



Degree project in Technology

First Cycle 15 credits

Fine-tuning a LLM using Reinforcement Learning from Human Feedback for a Therapy Chatbot Application

SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

DESIRÉE BILL

THEODOR ERIKSSON

FINE-TUNING A LLM USING REINFORCEMENT LEARNING FROM HUMAN FEEDBACK FOR A THERAPY CHATBOT APPLICATION

Desirée Bill, Theodor Eriksson

Svensk titel

Finjustering av en LLM med hjälp av förstärkande inlärning från mänsklig återkoppling (eng. RLHF) för en Psykolog-chatbot applikation

Abstract—The field of AI and machine learning has seen exponential growth in the last decade and even more so in the recent year with the considerable public interest in Large Language models (LLMs) such as chat-GPT. LLMs can be used for several purposes, but one possible application would be fine-tuning a model to perform a particular function in a specific field. The goal is therefore fine-tuning a LLM in the field of psychology using a new method called Reinforcement Learning from Human Feedback to determine if it is a viable method in such cases. The theory behind LLMs and RLHF as well as the ethical perspective on developing a psychological AI is presented. Previous studies on both RLHF and AI in psychology are presented, showing the goal is feasible. Then the method is explained for both training and evaluating the model which is done by comparing a pre-trained model with the fine-tuned one. The study is considered scientifically relevant as RLHF has been used to fine-tune LLMs earlier, but has not been done with the intent to make it more specified in a field. The result did not show any clear difference between the pre-trained and the fine-tuned model therefore, more tests are required. However, with the limitations regarding hardware, time to train, and available data, there is much improvement needed for future studies. An ethical framework applied to a digital psychology assistant is discussed and a suitable introduction to the market and division of responsibilities is proposed.

Sammanfattning—Området AI och maskininlärning har sett exponentiell tillväxt under det senaste decenniet och ännu mer under det senaste året med det stora allmänintresset för stora språkmodeller som chat-GPT. Stora språkmodeller kan användas till flera saker där en möjlig tillämpning är att finjustera en modell för att fylla in viss funktion inom ett specifikt yrke. Målet med arbetet är därför att finjustera en språkmodell inom området psykologi med hjälp av en ny metod kallad Reinforcement Learning from Human Feedback för att undersöka metodens tillämplighet. Teorin bakom stora språkmodeller och RLHF samt det etiska perspektivet på att utveckla en digital psykologi assistent förklaras. Därefter presenteras tidigare studier om både RLHF och AI inom psykologi som visar att målet är genomförbart. Metoden för att både träna och

utvärdera modellen förklaras som görs genom att jämföra den förtränade modellen med den finjusterade. Studien bedöms som vetenskapligt relevant även fast RLHF har använts för att finjustera språkmodeller tidigare, har det inte gjorts med målet att finjustera en språkmodell till ett visst yrke. Resultatet visade inte på någon tydlig skillnad mellan den förtränade och den finjusterade modellen, därför krävs fler tester krävs. Men med de begränsningar som fanns gällande hårdvara, tid att träna och tillgänglig data är det mycket som kan förbättras i framtida studier. Det etiska ramverket applicerat på en digital psykologi assistent diskuteras och en lämplig introduktion till marknaden och ansvarsfördelning föreslås.

Index Terms—Ethics, Fine-tuning, Large Language Models, Machine learning, Psychology, Reinforcement Learning from Human Feedback,

I. INTRODUCTION

IN recent years the field of machine learning and artificial intelligence (AI) has grown rapidly. Natural Language Processing (NLP), which is how computers can understand and process language and text like humans, has also developed considerably, resulting in tools such as ChatGPT. Large language models (LLM) such as ChatGPT can be trained using several different methods; recently, a new method for fine-tuning LLMs, named Reinforcement learning from human feedback, has emerged, RLHF for short.

RLHF is an approach to reinforcement learning that uses human feedback to improve the performance of an LLM. By incorporating human feedback, the model can learn from specific examples and adapt to a particular task, making it useful for applications such as customer service, language translation and much more [1]. One possible application for RLHF would be to specialise a LLM to perform a profession and this thesis will try to train an LLM to act as a virtual psychologist assistant. With the RLHF method, the virtual psychologist assistant can be fine-tuned to address the needs of individual patients better and garner better responses. This could potentially improve the field of psychology but also show that RLHF can be used to turn general LLMs into specialised tools in a chosen field.

With technological advancements comes great responsibility to ensure user safety and ethical practice. In such a delicate field as psychology, it becomes extra important to take this into account as it otherwise could lead to detrimental consequences.

A. Project Background

This study was part of a bigger project with the purpose of developing a Cognitive Behavioural Therapy-aimed AI chatbot

Desirée Bill is pursuing a degree in Industrial Engineering and Management with a specialisation in Computer Science and Communications at KTH Royal Institute of Technology, Stockholm, Sweden.(e-mail: dbill@kth.se). My main responsibility was to structure the dataset and together with Theodor Eriksson, we co-authored the thesis.

Theodor Eriksson is pursuing a degree in Industrial Engineering and Management with a specialisation in Computer Science and Communications at KTH Royal Institute of Technology, Stockholm, Sweden.(e-mail: theodor@kth.se). My main responsibility was to structure the training and together with Desirée Bill, we co-authored the thesis.

specialised in psychology. The mentor for this project was Birger Moëll, a licensed psychologist and PhD student at Speech Music and Hearing at KTH. He has developed a demo working as an initial chatbot, supervised the investigation, and provided help and resources when needed. Others involved include another psychologist and four peers at KTH who are writing their bachelor theses related to the topic. Both the psychologists involved in the project were the ones evaluating this investigation.

This thesis is of interest for the open-source community for machine learning which is constantly working to improve old and find new methods to train LLMs and advance the field of machine learning. Therefore, using RLHF to specialise a model to a field becomes relevant for them. For the same reasons, the research community for machine learning is a stakeholder in the project.

As this thesis investigates the ethical and social issues relating to the development of AI in the psychology field, it concerns the government agencies of regulation and society as a whole from a user perspective.

B. Goal

We aimed to examine the effects of fine-tuning LLMs with RLHF applied to a specific field, in this case, the psychological and mental health field. Can the training method be applied in a specific area and generate relevant results for that field? The results may show it can be possible to extend this method to other fields such as healthcare, educational purposes and more. It is important to highlight that this thesis was not made with the objective of creating a substitute for a licensed psychologist or expert. But instead to investigate the effects of RLHF and see it implemented within the field of psychology and counselling. This thesis aims to examine the possibility of innovation and implementation of AI within psychology and thus inspire ethical practice in psychology.

C. The Ethical Aspect

Fine-tuning large language models has risks, especially for text generation and assisting technologies. As the models are trained on datasets scraped from sites on the internet, including banned ones with harmful content, the responses to prompts can contain toxic language. These prompts do not necessarily need to be of toxic nature but still receive abusive output. Another issue is that data on websites used in the models training dataset is biased from a young male perspective from a developed country as that profile has been overrepresented in the texts compared to other groups in the world [2]. By using these language models that are built on this fundamental flaw, it limits what the trained model can be used for, as the results for this project may display responses along the same lines. As the chosen field for this project is mental health, the abusive output is the opposite of what is intended.

