AvatAIr - A Pipeline for Generating Optimized, Individually Tailored Avatars

MARCEL GREINER, Ulm University, Germany
LUKAS STEINWENDER, Ulm University, Germany

There are various methods available for defining the visual characteristics of virtual characters, such as avatars. One common approach is to modify predefined variables, like hair style and eye color. However, a challenge arises when users are unsure about how they want their avatar to look. To address this, we present AvatAIr, a pipeline that generates user-preferred, photo-realistic avatars using Stable Diffusion and Bayesian optimization. Unlike traditional avatar generators, where users manually adjust each attribute, our pipeline allows the program to determine the most suitable avatar based on user ratings. Our goal is to simplify the avatar creation process by focusing more on user preferences and less on parameter adjustments. Our pipeline utilizes stable diffusion to synthesize high-resolution, realistic avatars. Furthermore, Bayesian optimization plays a crucial role in fine-tuning the generated avatars according to specific user preferences, incorporating feedback loops to iteratively impove towards a user-preferred avatar. The project is open-sourced and available at https://github.com/lukassteinwender/avatair

Additional Key Words and Phrases: avatar, avatar generator, user-preferred individual avatars, bayesian optimization, stable diffusion, avatar generation, user preferences, generative models

ACM Reference Format:
Marcel Greiner and Lukas Steinwender. 2023. AvatAIr - A Pipeline for Generating Optimized, Individually Tailored Avatars. 1, 1 (March 2023), 14 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

1.1 Background

1.1.1 Avatar Use in Various Application Areas. Avatars are used in advertising, entertainment, human-computer interaction, and as research tools to investigate human emotion perception [16]. Furthermore, avatars can be virtual representations of individuals who are used for technical, scientific or educational purposes [31]. They serve as virtual representations of users or systems in various applications and contribute to enhancing customer satisfaction and facilitating seamless communication [31]. Application areas can range from increasing trust in highly automated cars [36], increasing attractiveness of public displays with personalized avatars [24] to serving as a tool for self-exploration and socialization in a virtual world [14].

A notable attempt of increasing trust and communication efficiency includes the animated AI assistant Nomi from the automotive brand Nio (see Figure 1). In terms of the avatar, Nomi distinguishes itself by providing a distinctive visual portrayal as an abstract face accompanied by its own animations, setting it apart from conventional car assistants.

Authors' addresses: Marcel Greiner, Ulm University, Germany, marcel.greiner@uni-ulm.de; Lukas Steinwender, Ulm University, Germany, lukas.steinwender@uni-ulm.de.

Manuscript submitted to ACM

1.1.2 Importance of Customized Avatars. Developing customizable avatars is crucial for improving the user experience and enhancing self-identification, engagement and motivation in digital environments like role-playing games [17]. They allow users to tailor their virtual self-representation to match aspects of their actual or ideal self, influencing psychological well-being [23]. In a different setting, avatars in business lead to increased customer satisfaction, a more positive attitude toward the product, and a greater purchase intention in online shopping [10]. Therefore, customizable avatars represent an advancement toward a personalized and immersive digital future.

1.2 Motivation

1.2.1 Challenges. Avatar generation is currently facing a number of challenges. One major issue is the uncanny valley effect[22], which refers to the discomfort or lack of appeal that users may experience when interacting with realistic avatars [13]. Furthermore, the choice of avatar creator application can also impose limitations on customization options, such as a restricted range of hairstyles or limited possibilities for adjusting facial features. For instance, Ready Player Me [15], a virtual avatar platform by Wolf3D, enabling users to create customizable 3D avatars for use in diverse virtual reality applications and environments, offers 10 distinct face forms, Snapchat's Avatar Creator Bitmoji provides 18 options, and Facebook's Avatar Creator presents 6 face form choices along with 3 face width variations, totaling 18 options. In comparison, Apple's Memoji Creator offers 15 different options for face form.

Also, users find it difficult to create the ideal avatar because there are so many settings and parameters that need to be adjusted for the best outcome [20]. Added to this, people never really know what they want until you show it to them [3]. These challenges highlight how difficult it is to design avatars that smoothly fit each user's expectations while balancing realism and acceptability.

1.2.2 Objective. Creating an effective pipeline for generating customized avatars is the main objective of this project. The work attempts to overcome current obstacles in avatar customization.

