An Empathetic AI Painter: A System for Computational Creativity Through Embodied Conversational Interaction

Abstract

This paper presents an investigation into the computational modeling of the creative process of a portrait artist, focusing on the incorporation of human traits like personality and emotions into the artistic process. The system includes an empathetic conversational component to discern the dominant personality traits of the user, and this information is then utilized by a generative AI portraiture module to create a personalized stylization of the user's portrait. The paper details the system and the outcomes of real-time interactions from a demonstration session.

1 Introduction

The incorporation of human traits in the creation of artworks has consistently held significant importance. Although there are differences between art and science regarding their goals and toolsets, these distinctions blur when artists use scientific understanding to inform their work and science examines art to comprehend the human experience. The idea of leveraging established psychological insights into human traits such as personality and emotion to guide the creation, critique, and informing of artwork is not novel. Traditional portrait artists employ their understanding of human perception and vision to create portraits from life or photographs. This process includes the arrangement of the environment, placement of the subject, and an interview to grasp their mental and physical characteristics. Artists also aim to convey their individual painting style while trying to express personal and universal ideas. An artist has several options in themes, brush style, color plan, edge and line plan, abstraction style, and emotional narrative to achieve the finished artwork. Computational creativity and generative art offer fresh avenues for modeling scientific knowledge to replicate this process and deepen our grasp of human creativity. This study uses AI techniques to begin emulating this artistic procedure. The Empathic AI Painter system seeks to discover novel approaches to balance diverse aesthetic and conceptual aspects.

2 System Description

The Empathic Painter System is created to mimic the interaction between a live portrait artist and a person, referred to as the sitter. It aims to understand the sitter's traits, such as personality and emotions, to create a unique portrait by selecting the appropriate abstraction techniques, color palette, and style that correspond to those traits. The system operates in a two-stage process; the first stage involves capturing the characteristics of the sitter, followed by the second stage, which uses the captured traits to generate a stylized artistic representation of their portrait. The initial stage of capturing the personality of the sitter occurs during the conversation with an embodied conversational agent, using empathetic interaction methods. This system utilizes the M-Path conversational agent, which has been developed previously. The M-Path system was modified for this demonstration to conduct an interview based on the Big-5 personality questionnaire to categorize the sitter into one of the established personality dimensions. This data is then used to map the personality traits to a particular artistic style. The mapping is transferred to the Generative AI Portrait Stylization system in

.

the second stage, which creates an artistic portrait. The interaction process includes several steps. First, a portrait of the sitter is captured under controlled lighting conditions, and a unique ID is assigned after consent is provided for participation and use of the portrait. The sitter is then given information about the M-Path system with instructions about how to interact. The sitter initiates the interaction until a complete conversation is concluded and the agent informs the sitter that the interaction has ended. The M-Path system uses the data collected to classify the sitter's personality into a specific dimension. This dimension is then used by the Generative AI Portraiture system to create a personalized portrait style. The generated portraits are showcased on a monitor for all participants and the crowd to observe and assess.

2.1 Big-5 Personality Mapping

The five-factor model of personality is also known as the "Big-5 Personality Model" and is designed as a categorization to capture the variations in personality traits among individuals. This model classifies personality variations across five dimensions: extraversion, openness, conscientiousness, neuroticism, and agreeableness. Each of these dimensions encompasses a wide range of psychological functions, which are composed of more specific traits. Extraversion pertains to the extent to which people are dominant, talkative, assertive, active, energetic and enthusiastic. Openness characterizes people who are curious, creative, innovative, imaginative, reflective, cultured, curious, original, broad-minded, intelligent, and artistically sensitive, seeking new experiences and exploring novel ideas. Conscientiousness indicates an individual's level of hard work, persistence, organization, and motivation in achieving their goals. Individuals high in conscientiousness tend to be organized, plan-oriented, and determined. Neuroticism, also referred to as Emotional Stability, represents differences in emotional stability and adjustment. Individuals scoring high on neuroticism tend to experience negative emotions, such as anxiety, depression, impulsiveness, self-consciousness, vulnerability, anger, hostility and worry. Agreeableness is linked to likability, conformity, friendliness, and social compliance. Individuals with high scores in agreeableness are characterized as trusting, caring, forgiving, altruistic, flexible, gullible, good-natured, soft-hearted, cooperative and tolerant. This model is based on factor analysis of descriptive words of human behavior. The questionnaire used is a shortened version of the Revised NEO Personality Inventory, which has 120 questions and takes 45 minutes to complete. For the online demonstration, one statement for each dimension was used, where the whole conversational interaction could be completed in under 5 minutes. Each question is further modified to align with the conversation setup in the demonstration environment.

Dimension	Question
Openness	How do you like the conference so far, is it interesting to you?
Conscientiousness	Don't you think the conferences are always a bit chaotic?
Extraversion	Do you normally talk and interact with a lot of people?
Agreeableness	How about agents? Do you trust me in sharing how you feel?
Neuroticism	How do you feel about your portrait being displayed on the screen?

Table 1: The questions used for the personality dimensions.

The answers to these questions are evaluated for their polarity and then mapped onto two-factor dimensions for personality adjectives. The mapping model is the Abridged Big Five Circumplex Model, in which facets of the Big Five dimensions are mapped as combinations of two factors. The AB5C mapping contains descriptive personality terms for each of the resulting 90 combinations, where the most distinctive trait of an individual is used to select the column, and the second most distinctive trait selects the row. These traits may be either negative or positive. The mapping from Big-5 traits to the Generative AI portrait styles was provided by art experts who independently mapped the styles to the Big-5 categories and reached an agreement.