Another ethical aspect that must be considered when developing a psychological bot is who is responsible for the bot's actions. To become a licensed psychologist in Sweden today, a five-year education at university is required and, after graduation, one year of practice under the supervision of an

experienced psychologist [3]. Using a bot as a supplement or substitute for a licensed professional psychologist creates the problem of who is responsible for the bot's actions. The same problem exists in other similar fields that experience automation, such as robotic surgery and self-driving cars. In robotic surgery, it can be problematic to determine who bears responsibility when an operation goes wrong. Depending on what went wrong, the doctor, the hospital or the robot manufacturer could be held liable, making it a complex problem [4]. For self-driving cars, currently, it is the driver who is responsible if there is an accident, [5] but in such an area where we can soon see fully self-driving cars, it will become increasingly unclear who is liable, the driver, the manufacturer or someone else? Similar dilemmas exist for a psychologist bot. This year there have been reports of individuals committing suicide following the use of chatbots; for instance, a Belgian man used a chatbot extensively for six weeks to discuss the climate crisis and ultimately ended his life after being encouraged by the chatbot [6]. Who is responsible if the bot generates toxic responses? If the answers harm the person seeking help, it can have severe consequences on people's lives. Are the developers liable, or is the company giving access to the bot or someone else? It is an issue that must be investigated to be ethically justified to develop such a tool.

D. Scientific Questions

What impact does Reinforcement Learning from Human Feedback (RLHF) have on a Large Language Model's (LLM) responses to prompts regarding mental health?

How can the chatbot be implemented ethically in psychological practice, and how is AI innovation within the psychological field affected by regulation frameworks?

E. Problem Definition

The output was evaluated as to how appropriate it is regarding support, listening, validating, insight and guidance, impartiality and respect. To measure this, a list of criteria has been developed with the help of licensed psychologists that will be used to evaluate the LLM before and after RLHF. The challenge with this investigation is to compare the fine-tuned language models to how well they apply to the criteria. The models will be evaluated with the help of psychology students from Karolinska Institutet (KI) and Stockholm University (SU), filling out a survey with a point system as to how they perceived the conversation.

F. Expected results

The evaluation of the results will be qualitative, where a survey will be given to the individuals using the pre-trained LLM and the LLM with RLHF training. Using RLHF on a suitable LLM is expected to result in the model working better as a psychologist. To measure this, an evaluation where the models were compared was carried out.

II. THEORETICAL BACKGROUND

A. Large Language Models

A large language model (LLM) is a machine learning model that generates output to a given input and can be used for many NLP tasks. The input can be different depending on the task the LLM has been trained for, but for a LLM trained to be a chatbot, the input is sentences, and the output is an answer. A LLM can be trained and based on datasets composed of large corpus text scraped from websites but might be based on other large corpora. [7].

Transformers are a type of LLMs which use a unique technique called self-attention. Transformers are a common artificial neural network architecture for modern language modelling and use attention blocks, becoming state-of-the-art for NLP tasks. A neural network consists of an input layer which is a vector of scalar values, then hidden layers and finally an output layer. The connections, called edges, between the nodes of the hidden layers have assigned weights. The weights are updated during the training of the model and determine the output of the model. When the model receives data as input it goes through each hidden layer where it is multiplied by the weights, summed and then passed through an activation function, typically sigmoid. The multi-headed self-attention and positional encoding that transformers use allow the context of the word rather than proximity to the word to be relevant in a sentence. This enables them to handle distant information and relationships between the tokens in the input. This is why this structure is used in modern LLMs since the LLMs can remember earlier parts of a conversation and thus work much more intelligently [2].

The transformer works in several steps, the first one being the encoder. The encoder works by tokenising the input and creating an embedding vector representing the input and the position it had in the sentence. With the embedding and three matrices called The Query, The Value and The Key, a self-attention score can be calculated, which measures how much focus to place on different parts of the sentence and different tokens. A self-attention head was just described, and it has a problem. A self-attention head gives the same output for two sentences as long as the sentences contain the same word, irrespective of order. To solve this problem, several self-attention heads can be combined into what is called Multi-head Attention, which is the base for the transformer architecture [8].

The model sizes typically range from 10B-280B machine learning parameters. The more parameters the large language model has, the better it usually performs. GPT-4, the successor to chat-GPT (GPT-3), reportedly has one trillion parameters, while chat-GPT has 175B [9].

B. Fine-tuning

Fine-tuning further trains the models on smaller adjustments to get a more desirable output [2]. LLMs used in machine learning are very large, where the number of parameters they are trained on can range up to hundreds of billions of parameters. Fine-tuning all these parameters can be costly for computation and storage and time-consuming.

A library called Parameter Efficient Fine-Tuning (PEFT) has several techniques that can be implemented to scale down the number of parameters to be fine-tuned, which decreases time and complexity when training [10]. One example of PEFT is through Low-Rank Adaption of Large Language Models, which is used in this thesis by freezing the weights, thus minimising the query and value layers in the attention matrices with fewer parameters than the initial model [11].

C. Reinforcement Learning from Human Feedback

For this thesis, the effects of reinforcement learning from human feedback (RLHF) on a LLM were investigated. RLHF is a method used to tailor a language model trained on a corpus of data to the user's preference. RLHF can be structured into three steps; pretraining a language model, collecting datasets to train the reward model and finally, fine-tuning the model with reinforcement learning. For this project, the language models are already pre-trained, meaning that they have processed a large corpus making them able to represent the meanings of words from their context. The pretraining enables the model to fine-tune - the better the pre-trained model, the easier it is to fine-tune.

With human feedback, the goal is to capture human sentiments in the LLM's responses. The objective of the reward model is to map the model's output to a scalar reward representing human preference; it can be from ranking using ELO scores or a point system. The model must return a numerical value because Reinforcement learning (RL) uses scalar rewards signals. When the reward increases over training iterations, that is the indicator that the model is learning to generate more preferred responses. The inputs for a reward model are responses to prompts that are ranked in relation to each other. The ranking methods differ depending on the reward model, but one of the more used ones is an ELO system where each answer is ranked in relation to all the other answers. Another common ranking system is where two answers are generated for each prompt, and one is preferred out of the pair. The output of the reward model is then a scalar reward for each response and the loss is then calculated between the human label and the scalar reward and used to update the reward model.

Lastly, reinforcement learning (RL) is used to fine-tune the initial language model concerning the reward model which can be seen in Figure 1 [1]. Reinforcement learning trains an agent in an environment to align its use towards the goal intended by using a policy. The policy is the strategy that the model uses to learn from previous actions; it is a language model that returns a text based on probability distributions as a response to a prompt.

