrossEntropyLoss$) were kept identical for all models. Table 2 shows the performance of each model for a random split 80-20 split on the same dataset. The models were then saved so that inference can be done at a fraction of the time when used in combination with the application. The datasets used, as well as the classifiers themselves can be found here.

House			
Model	**Epochs**	**Accuracy**	**Dataset size**
house_model_1	10	95.37%	
house_model_2	12	99.07%	538
house_model_3	15	96.30%	
Tree			
Model	**Epochs**	**Accuracy**	**Dataset size**
tree_model_1	10	99.07%	
tree_model_2	12	99.07%	514
tree_model_3	15	100.0%	
Person			
Model	**Epochs**	**Accuracy**	**Dataset size**
person_model_1	10	100.0%	
person_model_2	12	100.0%	542
person_model_3	15	100.0%	

Table 2: Performance of the models on random splits of the same dataset

4.2 The dataset

In subsection 3.3.2 it was said that Google's Quick, Draw public dataset was deemed inadequate for this use case. The dataset contained images far too small (28×28 pixels) and thus, not enough details could be recognized by the classifier. The solution was to create a dataset using manually drawn images for a house, a tree, and a person. The images were created in a way that all features that were tested for, were present and absent in a substantial amount of samples. This ensured that the classifier would be able to correctly identify if an image contains the appropriate features or not, hence the really high accuracy of the model. Due to the great amount of time required to create the dataset, to increase the accuracy of the model the dataset was extended by mirroring the images horizontally. This created "new" images as seen by the classifier and effectively doubled the dataset size.

4.3 The application

In subsection 3.3.3 it was mentioned that a python application was created to prompt the participant to draw, and to display the outcome of the test's adaptation. Figures 3 and 4 show what the drawing screen and the result screen look like respectively.

The participant is asked to draw individually a house, a tree, and a person in this order. The time limit for each drawing is set to 45 seconds after which the participant can no longer draw unless they reset the timer and start over. After the last drawing of a person, clicking on the "Next Step" button will start the model evaluation. The current drawing window will close and the result window will immediately pop up.

The result window will show the outcome for each of the drawings starting with the house. At the top, it is shown what mental states/personality traits are associated with each outcome. The box bellow provides some suggestions from the perspective of a fellow student that could improve the wellbeing of the participant, based on what the aforementioned outcome was. The participant can click on the "Next" button to transition to the next screen to view the remaining results.

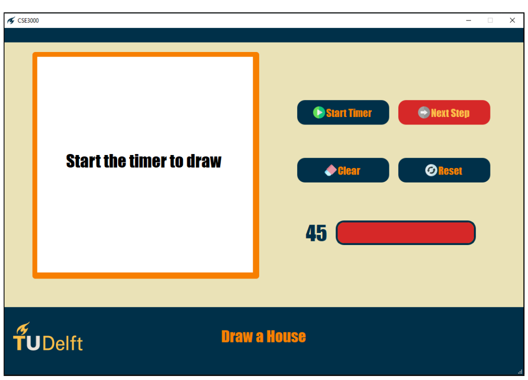

Figure 3: The drawing screen of the application

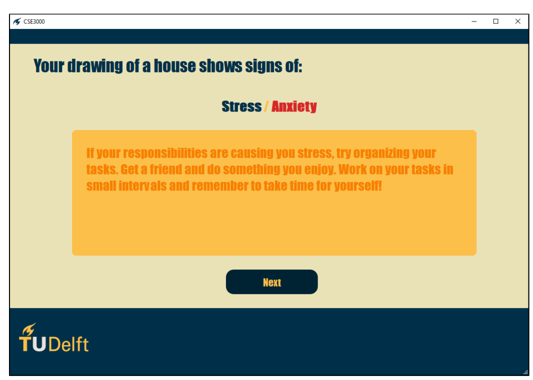

Figure 4: The result screen of the application

5 Responsible Research

This section will reflect on the ethical aspects of this research. Firstly, a look into the ethical implications, and last an assessment of the reproducibility and scientific integrity that this research poses.