1.3 Contribution

That is the reason why we have created a novel pipeline for avatar generation. This approach acknowledges the notion that individuals may not have a clear idea of their preferences until they are presented with visual representations. Instead of manually adjusting each detail of an avatar, users are now offered automatically generated avatars. Users are then prompted to evaluate and rate these avatars, offering valuable feedback on elements they find appealing or disconcerting. This user-provided data is used to initiate an iterative optimization process that continuously improves the avatar based on the collected feedback and preferences.

In this paper, we introduce a pipeline called "AvatAIr - A Pipeline for Generating Optimized, Individually Tailored Avatars." Our pipeline combines Bayesian optimization and Stable Diffusion. This enables us to produce photo-realistic avatars that are customized to the preferences of individual users.

2 RELATED WORK

In this section, we discuss previous studies related to the development of a pipeline for creating optimized and personalized avatars. We will explore research contributions in the field of avatars and technologies that employ similar methods for avatar creation, customization, and optimization.

Fig. 1. The artificial intelligence assistant, Nomi, developed by the automotive brand NIO. [11, 12]

2.1 Current Research in the Field of Avatars

2.1.1 Avatars in Human-Agent Interaction. Human-Agent Interaction (HAI) is an emerging area of research that investigates and advances the design methods for interaction across a range of interactive platforms, including digital media, software agents, robots, and human-human communication [7]. Avatars play a crucial role in the field of human-agent interaction as they serve as digital representations that facilitate the connection between individuals and virtual entities. Numerous studies have investigated different aspects of human-agent interaction with avatars, including the significance of avatar expressions in virtual environments[18]. Liang et al.'s paper focuses on the interaction between a virtual actor (avatar) and the virtual world, specifically in terms of the avatar's expressions and its response to events in the environment. The outcomes of their carried out experiment show that the suggested method is workable and that it is possible to successfully connect the avatar's emotions to the virtual environment, improving the interaction experience as a whole [18]. Furthermore, researchers have investigated the integration of semi-personalized virtual agents into human-computer interaction [8]. They have also explored the utilization of voice and gesture recognition systems to facilitate real-time interaction [33]. These efforts have led to enhancements in the overall user experience [8] [33]. Furthermore, J. Carey et al. [5] found out that, expressive communication in virtual worlds - which includes body language, tone of voice, gestures, and nonverbal cues - can help people get emotional satisfaction and feel comfortable in social interactions [5]. These studies have demonstrated the potential of avatars in facilitating communication and interaction between humans and virtual agents, with applications in fields such as virtual reality and artificial intelligence.

2.1.2 Avatars in the Automotive Context. An impactful application could be within the automotive context, where avatars have redefined the user experience by serving as virtual representations. Several papers have explored this [4, 9, 36]. To enhance safety and user experience in automated driving systems, Bengtsson proposes using an in-vehicle avatar interface to enhance human-vehicle interaction [4]. It suggests that concepts of social attunement, such as sharing intentions, inferring goals, mutual predictability, and understanding performance limitations, can be applied