The reward function is a combination of the initial model, the RL policy, the reward model and a constraint on policy shift. A prompt from the dataset is taken and a text from the current fine-tuned RL policy is produced. This text becomes input for the reward model and a scalar reward is returned. The RL policy's probability distributions of the tokens are then compared with the initial models to find the penalty between them with each batch of training. This is to ensure

that the RL policy does not diverge too quickly from the initial model. In this project, a policy-gradient algorithm Proximal Policy Optimization (PPO), was used and is a relatively mature policy. PPO uses small steps to improve the policy when training, for two main reasons; it is known that smaller policy updates are more likely to converge to an optimal solution and a smaller policy update is safer since a large update might result in getting a destabilised learning process which makes improving the policy difficult [12]. More specifically PPO-Penalty was used, which incorporates a Kullback-Leibler (KL) divergence penalty. KL constrains the update by measuring the distance between samples of the distributions of the RL policy and the initial language model, as without KL, there is a risk of fooling the model to display improvements by giving high rewards but not generating better results [13]. The PPO updates the parameters called the update rule which maximizes the scalar reward for the current batch of data.

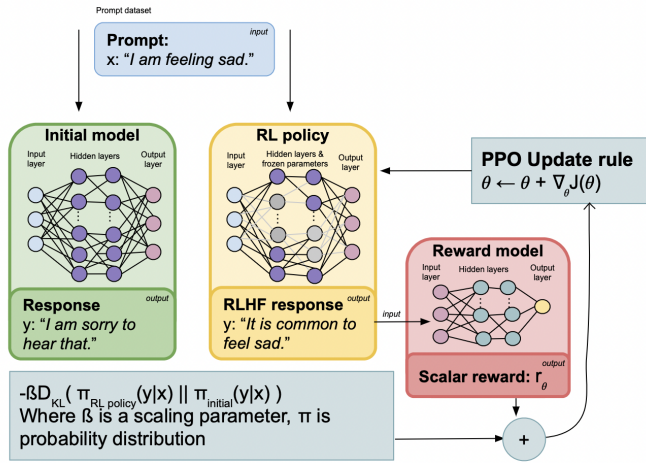


Figure 1. Shows the last step of RLHF, fine-tuning with RL. Adapted from [1]

D. Personal Data

With the introduction of an AI tool within the field of psychology and mental health, sensitive personal data is what the stored data would be comprised of as it would be the input for the system. Personal data is defined as any information connected to a natural person that can be identified through, for example, an identification number or a name [14]. Processing that information safely would be paramount as it is exposed to faults and potential leakage. For instance, in the Cambridge Analytica scandal, the data that was accessed was information on friend networks and activity, such as likes from 50 million Facebook users, to build user identities and influence their behaviours politically [15]. In this case, the content is explicit and not interpreted from clicks since it annotates individuals' thoughts and opinions; the personalities are more apparent and require high protection. The General Data Protection Regulation (GDPR) applies to data processing regarding personal data, and the regulation has it as one of its key issues. It becomes one of the regulations to adhere to regarding the implementation of AI within the field of psychology.

E. Ethics, Transparency and Accountability Framework for Automated Decision-Making

To help sort out the ethical issue with a psychology chatbot, we will use a framework created by the UK government called the "Ethics, Transparency and Accountability Framework for Automated Decision-Making" [16]. The framework was developed to clarify how artificial intelligence should be used correctly, safely and ethically in the public sector. The framework consists of seven points developed in line with guidelines from relevant authorities (for example, the Data Ethics Framework) and legislation. This makes the framework an excellent tool to sort out the ethical issue with a psychologist chatbot. The framework and its simplified steps can be found in Figure 2.

The first step is to prototype and test the system being built so it is fully understood, sustainable and functions as intended. Step two is to have a diverse team when creating the system and thus ensure that the delivered product is fair for all different kinds of users. This step is not as relevant to this thesis compared to a governmental function, but the idea of equality and inclusion is, of course, something that should always be applied. The third step is to be clear about who is responsible for the system. This is an extra important step as misuse of a psychologist chatbot can have serious consequences, and it is an issue that must be well thought over. The next step is to manage personal and sensitive data securely. The fifth step is to be transparent with the users about how the tool works, what it is meant to do and how they can influence them to ensure that the expectations of the tool are reasonable and correct. The next step is to ensure that tools comply with the law. The final step is to make sure to build something that is future-proof to ensure that the tool has a sustainable future [16].

III. RELATED WORK

A. RLHF

In a study by Bai et al. in 2022 [17] the authors used RLHF to fine-tune language models to act as honest, helpful and harmless assistants [HHH]. The study used datasets of more than 86 000 datapoints structured in two categories; harmless and helpful, and showed successful results in fine-tuning a model with a method where the models were updated on a weekly basis following the feedback given by crowd workers online. The final model showed that it was possible to train a helpful and harmless assistant using RLHF and that the models could be trained for specialised skills such as Python coding or summarisation without decreasing performance. In an earlier study by Ouyang et al. [18] in 2021 for OpenAI, human feedback was used to fine-tune a 1.3B parameter LLM, named InstructGPT, and compared to non-RLHF trained LLM GPT3 with 175B parameters. The dataset for the reward model consisted of 33 000 prompts and when the output from the two LLMs was compared in a human evaluation, InstructGPT was preferred despite GPT3 having more than 100 times more parameters. The result of the study showed that RLHF is a promising method for aligning language models with human



Figure 2. Shows the simplified steps of the Ethics, Transparency and Accountability Framework for Automated Decision-Making

intent, which was based on the parameters Helpful, Harmless and Honest.

Guo et al. [19] did further research from Anthropic's HHH Assistant by comparing a fine-tuned RLHF model, ChatGPT, with Human Experts on HHH. This was done by collecting 40 000 questions and answers from both subjects covering different topics, including psychology. From human evaluations, patterns could be drawn, and the future use of language models could be discussed. The authors presented that in the majority of the answers, ChatGPT's completion was considered more helpful than humans, especially in the psychology areas. Other findings were that the fine-tuned model showed less bias, less emotion, and more formal language, and it refused to answer when it was out of knowledge. However, ChatGPT may fabricate facts in its response, making it risky to use in professional consultations.

With the previous studies in mind, we believe it is scientifically interesting to investigate whether it is possible to take the fine-tuning a step further and specialise a LLM to function as a tool in a complex field, in this case, psychology.

B. AI in Psychology

Inkster, Sard and Subramanian address the effectiveness and engagement of AI-enabled conversational agents for mental health. The authors suggest these platforms could help fill a gap between the people needing help and the lack of professionals [20]. Their study was conducted by having an anonymous group of volunteers to use Wysa, an AI-based emotionally intelligent chatbot app, and after messaging, they filled out the Patient Health Questionnaire-9 to report symptoms of depression. From the participants, two types of

users emerged, distinguished by their engagement with Wysa; high and low users. Both user groups showed improvements in the score comparison before and after the study period of three weeks. However, the high user group had a more significant improvement. The study demonstrated promising results as to the implementation of artificial intelligence in mental health but also raised the importance of future research to validate the results.