5.1 Ethical implications

Naturally, wellbeing assessment involves recognizing one's feelings, personality traits or mental health states. The correct recognition of these features is essential; a participant could be particularly vulnerable if confronted with a false assessment. Consequently, this makes the data used for assessment sensitive, hence why the privacy of a user is equally important.

For the context of this research, an exploration of wellbeing assessment through a graphical medium was the primary focus. A transformation of a projective test used in psychology, the H-T-P test, formulated the basis for this assessment. This raises the question about the validity of this transformation - can the designed system assess one's wellbeing status sufficiently well.

As it was previously stated in subsection 3.1, the system does not output by any means a psychiatric diagnosis; hence, the results are merely a suggestion to be considered by the participant to improve their wellbeing status. The experiment that is described in detail in subsection 6.2 is a validity verification method for the system that was designed in this research.

Regarding the privacy of the participants, the designed system does not store any sensitive information such as name, age, or gender. The current system architecture saves any input from the user only over the span of run-time of the application. Thus, the results and the input for a single run are deleted immediately after the termination of the application.

5.2 Reproducibility

The results that are reported in this paper were generated solely with the available material in this GitHub repository. The models that the current system implementation uses are pre-trained and during run-time only inference is performed. This means that if the same image is fed through the application multiple times the output will remain unchanged. Additionally, the resources used to train the classifiers (datasets, model parameters, code base) will be made publicly available.

5.3 Scientific Integrity

A significant fraction of this research was formed on the basis of previous research and related work. All material used was publicly available, or accessible through TU Delft resources; references are provided to the source of the original author or creator.

6 Discussion and Limitations

The focus of this study was to explore a solution for wellbeing assessment using graphics in a CUI, with the goal being to provide a better experience and increase user engagement. Even with an exploratory approach, the results of adapting a psychological projective test as a solution, prove to be promising. The performance of the models and the usability of the prototype application are signs that a graphical solution for wellbeing assessment is worthwhile developing. While no experiments were conducted in the span of this research project, the verification of the underlying hypotheses and assumptions is feasible, now that the solution posed in this study can be used as a starting point.

This section further discusses some design choices of this project through experiments that can additionally reinforce key elements that form the basis of this study. Furthermore, future improvements and considerations are presented, that can improve the quality of such an approach to wellbeing assessment.

6.1 Research hypothesis experimentation

Following the research hypothesis that was formulated in subsection 3.1, this research was based on the fact that a graphical modality would prove to be more engaging than a normal survey when assessing one's wellbeing. Here, an experiment is proposed that can confirm or reject this original hypothesis.

To properly assess if the suggested graphical solution to assessing wellbeing is preferred over a traditional survey, a quantitative metric can be used. Such metrics include the User Engagement Scale (UES) [20], and NASA's Task Load Index (TLX) [21]. In a controlled setting, a number of participants completes the assessment through both tools; the proposed graphical solution, and a traditional survey. The order of completion would be randomized so that biases towards one approach are omitted. Immediately after the completion of the assessment the participants would be asked to fill in the details for the UES or the TLX, so that a score can be generated based on their experience. Trivially, if the scores of the proposed graphical solution are higher, then the hypothesis is confirmed.

6.2 Validity of the H-T-P test adaptation

In 3.3.1 an overview was given of how the H-T-P test was adapted to be used in the context of a CUI. The original H-T-P test was designed to be used as a psychological evaluation and its administration in a normal setting is much more elaborate than the graphical solution that this research explores. For the context of this project, an assessment of that calibre was not necessary; hence why the adaptation is much simpler. This raises the concern of how well this adaptation can capture the participant's mental state and feelings. The experiment described here would verify that the adaptation performs as intended.