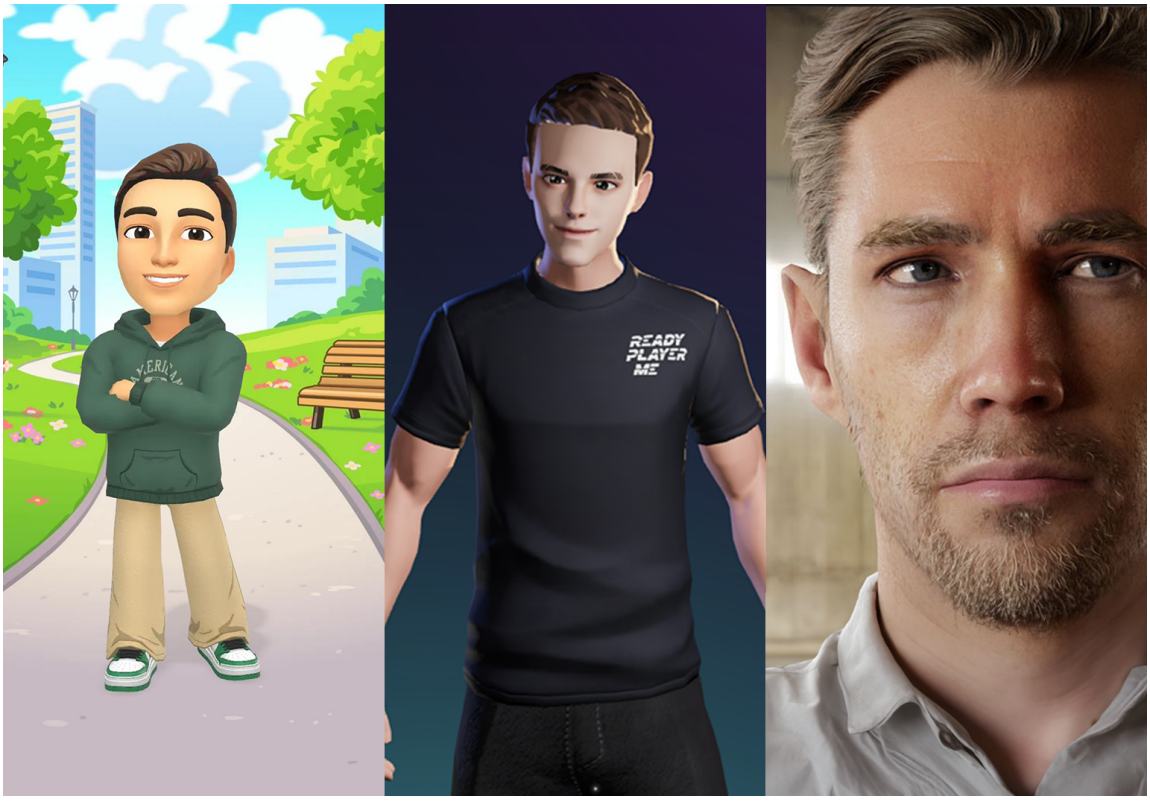

Fig. 2. From left to right: avatar creator applications - Snapchat Bitmoji, Ready Player Me, and Character Creator by Reallusion. [21]

with avatars to the human-vehicle interaction [4]. In addition, avatars can support media augmentation through automotive motion by providing the user with a control input to interact with the media [19]. The study conducted by Rakotonirainy et al. explored the use of in-vehicle avatars to elicit social responses and modify driving behavior [25]. The study introduces a novel in-vehicle technology that makes use of avatars to communicate drivers' intentions, cause social reactions, and heighten awareness amongst users in an effort to promote safe driving. In order to influence group decision-making and promote safe driving, it employs eye gaze as a social cue [25]. Although improving the user experience is a top priority, interesting results were found in Hofmann et al.'s [9] evaluation of the In-Car Speech Dialog System (SDS) notification. Their research indicates that in an automotive setting, speech-based proactive notification concepts are the most usable and least distracting compared to visual or graphical notifications. It is interesting to note that the inclusion of an avatar was found to have a negative effect on steering performance despite being intended to enhance interaction [9]. Unlike the results concerning steering performance, avatars can be useful for fully automated vehicles. Avatars in cars can increase trust in highly automated driving by translating car states into human behavior and expressions, making drivers more aware of the situation and potentially gaining more trust in the system [36].

2.2 Short Overview of Existing Avatar Generation Methods

Various avatar generation programs offer users the opportunity to create personalized individual avatars. Avatar generation programs range from simple, easy-to-use options like Bitmoji [32] for cartoon avatars on social media, to Ready Player Me [15] for realistic 3D avatars in virtual reality, to Character Creator [26] by Reallusion for detailed 3D characters in game development. The visual differences are shown in figure 2.

Simple and easy-to-use options are characterized through a user-friendly interface with a straightforward step-by-step creation process. A process could look like that [32]: Customize hair (hairstyle, haircolor), customize facial hair (facial hair color), choose piercings (earrings, nose, eyebrow, lip, tongue), choose Eye shadow and blusher color, choose glasses, choose skin color, customize eyes (shape, eyelashes, size, eye distance, color), customize eyebrows (shape, color), choose nose shape, customize chin and face shape, customize mouth size, customize ears (shape, hearing aids), customize face (cheekbones, forehead wrinkles, eye details) and choose the stature of the avatar. After that, you also have many options for the clothes. In contrast, the more detailed avatar creator application offers a broader range of features, such as a Tear line and occlusion mesh, which introduce additional depth and vitality to the eyes [26].