Kretzschmar et al. raise the social and ethical concerns with conversational agents in mental health support by investigating the possibilities and limitations of the bots Woebot, Wysa and Joy [21]. The article focuses on young people age 14-18 as they are the largest group with mental health issues and the most significant consumer of the digital world. The bots were deemed beneficial for destigmatising mental health issues and could prove to be the first step for users to get help. Limitations include the chatbot not being able to comprehend a user's life history and current situation, which are critical factors regarding mental health issues. This limitation can lead users to feel that the chatbot is not 'listening' to them. Ethical issues regarding privacy concerns, such as access to personal information and protecting users' safety, were also declared. The possibility that the bot results in isolating people in need of help is also introduced and is something the authors consider should be investigated next. It is determined that the bots need to be evidence-based and stressed that they need empirical testing of the clinical effects as well as performing controlled trials. The study evaluated the use of fully automated conversational agents in mental health support, where both the AI technology and the psychology field are relevant to this thesis.

IV. METHOD

A. Dataset

The dataset used for this investigation was a set of questions and answers from Counsel Chat. Counsel Chat is a platform where users can contact verified therapists and ask mental health-related questions. There are 307 therapists from the U.S.A on the platform, and they have licenses as PhD level psychologists, social workers or mental health counsellors. The key advantage of the dataset was that it was an actual interaction between a therapist and a client. The dataset consisted of 31 therapy-related topics and was structured with the question id, question title, question text, the answer text, therapist information (name and speciality), topic, number of views and number of upvotes [22]. From this, the question title, question text, answer text, number of views and number of upvotes were used. The dataset was labelled regarding the rating derived from the number of upvotes divided by the number of views the answer had gotten. As there were multiple answers to the same question and each had a rating, a dataset could be processed with two answers to a question where one was preferred over the other based on the ranking. For this thesis, the first 4 000 datapoints of the total 10 000 datapoints produced were used due to limitations in hardware when training. A datapoint consisted of a prompt, then two responses; a better-ranked response and a ranked response.

An example of a prompt in the dataset can be seen in Table I, with a related better response in Table II and three of its ranked worse responses in Table III.

Table I
SHOWS AN EXAMPLE OF A PROMPT IN THE DATASET

Prompt
"Do I have too many issues for counselling? I have so many issues to address. I have a history of sexual abuse I'm a breast cancer survivor and I am a lifetime insomniac. I have a long history of depression and I'm beginning to have anxiety. I have low self-esteem but I've been happily married for almost years. I've never had counselling about any of this. Do I have too many issues to address in counselling?"

Table II
SHOWS A RESPONSE TO THE PROMPT IN TABLE I RANKED BETTER COMPARED TO EACH RESPONSE IN TABLE III

Better response
"It is never too late to get help and begin making significant changes and improvements in your life. The right time is when you feel ready and open to change. Sometimes when we have a history of trauma, like sexual abuse, the impact of that trauma can affect many areas of our lives as adults. Working with a therapist who specializes in trauma is a great way to begin developing skills to manage present-day life, learn to regulate your emotions and nervous system when you become triggered, and then in a safe environment have the opportunity to process past traumas so you can live your present life with more ease and joy. People can and do recover from the impacts of trauma at any stage in their life. It is never too late and you can feel better."

Table III
SHOWS THREE RESPONSES TO THE PROMPT IN TABLE I RANKED WORSE THAN THE RESPONSE IN TABLE II

Worse response
"Absolutely not. I strongly recommending working on one issue/need at a time. In therapy you will set smart goals and objectives that will help you reach your goals. I see you as a survivor and not a victim. Best wishes to you."
"Hi, people who have only a few issues and people who have several issues are perfect for therapy. There is no such thing as having too many problems for therapy. Therapy is great because it will help you process all that you are going through, heal from past and present pain, and gain clarity."
"Yeah, you might :)At least, it certainly feels like that, I bet...Not to mention that most counsellors out there might generally feel overwhelmed with a "case" with so many problems. And, so, you will slog along week-to-week with gradual improvements occurring as she slowly does her best to chip away...The real issue is that these folks don't yet comprehend the integrative nature of each of these subjects and how generally simple it is to understand them, and then eradicate them."

B. Performing RLHF

For this study, we performed RLHF with the earlier described data and with the help of a GitHub library named Alpaca-LoRA-RLHF-PyTorch. Some changes were made to the code to make it able to run with the data, and some minor bug fixes were made to make it runnable. A Google Colab notebook was set up to be able to run the training since it requires powerful hardware, which was accessed via a cloud service using Google Colab Pro+. With the correct environment set up, a reward model was trained from the base model Llama-7b-hf, an open-source LLM developed by META AI.

For the reward model to be trained within a reasonable timeframe, PEFT was used during the training, which meant the result from the training were PEFT layers which can be used to make a complete reward model. To complete the reward model, a PEFT script was used to merge the layers that had been trained earlier with a base Llama-7b-hf, resulting in the final complete reward model. The loss function for the reward model was the log loss between the different scalar rewards from the reward model for prompt x and completion y with parameters θ where y_w is preferred out of the pair of y_w and y_l and D is the dataset. The GitHub repository therefore used Pytorch's non-linear activation function logsigmoid that applies element-wise logsigmoid.

$$loss(r_\theta) = E_{(x, y_w, y_l) \sim D} [\log(\sigma(r_\theta(x, y_w) - r_\theta(x, y_l)))]$$

Finally, the reward model was used to train the final model, which was done using a training script which implemented PPO, marking the end of the RLHF training. The RLHF training was performed twice since the first attempt with 5 000 datapoints divided into 625 iterations as the batch size was set to 8, which came to a halt in the 587th iteration. The dataset was lowered from 5 000 to 4 000 datapoints and split, so the evaluation set was set to 700 and the training set to 3300. The dataset of 4 000 datapoints gave 500 iterations in the second execution of the reinforced learning training which amounted to one epoch. For this thesis, there was one epoch of training, and the number of iterations for that epoch was set due to hardware limitations. The time trained was decided by the number of iterations required for one epoch.

C. Evaluation

When the model was fine-tuned, it was later evaluated on its success. The LLMs were solely evaluated by individuals studying psychology and counselling and legitimised experts within the field of psychology. This is to comply with the Swedish Ethical Review Act (2003:460) 4.2§, which restricts research that is performed to affect the users physically or mentally.

The test was a comparison between the pre-trained and fine-tuned models, in which each generated an answer to five different therapy-related prompts using a python script. In the Appendix the prompts can be seen in Table V, the responses from the pre-trained model in Table VI and the responses from the fine-tuned RLHF model in Table VII. Using the responses, a form was created which included the prompts, each with both the answers from the models in a random order (it was not displayed which answer belonged to which model) and eight questions were asked, shown in Table IV. The questions were based on a list of criteria for how a psychologist should perform and were formulated with one of the psychologists who evaluated the models and were based on a counselling session analysis [23]. The questions were qualitative but compared quantitatively as they were rated on a scale of one to five, excluding the last two questions, which were answered "yes" or "no" and then calculated with the values 1 for "yes" and 0 for "no".