In a controlled setting, a sufficiently large number of participants would use the adapted H-T-P test as normal. After completing the drawings and receiving their results for each one, the participants would be asked a small number of questions. For every drawing there are three possible classifications, as seen in Table 1 and in more detail in Appendix A Table 3. The participants would be asked to answer if that outcome is representative of their mood and feelings and if not, select which outcome in that subset would be a better fit for them. If no option from that subset suits the participant, then they indicate that through another option that there is no suitable choice. This would show that the adaptation is too restrictive and should be expanded to outcomes with greater variety.

A similar experiment can be conducted to compare the solution in this research to the H-T-P test directly. Projective tests, such as the H-T-P test, were deemed controversial and inaccurate, as it was previously mentioned in subsection 3.1. Since the system that was designed for this research project is based on a projective test, similar comments can be made about its accuracy and usefulness. If the original H-T-P test format proves to be inadequate for the case of an arbitrary participant, then the system that is based on that will most likely also be inadequate. To investigate this aspect, the aforementioned test can be expanded to provide further insight. In an ideal scenario, the participant would perform the H-T-P test properly with the presence of a trained psychologist. Subsequently, they would perform the adapted version of the test that was created for this research project. If a resemblance exists between the two means of assessment, then the adaptation was successful. It could very well be the case that the H-T-P test fails to capture the state of a participant entirely, hence why a sufficiently large number of participants is necessary for this experiment.

6.3 Dataset size and quality

The datasets used for the training of the classifiers contain combined nearly 1600 images. Due to the fact that the images had to be created manually, having a larger dataset was not feasible with the given time constraints of this study. Each image was drawn in a way that incorporates various combinations of features that correspond to different outcomes from the H-T-P adaptation. The classifiers are then able to detect these features and properly assign the correct outcome to the image that the participant draws. Since the images used for training were drawn by a single person, the classifiers could have difficulties generalizing and recognizing features in the input images if they are drawn drastically different. This could be mitigated with a larger dataset that includes more noise and variety in the way the images are drawn; ideally the images would be drawn by multiple individuals.

6.4 Future work and considerations

This study explored the possibility of utilizing the graphical modality of a projective test to provide a wellbeing assessment. With the results being primarily positive, such an approach proves to be feasible. Nevertheless, with the given time constraints there is still room for enhancement. Here, some aspects of the designed system are identified that can better the quality of the solution proposed in this study.

6.4.1 Python vs Web-application

The app that was designed as a user interface for the participant's input was created in Python; this limits the usage of the system that was designed to a computer. A web application as a front-end would be preferred, since the interface would be significantly easier to develop for additional platforms such as a smartphone.

As a result of security protocols, a web-app cannot save a file (in this case an image) to a user's system without prompting them with a pop-up window [22]. Thus, a web-app interface that saves the participant's drawings for inference without asking for permission was impossible to create. A viable solution to this issue, is to run the models in the web browser using ONNX [23] and ONNX Runtime [24]. The models that were created with PyTorch can be converted to an ONNX model and subsequently ran in the browser using ONNX Runtime. Therefore, the entire system can function in a web browser providing more flexibility through the support of more devices. Such an approach was outside the scope of this study, but would constitute a substantial future improvement.

6.4.2 Continual Learning models

This research project presents a system that uses pre-trained models for inference; the idea of such an approach is to train a model sufficiently on a dataset so that it can label unseen data accurately and reliably. In the context of the proposed adapted H-T-P test, the models in question are labelling drawings. A drawing of a house, a tree, or a person, can vary drastically from person to person. A pre-trained model may not be able to identify the absence or presence of features - windows on a house for example - if the input image is far different than the samples the model was trained on. A system that makes use of Continual Learning (CL) [25] can overcome this problem. Rather than using the models just for inference at runtime, the models retrain as new images get drawn by participants. The drawings have to be assigned a correct label so a similar approach could be followed as subsection 6.2, where the participants are asked after completing the assessment to choose from the subset of outcomes (or labels) the option that best represents them. The accuracy and reliability of such an approach would keep improving the more the system is being used. However, the degree of complexity for this solution is far greater than the exploratory approach of the proposed system for this study. Nonetheless, continual learning is a noteworthy consideration for a larger and more robust future implementation.