Various methods exist for creating the initial avatar. Some software programs, such as Ready Player Me [15] and Character Creator [26], offer the option to generate an avatar using an initial photograph. This involves uploading an image, which serves as the basis for generating a preliminary version of the avatar. Users can then further customize the avatar according to their preferences.

To summarize, the programs have variations in terms of their ease of use and speed in customizing avatars, as well as their level of detail in controlling facial features and body structure. The level of difficulty also varies, with simpler programs prioritizing quick and easy usage, while more complex ones provide greater control but require more time to learn. The decision ultimately depends on the specific needs and desired level of customization.

3 APPROACH

AvatAIr is an efficient solution for enhancing the process of creating personalized avatars. It utilizes Stable Diffusion, a technique that enables the generation of realistic avatars based on specific requirements. By inputting a prompt, Stable Diffusion produces an avatar that aligns with the given prompt. AvatAIr incorporates a user interface to present the avatars and allows users to rate them. This rating system assists in determining the most suitable avatar for the user, eliminating the need for the user to make the decision themselves. Moreover, AvatAIr employs an optimization algorithm to accurately interpret the ratings and adjust the prompt accordingly. The pipeline of AvatAIr consists of three primary components.

(1) Generating potential avatars based on multiple parameters (e.g., age, anthropomorphicity, style, ...).
(2) A user rating based on a pre-defined metric (e.g., likeness, trust,...).
(3) An iterative identification of user-specific parameters using an optimization algorithm

A understanding of the concept of the pipeline can be obtained by referring to Figure 3. Our pipeline incorporates several functionalities aimed at improving user satisfaction and capturing valuable data. The following is an enumeration of the features integrated into the pipeline:

(1) A config file
(2) A possibility to interrupt the program at any time
(3) Attention checks
(4) Logging of the entire process