D. Reworking the Ethics, Transparency and Accountability Framework

The framework for Ethics, Transparency and Accountability Framework for Automated Decision-Making was applied to answer the second scientific question about ethics in psychology and AI innovation. However, The framework is generalised to all sorts of automated systems and not specified to a psychological chatbot, which is a complex problem requiring more specific guidelines. To navigate the framework for this situation, it was reworked, allowing for a more relevant discussion about the implementation. An analysis of the existing psychologist assistant and the current accountability structure for that profession was also carried out to find the possibilities of AI innovation within the field and apply the framework. Further discussed was the competence required to practice therapy and the regulation of AI to determine how AI innovation within the psychological field is affected.

V. RESULTS

A. Comparing the pre-trained Model and Fine-tuned Model

Our test showed some differences between the pre-trained and the fine-tuned model, but overall the results did not show a clear indication of an improvement between the pre-trained and fine-tuned model.

In total, two psychologists answered the form, which consisted of eight questions. The results can be seen below, where the questions are in the same order from left to right, as stated in IV.C. The first six questions are rated 1-5, and an average is displayed, and the last two are "yes" and "no" questions where the fractions of "yes" answers are shown.

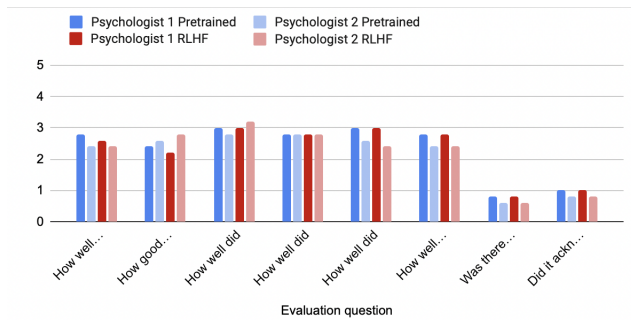


Figure 3. Shows the respective psychologist's evaluation of the pre-trained model and the RLHF fine-tuned model

In Figure 3 the individual evaluations can be observed and can be compared to each other, where psychologist 1 generally rated both the pre-trained and the fine-tuned RLHF model a higher score than psychologist 2. Psychologist 1 also rated the pre-trained model better or the same as the RLHF model. In contrast, psychologist 2 only rated the pre-trained model better once regarding how well the model showed empathy through its response.

For the third question in Table IV concerning how well the response reflected on the prompt, the fine-tuned model

Table IV

SHOWS THE AVERAGE SCORES FOR EACH QUESTION'S PRE-TRAINED AND FINE-TUNED MODEL RESPONSES.

Question	pre-trained	Fine-tuned
How well did this response resemble a psychologist's response?	2.6	2.5
How good was the psychological advice that was given?	2.5	2.5
How well did the response reflect on the prompts written?	2.9	3.1
How well did the response validate the feelings and thoughts?	2.8	2.8
How well did the response show empathy?	2.8	2.7
How well did you perceive the understanding of the situation in the response?	2.6	2.6
Was there normalising of the situation in the response?	0.7	0.7
Did it acknowledge the problems?	0.9	0.9

was ranked slightly higher with a value of 3.1 compared to the pre-trained model with a value of 2.9. However, for the rest of the questions, the pre-trained model had the same score or greater than the fine-tuned model.

As seen in Table IV, no significant differences were found between the pre-trained and fine-tuned models.

B. Psychologist Assistant and Accountability

To find a possible role for AI innovation within psychology, the field and a few current professions were investigated to compare the competence and assess the suitability for its implementation regarding the framework.

In Swedish health care, only PTP- and legitimised psychologists have the right to call themselves psychologists as it is a protected professional title. PTP psychologists have taken their psychology degree but are not legitimised yet. The legitimisation is to ensure that the patients receive safe health care and rely on scientifically based academic competence. It also protects the client's safety and fulfils the related requirements. More importantly, legitimised psychologists are responsible for patients' investigation, diagnosis and treatment in the psychological field [24].

Another professional title is a psychological assistant, which can be taken by someone pursuing a psychology degree and who has passed the first three years. It is imperative to know that this employment is temporary. An important part to distinguish and determine is the formal and actual competence of the assistant to perform duties. The psychologist assistant must have the competence required to complete the tasks and is responsible for how they fulfil their assignments delegated by the licensed psychologist. The assistant does not have formal competence to independently make decisions regarding diagnosing, assessment, and conclusions of investigations, nor to give treatment [24].

By actual competence, the knowledge and experience from education or prior professional experience work is meant. The legitimised expert and the operations manager are responsible for assessing what tasks are considered suitable and whether

the assistant has the competence required for the task. The psychologist must be employed at the same workplace as the assistant to observe the assistant's performance. The assignments need to be clearly defined and thoroughly described, and the assistant needs to deem themselves qualified for them. The psychologist's assistant is also accountable for carrying out the tasks per the rules and regulations within health care and the association's ethical guidelines [24].

Tasks include exposure exercises, social skills training and internet treatment for patients with mild psychological problems. Other tasks can be collecting and compiling data from investigations such as medical history, observations and test results [24].

C. Applying the Ethics, Transparency and Accountability Framework

To answer the second scientific question, the framework described in the theoretical part will be adapted to this particular situation so that the framework can be applied as effectively as possible. The first step in terms of the framework for a psychologist bot is to make sure that it is possible, which can be done using prototyping and ensuring the prototype is functional. This thesis attempts to implement AI in psychology and thus test prototyping possibilities by applying RLHF. The second step is best adapted by developing the bot so that it can be used by all kinds of users, even if the team behind the bot does not represent all users. In this case, the dataset used was scraped from Counsel Chat, and the questions asked were anonymous; therefore, it is not sure that they are representative of all different kinds of users. The relevant users for the psychology bot depend on where it will be implemented. The third step is the same as for the original framework; make sure it is clear who is responsible for the bot and its actions. The same applies to the remaining points. The fourth step is that data and personal information must be processed correctly to ensure it is safe for developers and users; for this specifically, GDPR is relevant. In the fifth step, the user must be informed and understand how the bot should and can be used. This aligns with the limitations for which the legitimised psychologist can use a human psychologist assistant. The last two steps, that the bot should adhere to existing laws and regulations and that the development of the bot should be made with the future in mind, are also the same as for the original framework.

VI. DISCUSSION

A. The RLHF Fine-tuned Model

The results from the investigation revealed no significant difference between the pre-trained and fine-tuned models regarding the set criteria a psychologist should meet in an interaction with a patient. This contradicts the hypothesis that the RLHF fine-tuned model would outperform the pre-trained model.

A reason why RLHF might not have worked as well as we hoped could be that the problem is too complex for the resources available. Fine-tuning an LLM to be a psychologist is a significant task, and fine-tuning it to act as a psychologist

assistant proved difficult. In the previous studies presented in the related works section, the behaviour to align the model with a quality such as helpful and honest was shown to be possible. The role of a psychologist assistant is incredibly complex and nuanced, making learning the behaviour to a machine difficult. The results reinforce the generally accepted idea that complex machine-learning tasks require large amounts of data, time and hardware, and future studies with better access to these resources may uncover more decisive results.