2.2 Empathic Conversational Avatar

The starting point of interaction is the empathetic conversational agent, M-Path, which was developed using a framework based on a computational model of empathy. M-Path is a human-like avatar capable of initiating and maintaining an emotional conversation, based on the predetermined goal of the dialogue. The interaction involves a face-to-face conversation with a human interaction partner,

similar to a video-conference with audio and visual input and output. The agent processes the real-time inputs in terms of their linguistic and affective properties to generate empathetic verbal and non-verbal behavior. The main objective of the interaction is to complete the modified Big-5 questionnaire to categorize the partner's personality and send it to the generative art system. The system has three distinct modules: a perceptual module, a behavior controller and a behavior manager. The perceptual module gathers the video and audio signals when the conversation partner is speaking. This process was triggered with a push-to-talk system. M-Path enters a listening state when the user speaks. During the listening state, speech and facial expressions are processed in real-time for speech and emotion recognition. The video input is used in the facial emotion recognition module, which uses an OpenCV face-recognition algorithm to identify the face. Emotions are categorized using a CNN model, trained on the CK+ Dataset, into 6 basic emotion categories. The speech input is sent to the speech-to-text module which uses a service to get streaming speech recognition. Sentiment analysis evaluates the text for its polarity using the SO-CAL Sentiment Analyzer, which was trained on the NRC-Canada lexicon. The text is sent to the decision-making module for creating conversational responses. This process continues until the partner finishes speaking, which concludes the listening state. The information is then sent to the decision-making module, and the agent enters a thinking state. The behavior controller module creates goal-directed verbal and non-verbal responses in all states of the conversation: listening, thinking, and speaking. This is done by analyzing the user's emotional response from the listening state. The conversation begins with the user's greeting and finishes when the agent receives suitable answers to the personality survey questions. The listening, thinking, and speaking states of the agent loop until the user is categorized. During the listening stage, the agent shows a non-verbal affect matching response and backchanneling behavior. Affect matching is a facial expression that mirrors the user's facial expressions in real-time, chosen by empathy mechanisms. Backchanneling is created by a nodding behavior when pauses are detected in the user's speech. These behaviors are combined to create an empathic listening behavior. After the conversation with the participant ends, the final text received and the user's overall sentiment are sent to the Dialogue Manager (DM), and ultimately to the Empathy Mechanisms (EM). The DM completes the Big-5 personality questionnaire to assign a personality category. The EM ensures that the DM generates empathetic responses while reaching its goal. The DM gathers the appropriate emotional response from the EM to generate an emotionally appropriate verbal reaction to the user, followed by a survey-related coping response, and then the next survey question. The system uses the scikit-learn library in Python for the TF-IDF vectorizer model, and the NLTK Lemmatizer. A second model is created by fine-tuning BERT for the classification of user responses according to sentiment and the Big-5 questionnaire answers. The Big-5 questionnaire answers are collected to select the most dominant personality dimensions of the user, based on their probability values and polarity. The Big-5 mapping is used to select a category for the user, with adjectives. This categorization is then sent to the generative art cycle to produce a personalized portrait. After each response is generated by the dialogue manager, it is sent to the behavior manager to be performed by the conversational agent during the speaking state. To achieve a natural conversation, the system continuously produces non-verbal and verbal behaviors. Lip movements, facial expressions, head gestures, body gestures, and posture are synchronized with the agent's speech. The animation is sent as a BML message to the Smartbody character animation platform, to display the generated behaviors.

2.3 Generative AI Portraiture System

The stylistic rendering of the portraits is generated by the generative art component of the system. The portrait goes through three processing phases. The first phase preprocesses the original portrait by using an AI tool to separate the foreground from the background, which will be used to stylize the portrait. Then, the light and color balance of the face are adjusted to achieve a lighting effect, where one side of the face is dramatically shown. The next phase uses this image and the personality category as inputs to a modified Deep Dream (mDD) system with multiple passes on the image to create the base style. While most DD systems use pre-trained networks with object recognition data, the modified system uses artistic paintings and drawings as training data. The system has a dataset of 160,000 labeled and categorized paintings from 3000 artists. A method called hierarchical tight style and tile was developed to overcome the problem that most artists create fewer than 200 paintings in their lifetimes. In the last phase, the source image from the previous phase is further enhanced using the personality category. The ePainterly system combines Deep Style techniques as a surface texture manipulator, and a series of Non-Photorealistic Rendering (NPR) techniques like particle systems, color palette manipulation, and stroke engine techniques. This iterative process enhances

the portrait, and the final result is shown in an online gallery. The ePainterly module is an expansion of the Painterly painting system, which models the cognitive processes of artists based on years of research. The NPR subclass of stroke-based rendering is used as the final part of the process to realize the internal mDD models with stroke-based output. This additional step reduces noise artifacts from the mDD output, creates cohesive stroke-based clustering, and a better distributed color space.

3 Conclusion

The Empathic AI Painter was presented at a conference demonstration session. Forty-two participants tested the system, with 26 of them completing the portrait-taking and interaction. Each conversation with the M-Path system took approximately 5 minutes. The performance of the M-Path system was evaluated individually. On average, 84.72