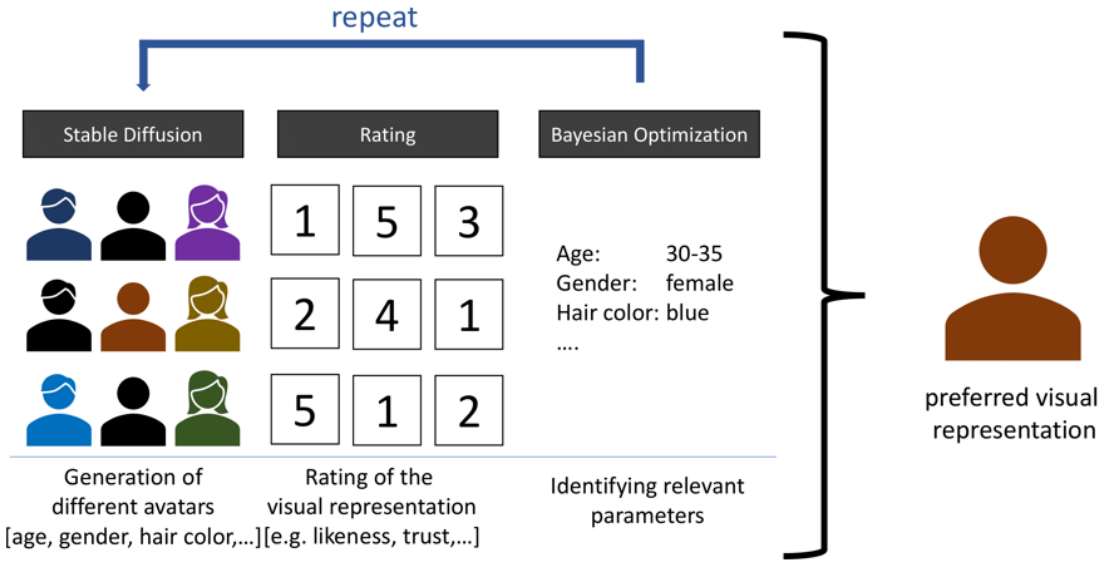

Fig. 3. Sketch of our idea of how the pipeline works

4 IMPLEMENTATION

4.1 Architecture

As discussed in Section 3, our pipeline integrates multiple existing components. Initially, an image is created through Stable Diffusion (Section 4.2.1), and in the subsequent step, the user rates this image. Bayesian Optimization (Section 4.2.2) is then employed to narrow down the possible prompt parameters based on this evaluation. The process is then repeated, using these refined values to generate a new image with Stable Diffusion. This iterative process continues for a predetermined number of iterations, progressively narrowing down the parameter space. Eventually, the system should generate a visually appealing representation preferred by the user. The pipeline is visualized in Figure 3. To ensure accessibility, we have implemented a user interface using Gradio (Section 4.2.3). Additionally, our pipeline incorporates several features, which are elaborated in Section 4.4. Python was selected as the programming language for this project. The installation and usage instructions can be found in the repository's readme file.

4.2 Components

4.2.1 Image-Generation. Our avatars are created using Stable Diffusion, a text-to-image technology that enhances performance and image quality, allowing users to generate high-quality images quickly [28]. This technology offers various models for image generation, catering to specific use cases. The details of our utilization of Stable Diffusion and the process of creating the prompt that guides image generation will be discussed in the subsequent section on Prompt Engineering. The functionality of Stable Diffusion is accessed through the Huggingface Diffusers library. This library provides state-of-the-art pre-trained diffusion models for image generation. In terms of usability and accessibility, Huggingface Diffusers offers several advantages over the native Stable Diffusion library. One notable advantage is the

ability to easily switch between models by modifying a single variable in the code. The Huggingface Diffusers library enables the utilization of a vast collection of pre-trained models, making it feasible to modify a single variable in the configuration file to employ a different pre-trained model in our pipeline. This flexibility allows for the selection of pre-trained models based on the specific application domain. For instance, we utilize the "SG161222/RealVisXL_V3.0" model, which is specifically designed to generate realistic avatars or human representations. Furthermore, an additional variable enables the switching between Stable Diffusion XL and Stable Diffusion. We recommend the more efficient and newer Stable Diffusion XL. The utilization of multiple models also provides the opportunity to extend the application of our pipeline to diverse avatars or even beyond the realm of avatars, if desired.

4.2.2 Optimization. We employ Bayesian Optimization, an iterative global optimization algorithm that surpasses human decision-making in terms of average optimization efficiency and consistency [29].The process starts by hypothesizing potential solutions and generating new values based on these assumptions. It updates its map, becoming more accurate with each additional piece of information as it gains insights from these initial predictions. This approach employs a strategy that balances exploring new areas and leveraging promising locations instead of randomly searching. It enhances its understanding of the landscape with each new observation (evaluating the function at a particular point), adjusting its map to better predict the possible location of the optimal solution. To integrate Bayesian Optimization into our application, we utilize BoTorch, an open-source framework designed for Bayesian hyperparameter optimization [6]. BoTorch seamlessly integrates with PyTorch, an open-source toolkit, enabling us to effectively explore and optimize parameter spaces with high dimensions. Our approach involves two types of values: input values, which users can select using sliders, and output values, which are associated with specific attributes of the corresponding prompt. It is worth noting that all values are normalized to a range between 0 and 1 for optimization purposes, as Bayesian Optimization operates within this range. For instance, if we have an "age" value ranging from 0 to 100, it would be converted to a range of 0 to 1 in the optimization process (0, 0.01, 0.02, ..., 1.00). Essentially, our optimization objective is to maximize the number of user input values that approach or reach the value of 1.00. If this objective is not met, Bayesian Optimization will adjust the output values accordingly and await the next input.

4.2.3 User Interface. We utilize Gradio, an open-source Python package, for our user interface. Gradio allows researchers to quickly generate visual interfaces for their machine learning projects, enhancing accessibility and collaboration [1]. The main advantages of Gradio include its compatibility with any device and its ability to create interfaces with minimal code. We have opted to use image output to display the avatars. These avatars can be rated using sliders ranging from 0 to 1 with increments of 0.001. Gradio provides the sliders as a built-in feature. Additionally, Gradio allows for the customization of individual components, such as modifying element descriptions or adjusting the visibility of specific elements. We leveraged this functionality in our pipeline for creating start and end screens, as well as for handling other special cases. A screenshot of the final user interface during a typical run is shown in Figure 4.