B. Possible Sources of Error

Flaws in the conducted experiment are, for example, the use of only 4 000 datapoints out of the total 10 000 datapoints available. The number of datapoints used was limited by the capacity of the hardware available in Google Colab and had to be lowered to 4 000 for the second trial for it to be able to complete without error, which was successful. A different result could be obtained with more than double the number of RLHF datapoints for the reward model, considering previous works mentioned with RLHF have had datasets consisting of 33 000 to 86 000 datapoints.

In the RLHF process, the first step is pretraining the model; as the LLM used was already pre-trained, it was not done further on prompts and responses related to mental health. It could be interesting to do this in further studies and see if there is a significant difference when pretraining on a psychologically inclined corpus.

In the evaluation, there are also causes for inaccuracies, an example being the use of only five prompts, each with a set of responses from the respective model. For a more thorough study, there should be a more significant number of prompts as these could represent more patients seeking mental health-related questions. The prompts were also longer, presenting a situation to give more information for the model; it would also be beneficial to include shorter questions in the evaluation. That there were only two psychologists who filled out the evaluation form could cause a more significant presence of bias as there can be differences in interpretation of values, mainly since their answers differed; psychologist 1 generally rated the pre-trained model and the fine-tuned model the same except for two questions where it preferred the pre-trained model, whereas psychologist 2 only preferred the pre-trained model when it came to the question regarding how well the response showed empathy. The reason for having solely two evaluators was because they were the ones available and involved in the larger project mentioned in the Project Background of this thesis. This could also increase the bias, as they could have expected one of the models to outperform the other, but their results indicated that the pre-trained and RLHF models performed equally. Also, the fact that the evaluators did not know which response belonged to which model and that they were placed in random order in the evaluation form, the presence of bias due to the psychologists' involvement is less likely. Including more psychologists in the evaluation could have contributed to better representing the consensus and giving different results.

Another possible source of error is that there are dimensions to the psychology field that are not included in the

eight questions in the evaluation form. As the authors of this thesis do not have an academic background in psychology, many factors could have been missed to be incorporated in the evaluation. For example, this evaluation only has prompts and responses, whilst there are conversations when a patient sees a psychologist. However, this evaluation indicates how well the models perform in some critical aspects of psychology.

C. Competence of the Digital Psychologist Assistant

The results indicate that the fine-tuned model was unsuitable for implementation in the psychology field and real-life situations. The following discussion was thus built on the assumption that the model improved significantly to an acceptable level.

As written earlier, the problem with an AI psychologist is the difficulties in controlling the quality and determining competence. A psychological assistant helps psychologists with their workload and is the closest function this model could fulfil in the field. They have a formal education and prior experiences evaluated through rigorous exams and supervised practical work in pursuit of a degree from an institute. As previously mentioned, if a psychological assistant is to perform a task, the assistant must assess themselves that they are qualified for the task or otherwise step back. Comparing this to the language model that does not possess the self-awareness to determine the ability to perform a task unless explicitly defined in the instruction and development of the model. This creates uncertainty in how the model can be used within psychology.

The questions in the evaluation form were an attempt to measure the competence of the LLMs. Still, they are not equivalent to the amount of training and education a human psychologist assistant acquires. It is imperative to further work on assembling more legitimate criteria and perhaps evaluate the model similar to how a psychology student is evaluated. This is to determine if it is possible to map the competence of a LLM and equate it to the competence of a human psychologist assistant for it the model to be a tool in the field of psychology as a digital psychology assistant.

D. Discussing the Ethics, Transparency and Accountability Framework

Using the Ethics, Transparency and Accountability for Automated Decision-Making framework on how AI is used safely, sustainably and ethically can help clarify the limits of a digital psychologist assistant and the advantages and disadvantages depending on where it is implemented and who is held accountable for it.

1) *Testing to avoid unintended outcomes:* With the first step of the policy framework, it is important that further empirical testing and research are conducted before establishing and introducing a prototype to the field. From the results, it can be gathered that it also is crucial to define clear limits in the model and its tasks, just like the psychologist assistant needs defined assignments. If the task is to help patients with mild psychological issues, the limit could be incorporated into the instruction of the language model to the responses that specifically pertain to the psychological field.

2) *Fair services for all users:* For the second step, it is important to determine in what scope this tool can be used. Depending on what it is capable of, it has to be assessed whether it should be introduced to let patients use it on their own or under the supervision of psychologists or whether it even should be used by psychologists. How can it be implemented ethically? Considering the model would be a digital psychologist assistant, the same rules that apply to the human psychologist assistant must also be used if not more regulations. Therefore only the tasks deemed suitable for a psychologist assistant should be considered appropriate for the digital alternative. This includes therapy for individuals regarded as having moderate and mild psychological issues. There needs to be rigorous testing before it is ethical to let patients use it independently, as the technology is new. Therefore, it is not guaranteed that the bot's response will not be toxic. One strategy could be first to implement it for the experts, then observe it by the experts and lastly, introduce it directly to patients. At first, the psychologists can test out the digital psychologist assistant and see how well it performs and aids in their practice. The bot could be used as a co-therapist for treating the patients just as a human psychologist assistant, where the psychologist could get input from the model in their analysis. Here it is not only the responsibility of the developer of the product but also the psychologist as to how they incorporate the bot into their practice. It is also the organisation's and the developer's responsibility to share clear guidelines on how the bot should be used and to be clear with the limitations. If the psychologist later observes the bot in its interaction with the patient, the danger of the model releasing harmful content to the patient is still avoided. The psychologist is more accountable as they supervise the bot's response. Lastly, the implementation of individuals getting direct access to it unsupervised also moves a large part of the accountability to the distributor, that is to say, the therapist organisation. It can be dangerous to release a model that can have toxic output to the public. Here it is the organisation's responsibility to give access to the digital psychologist assistant to be clear on what purposes the model can be used for. However, if the model is left unsupervised, it cannot fulfil the duty of a digital psychologist assistant according to the criteria for a psychologist assistant.

If used long enough in each phase, this strategy would let the implementation's effects gradually reveal themselves; if there are problems, they would hopefully be noticed early. No matter which implementation is chosen, all of them give the model access to sensitive information about individuals. This and compliance with the GDPR will be discussed further in the fourth step. The bot needs to be adapted to large amounts of data to be used by all potential users.

3) *Be clear about who is responsible:* The third step is paramount as this implementation can consequently have various negative repercussions. Initially, the developers of the tool have the responsibility and are held accountable for their released product. This includes the use and storing of data, the output that the model has, whether it is toxic or not and communicating exactly what the bot can be used for and its limitations. Then the accountability is also placed on

the Swedish Association of Psychologists to approve it and organisations that distribute it, as they have introduced the bot in their practice. From there, the psychologist is also responsible for how they use the tool if they are either using it themselves or observing the patients using the bot.

4) *Handle data safely & protect users' interests*: The fourth step concerns patient safety and where GDPR is vital as the model would get access to sensitive information of individuals and be exposed to leakage. This is further the responsibility of the organisation using the tool and gathering data from their patients. This affects AI innovation within psychology and several other fields extensively since much more care must go into data handling and security. If the data is considered too sensitive, it could limit what the chatbot could be used for and thus creates hurdles for AI innovation within psychology. On the other hand, it is a trade-off that must be done to protect privacy, even if innovation proceeds slowly.