7 Conclusion

This research was aimed at finding a graphical approach to wellbeing assessment and designing its implementation in the context of a CUI/survey. Literature research led to the adaptation of a psychological projective testing technique, the House-Tree-Person test. The test exploits human subconsciousness through the usage of ambiguous stimuli (House, Tree, Person) to gauge one's personality traits, mood or feelings. A simplified version of the H-T-P test was created to be used in the context of a CUI. After obtaining a means to assess wellbeing through drawing, a prototype system utilizing trained image classifiers was created as a proof of concept for such an approach to wellbeing assessment. Moreover, means of (in)validating the soundness and reliability of this solution are presented through experiments regarding the underlying hypotheses and assumptions. Some worthwhile future considerations are highlighted that can further improve the quality of the proposed system.

The introduction of a graphical modality to a system for wellbeing assessment proves to be a viable option. An implementation of a hybrid setup combining graphics and traditional means of assessing wellbeing could be the next step towards better user experience and increased awareness of personal wellbeing.

A Appendix

House					
Outcome 1	**Features**	**Outcome 2**	**Features**	**Outcome 3**	**Features**
• Stress • Anxiety	• Excessive amount of smoke • High details • Shaded roof	• Low self-esteem • Introverted • Withdrawn	• Absence of door • Absence of windows • Door located high above base line • Missing a chimney	• High self-esteem • Extroverted • Fantasizing	• Very large door • Very large roof • Very large windows • Open door

Tree					
Outcome 1	**Features**	**Outcome 2**	**Features**	**Outcome 3**	**Features**
• Depression • Low energy	• Downward facing branches • Thin trunk • Absence of leaves	• Introverted • Low ego strength	• Short or no branches • Thin and small trunk	• Extroverted • Ambitious • High ego strength	• Thick trunk • Upward facing branches • Large number of branches

Person					
Outcome 1	**Features**	**Outcome 2**	**Features**	**Outcome 3**	**Features**
• Depression • Low energy	• Miniature size • Lack of detail	• Withdrawal • Lack of motivation • Boredom	• Overly simplistic • Significant lack of detail • Stickfigure-like	• Anxiety • Obsession	• Highly detailed • Facial characteristics present • Proportional dimensions of limbs/body

Table 3: Analysis of image classification based on absence/presence of features

References

[1] OECD, *COVID-19 and Well-being*, 2021. [Online]. Available: https://www.oecd-ilibrary.org/content/publication/1e1ecb53-en

[2] W. van der Maden, D. Lomas, S. Fonda, and P. Hekkert, "Designing a feedback loop for community wellbeing," April 2022, the paper was shared with the group for literature research.

[3] P. N. Dawadi, D. J. Cook, M. Schmitter-Edgecombe, and C. Parsey, "Automated assessment of cognitive health using smart home technologies," *Technology and Health Care*, vol. 21, no. 4, pp. 323–343, 2013.

[4] R. R. Dedeepya, L. C. Koneru, A. Lakshmi, and G. K. Kumar, "Stress management system," in *2022 IEEE Delhi Section Conference (DELCON)*, 2022, pp. 1–5.

[5] J. Baumgartner, N. Ruettgers, A. Hasler, A. Sonderegger, and J. Sauer, "Questionnaire experience and the hybrid system usability scale: Using a novel concept to evaluate a new instrument," *International Journal of Human-Computer Studies*, vol. 147, p. 102575, 2021.