5) *Help understand how it affects users*: The fifth step is also crucial, as the user has to be aware of what the bot is capable of and in what areas it can be used. The psychologist and the individual patients need to be notified to prevent misuse and false expectations of the model.

6) *Follow laws & regulations*:

7) *Build future proof & sustainable*: The sixth and seventh step highlights the importance of adhering to laws and creating a future-proof product. This means that the developers and organisations that employ the bot must work with the regulations at hand to ensure an ethical implementation. It is worth noting that this is a new technology, and AI is relatively unregulated regarding how it can be used. Still, it could be a potential market if ethically implemented. As previously mentioned, there are regulations regarding processing and storing personal data, for example, GDPR, which may stifle innovation. Another regulation that may slow down innovation is the Swedish Ethical Review Act (2003:460) 4.2§, which restricts research that is performed to affect users physically or mentally. This act ensures the quality of research is held high but, at the same time, makes it harder to conduct it, as resources can be more challenging to acquire. Same as for GDPR, it is a necessity that slows innovation for the price of something equally important. Nevertheless, there is a general lack of regulation regarding accountability which is crucial for the research and development of AI. If there are not enough constraints for the developers, there is a risk that the quality of the product does not reach a sufficient level, as there is no incentive for them to care about quality. This could lead to people disregarding the technology and losing trust in the product. Whereas placing too strict regulations can result in the loss of developers, as there can be a fear of liability in the case of negative repercussions, which also hinders innovation. If too strict regulations are placed on the distributor, the product might not be as widespread, inhibiting the integration and acceptance of AI innovation in the field. It is essential to be mindful when creating regulations since it is an act to balance innovation and safety. An additional aspect to consider is the different rules and regulations in other countries, as national laws associated with psychology

do not necessarily correspond internationally. Developing a digital psychologist assistant could therefore require different versions that are tailored to adhere to federal regulations.

VII. CONCLUSION

From the results gathered in this study, no significant differences were found between the pre-trained model and the fine-tuned model to conclude that fine-tuning using RLHF results in better performance in the field of psychology. Some rules and regulations can be applied to this field of research. However, there are not enough specific ones for implementing a digital psychologist assistant or AI in general. Therefore, governments must keep introducing relevant regulations for AI innovation, including but not limited to ethics, accountability, and personal data, especially since the chatbot can encourage changes in patients' behaviours and affect individuals' lives. Suppose a thoroughly tested product was to be introduced to the market. In that case, the Ethics, Transparency and Accountability for an Automated Decision-Making framework should be complied with to ensure an ethical implementation.

ACKNOWLEDGEMENT

We want to express our sincerest gratitude to our supervisors, Birger Möell and Jonas Beskow, for the support provided through motivation, coaching and direction. We would also like to thank our peers, Samuel Höglund and Josef Khedri, for the constructive feedback and help provided throughout the completion of this bachelor thesis.

REFERENCES

- [1] N. Lambert, L. Casticato, L. von Werra, and A. Havrilla, *Illustrating reinforcement learning from human feedback (rlhf)*, Last accessed 3 March 2023, 2022. [Online]. Available: <https://huggingface.co/blog/rlhf>.
- [2] D. Jurafsky and J. H. Martin, "Speech and language processing", in 2023, ch. 10, 11.
- [3] SACO, *Psykolog*, Last accessed 18 April 2023, 2022. [Online]. Available: <https://www.saco.se/studieval/yrken-a-o/psykolog/>.
- [4] L. Rozbruch, *Litigation robotic surgery: Product liability or medical malpractice?*, 2018. [Online]. Available: <https://www.pulj.org/the-roundtable/litigation-robotic-surgery-product-liability-or-medical-malpractice>.
- [5] G. Talton, *Who is at fault in a self-driving car accident?*, Last accessed 3 May 2023, 2023. [Online]. Available: <https://www.farrin.com/blog/who-is-at-fault-in-a-self-driving-car-accident/>.
- [6] L. Walker, *Belgian man dies by suicide following exchanges with chatbot*, Last accessed 14 June, 2023, 2023. [Online]. Available: <https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt>.
- [7] C. Dilmegani, *Large language model training in 2023*, Last accessed 6 March 2023, 2023. [Online]. Available: <https://research.aimultiple.com/large-language-model-training/>.

- [8] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, “Attention is all you need”, *Advances in neural information processing systems*, vol. 30, 2017.
- [9] T. Decoder, *Gpt-4 has a trillion parameters – report*, Last accessed 2 June, 2023, 2023. [Online]. Available: <https://the-decoder.com/gpt-4-has-a-trillion-parameters/>.
- [10] S. Mangrulkar and S. Paul, *Peft: Parameter-efficient fine-tuning of billion-scale models on low-resource hardware*, Last accessed 11 May 2023, 2023. [Online]. Available: <https://huggingface.co/blog/peft>.
- [11] E. Beeching, Y. Belkada, L. von Werra, S. Mangrulkar, L. Tunstall, and K. Rasul, *Fine-tuning 20b llms with rlhf on a 24gb consumer gpu*, Last accessed 2 June, 2023, 2023. [Online]. Available: <https://huggingface.co/blog/trl-peft>.
- [12] T. Simonini, *Proximal policy optimization (ppo)*, Last accessed 16 June, 2023. [Online]. Available: <https://huggingface.co/blog/deep-rl-ppo>.
- [13] C. Phelps, *Transformer reinforcement learning*, Last accessed 6 March 2023, 2020. [Online]. Available: <https://wandb.ai/lvwerra/trl-showcase/reports/Transformer-Reinforcement-Learning--VmlldzoXMDYyNjQ>.
- [14] I. Consling, *Personal data*, Last accessed 5 May 2023, 2021. [Online]. Available: <https://gdpr-info.eu/issues/personal-data/>.
- [15] K. Granville, *Facebook and cambridge analytica: What you need to know as fallout widens*, Last accessed 5 May 2023, 2018. [Online]. Available: <https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html>.
- [16] U. government, *Ethics, transparency and accountability framework for automated decision-making*, Last accessed 5 May 2023, 2022. [Online]. Available: <https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making>.
- [17] Y. Bai, A. Jones, K. Ndousse, *et al.*, “Training a helpful and harmless assistant with reinforcement learning from human feedback”, *arXiv preprint arXiv:2204.05862*, 2022.
- [18] L. Ouyang, J. Wu, X. Jiang, *et al.*, “Training language models to follow instructions with human feedback”, *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.
- [19] B. Guo, X. Zhang, Z. Wang, *et al.*, “How close is chatgpt to human experts? comparison corpus, evaluation, and detection”, *arXiv preprint arXiv:2301.07597*, 2023.
- [20] B. Inkster, S. Sarda, V. Subramanian, *et al.*, “An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study”, *JMIR mHealth and uHealth*, vol. 6, no. 11, e12106, 2018.
- [21] K. Kretzschmar, H. Tyroll, G. Pavarini, A. Manzini, I. Singh, and N. Y. P. A. Group, “Can your phone be your therapist? young people’s ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support”, *Biomedical informatics insights*, vol. 11, 2019.
- [22] N. Bertagnolli, *Counsel chat: Bootstrapping high-quality therapy data*, Last accessed 14 June, 2023, 2020. [Online]. Available: <https://towardsdatascience.com/counsel-chat-bootstrapping-high-quality-therapy-data-971b419f33da>.
- [23] L. Vroom, *Counselling session and analysis*, Last accessed 14 June, 2023. [Online]. Available: <https://eportfolio.ucalgary.ca/vroom/wp-content/uploads/sites/140/2017/07/Counselling-Session-and-Analysis.pdf>.
- [24] S. Psykologförbund, *Anställning som psykologassistent*, Last accessed 31 May, 2023, 2023. [Online]. Available: https://www.psykologforbundet.se/globalassets/psyk-ass_2016_lr.pdf.

APPENDIX

Table V
SHOWS THE FIVE PROMPTS USED IN THE EVALUATION FORM

Number	Prompt
1	"I have been struggling a lot with anxiety lately. I often get stuck with the same thoughts repeatedly, and it feels like a downward spiral. I have a hard time finishing schoolwork because I procrastinate a lot and I am starting to have problems finding time and energy to hang out with my friends. How can I break my bad habits and deal with my anxiety?"
2	"Being a full-time single mom trying to balance everything can get pretty hectic sometimes. Whenever things pile up too much or become chaotic, I start feeling really guilty because it makes it harder to take good care of both home duties and kids. I don't really know what to do and feel like I am disappointing everyone. How can I please everyone and find a balance in my life?"
3	"Feeling isolated socially has become routine. I find it challenging connecting with people since starting college, particularly amidst all the chaos and uncertainties. I feel insecure and struggle to find courage to make new friends. I've realised that this has started to affect my family life. I often eat lunch alone at school, making me doubt my self-worth. Why is it so hard and what can I do to improve my social life?"
4	"Having grown up in an abusive household, I frequently experience flashbacks, nightmares, and intrusive thoughts that cause intense distress. Despite understanding what happened was beyond my control, I am struggling to come to terms with my childhood experiences. Now I have developed trust issues, even to my closest friends - and I struggle in my social life. How can I cope with my past trauma and heal from it?"
5	"I've been dealing with body image issues for quite some time now and it's been eating me up inside. I always feel so disappointed when I look in the mirror and often compare myself to others. I've noticed that I tie my self worth to my looks as I felt amazing when someone complimented my appearance. I try to be body positive but it is just not working. How do I deal with this? It affects my life in more ways than I let on."

Table VI
SHOWS THE FIVE RESPONSES TO THE RESPECTIVE PROMPTS IN TABLE V
FROM THE PRE-TRAINED MODEL, USED IN THE EVALUATION FORM

Prompt	Response
1	"1) You seem very anxious about your academic performance. It is important for students to feel good about their academic achievements. However, if you are feeling anxious about your academic performance, then this may be causing you to avoid doing work or studying. This could lead to further feelings of anxiety and stress. If you find yourself avoiding things due to fears, then you should try to face these fears head on. For example, if you are afraid of failing an exam, then you should study more so that you don't fail the exam. "
2	"1) You have a lot on your plate right now! It is not uncommon for parents who work outside the home to experience guilt about their responsibilities at home. Parents often worry that they aren't doing enough for their children, even when they are working hard to make sure that all needs are met. This can lead to feelings of guilt, which may then cause more stress. Stress can also affect our ability to cope with other situations, such as parenting challenges. In addition, we tend to focus on the negative aspects of life rather than the positive ones. "
3	"1) You have identified an important issue for yourself. It is not unusual for students to experience some degree of social isolation during their first year at college. This may occur because they are away from home for the first time or because they are experiencing academic stressors such as taking more courses than before or having difficulty adjusting to a different learning style (e.g., lecture versus small group discussion). In addition, many students report feeling overwhelmed by the large number of choices available on campus and the pressure to decide which activities to participate in. These factors may contribute to feelings of loneliness and isolation."
4	"1) You may want to consider seeking professional help for your symptoms of post-traumatic stress disorder (PTSD). There is no shame in getting help when needed. 2) It sounds like you have been through a lot in your lifetime. While there is nothing wrong with having trust issues, they could also prevent you from forming healthy relationships. If you feel comfortable doing so, you might try talking about this issue with someone close to you. 3) In addition to seeking professional help, you should also work on building self-esteem by focusing on positive aspects of yourself."
5	"1) You have identified your issue which is great! 2) You have also recognized how your feelings about yourself impact other areas of your life such as relationships or work performance. This is important because it helps us understand what might be causing these negative thoughts and behaviors. 3) You have acknowledged that you want to change your behavior around your body image. This is an important step towards making changes. 4) You have asked for help from our community. This shows that you recognize that you need support to make these changes. 5) You have provided information regarding your situation. "

Table VII

SHOWS THE FIVE RESPONSES TO THE RESPECTIVE PROMPTS IN TABLE V
FROM THE FINE-TUNED RLHF MODEL, USED IN THE EVALUATION FORM

Prompt	Response
1	"1) You seem to be very aware of your problematic behaviors and how they affect you. This is an important first step! 2) It sounds like you might benefit from some CBT (cognitive behavioral therapy). There are many resources available online for this type of treatment. Here s one example: 3) If you feel comfortable doing so, please consider reaching out to me directly via email or phone. I m happy to help if needed. 4) Please let me know how things go! "
2	"1) You have a lot on your plate right now! It is not uncommon for parents who work outside the home to experience guilt about their time away from their children. This may be especially true if you are working long hours or travel frequently. In addition, being a single parent means that you must also manage all aspects of household management (e.g., cooking, cleaning, laundry), childcare, and other responsibilities such as paying bills, managing finances, etc. While this may seem overwhelming at times, there are many resources available to help you with these tasks."
3	"1) You have identified an important issue for yourself. It is not unusual for students to experience some degree of social isolation during their first year at university. This may occur because they are away from home for the first time or because they are experiencing academic stress. In addition, many students report feeling overwhelmed by the amount of work required in order to succeed academically. These factors combined with the increased demands on your time (e.g., classes, studying, etc.) may contribute to feelings of loneliness and isolation. However, these issues should not go unaddressed. "
4	"1) You need to understand your own history of being abused by your parents. This is important because if you do not know how this has affected you then you cannot begin to work on yourself. 2) You also need to learn about the effects of post-traumatic stress disorder (PTSD). There are many resources available for learning more about PTSD including books, websites, and support groups. 3) It may help you to talk to someone who knows something about PTSD. A good place to start might be talking to your doctor or counselor. "
5	"1) You have identified an issue which many people struggle with. It can be very difficult to overcome these feelings of low self-worth. However, there are things you can do to help yourself. 2) Firstly, remember that your value does not come from how you look or what other people think about you. Your value comes from who you are on the inside – your personality, your character, your values, your interests etc. 3) Secondly, focus on doing activities which make you happy rather than focusing on your appearance. For example, if you enjoy playing sports then play more sport! "