[6] F. Chiarugi, E. Christinaki, S. Colantonio, G. Coppini, P. Marraccini, M. Pediaditis, O. Salvetti, and M. Tsiknakis, "A virtual individual's model based on facial expression analysis: A non-intrusive approach for wellbeing monitoring and self-management," in *13th IEEE International Conference on BioInformatics and BioEngineering*, 2013, pp. 1–4.

[7] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. ClÃ©menÃ§on, and G. Cordasco, "Emothaw: A novel database for emotional state recognition from handwriting and drawing," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 273–284, 2017.

[8] S. Miller, "Using drawings for formative assessment," https://www.edutopia.org/article/using-drawings-formative-assessment, September 2021, Accessed: 14-05-2022.

[9] S. A. Kolakowsky-Hayner, *Projective Technique*. New York, NY: Springer New York, 2011, pp. 2046–2047. [Online]. Available: https://doi.org/10.1007/978-0-387-79948-3_2096

[10] J. N. Buck, "The h-t-p technique. a qualitative and quantitative scoring manual," *Journal of Clinical Psychology*, vol. 4, no. 4, pp. 317–317, 1948.

[11] C. D. Morgan and H. A. Murray, "A Method for investigating fantasies: The Thematic Apperception Test," *Archives of Neurology & Psychiatry*, vol. 34, no. 2, pp. 289–306, 08 1935. [Online]. Available: https://doi.org/10.1001/archneurpsyc.1935.02250200049005

[12] R. Bommireddipalli, "Council post: A data-informed, design thinking approach for user retention," March 18 2022, Accessed: 30-05-2022. [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2022/03/15/a-data-informed-design-thinking-approach-for-user-retention/

[13] R. G. KLEIN, "Questioning the clinical usefulness of projective psychological tests for children," 1986.

[14] natsmagbalon, "House, tree, person," Accessed: 13-05-2022, author name substituted by online username. [Online]. Available: https://pdfcoffee.com/house-tree-person-2-pdf-free.html

[15] nicklan, "Preliminary drawings art all levels," Accessed: 13-05-2022, author name substituted by online username. [Online]. Available: https://pdfcoffee.com/house-tree-person-assessment-pdf-free.html

[16] nesuma, "House tree person interpretation elements," Accessed: 13-05-2022, author name substituted by online username. [Online]. Available: https://pdfcoffee.com/house-tree-person-interpretation-elements-pdf-free.html

[17] A. Verma, "Pytorch[vision] - multiclass image classification," May 9 2020, Accessed: 26-05-2022. [Online]. Available: https://towardsdatascience.com/pytorch-vision-multiclass-image-classification-531025193aa

[18] B. Droste, "Multilabel classification with pytorch in 5 minutes," January 13 2022, Accessed: 14-05-2022. [Online]. Available: https://towardsdatascience.com/multilabel-classification-with-pytorch-in-5-minutes-a4fa8993cbc7

[19] "The quick, draw! dataset," GitHub, 10 2021. [Online]. Available: https://github.com/googlecreativelab/quickdraw-dataset

[20] H. L. O'Brien, P. Cairns, and M. Hall, "A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form," *International Journal of Human-Computer Studies*, vol. 112, pp. 28–39, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581918300041

[21] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, ser. Advances in Psychology, P. A. Hancock and N. Meshkati, Eds. North-Holland, 1988, vol. 52, pp. 139–183. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166411508623869

[22] "The file system access api: simplifying access to local files," June 2 2022, Accessed: 17-06-2022. [Online]. Available: https://web.dev/file-system-access/#security-considerations

[23] "ONNX | Home," Accessed: 12-06-2022. [Online]. Available: https://onnx.ai

[24] "ONNX Runtime | Home," Accessed: 17-06-2022. [Online]. Available: https://onnxruntime.ai

[25] Y. Ettun, "How to apply continual learning to your machine learning models," July 11 2019, Accessed: 16-06-2022. [Online]. Available: https://towardsdatascience.com/how-to-apply-continual-learning-to-your-machine-learning-models-4754adcd7f7f