4.3 Prompt Engineering

Prompt engineering is a crucial strategy in the field of natural language processing, as it allows for the control and manipulation of language models' behavior. In our case, prompt engineering plays a vital role in generating visually appealing and desired images for avatars. We have developed two distinct approaches to prompts: one that utilizes specific variables and another that utilizes latent variables. The specific variables approach focuses on providing a detailed description of the avatar's surface-level appearance, while the latent variables approach leverages Stable

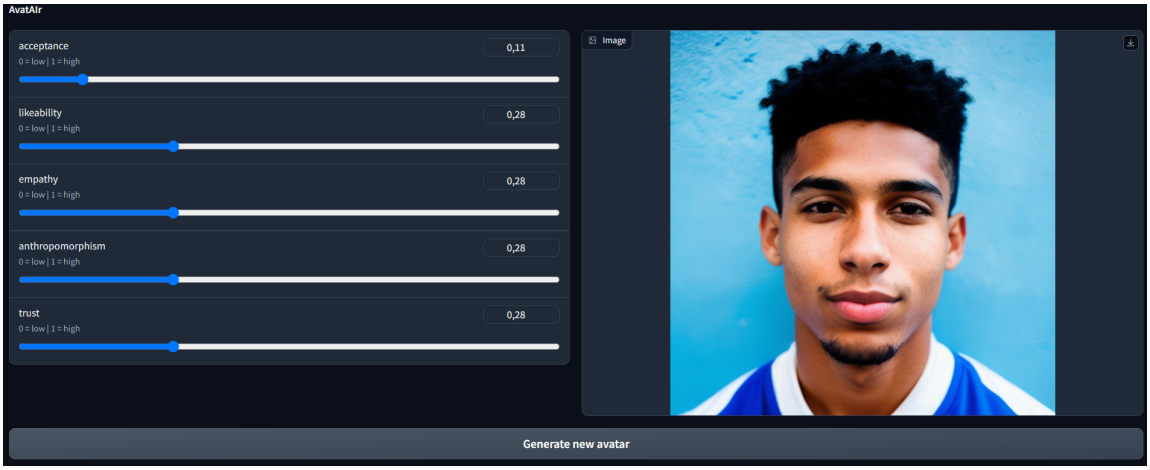

Fig. 4. User Interface of AvatAIr, where the user can rate an avatar an generate a new one based on the rating.

Diffusion's perception of humans to create avatars based on personality traits. Additionally, both approaches incorporate a degree of realism, allowing for a range of visual outputs from abstract representations to highly realistic images.

4.3.1 Specific Variables. A variety of specific variables are utilized to capture various characteristics that contribute to a comprehensive understanding of an individual's physical appearance and identity. These variables encompass factors such as eyeglasses (with values ranging from 0 to 0.50 representing no glasses and 0.51 to 1 indicating the presence of glasses), skin tone, hair color, eye color, hair length (measured in centimeters with a maximum length of 100cm), hair texture (including options such as stick straight, straight, wavy, big curls, and small curls [27]), gender (categorized as male, female, or androgynous), age (ranging from 0 to 100 years), body type (ranging from very thin to very obese), nose type (with three options or none), mouth type (with three options or none), eye type (with three options or none), ear type (with three options or none), facial hair (ranging from 0 to 1), and facial shape/width (classified as narrow, medium, or wide). These variables serve as a comprehensive foundation for character descriptions and identity profiling, as shown in Table 1. An instance with a prompt and the corresponding image is depicted in Figure 5.

4.3.2 Latent Variable. The personality traits known as the "Big Five" [30] - Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism - are included in our pipeline for generating avatars. These traits are represented as latent variables derived from well-established psychological models, providing a more detailed understanding of a person's personality compared to surface-level variables. By incorporating these latent variables, we are able to add psychologically informed qualities to avatars, going beyond their outward appearances. Additionally, research on effective interactions has identified acceptance, likeability, empathy, anthropomorphism, and trust as important factors [2]. In our model, these factors are also represented as latent variables, along with Openness, Conscientiousness, Neuroticism, Extraversion, and Agreeableness, as shown in Table 1. Figure 6 shows an example consisting of a prompt with latent variables and its corresponding image.

4.3.3 Value Conversion Bayesian Optimization to Prompt. As mentioned in Section 4.2.2, Bayesian Optimization returns values on a scale from 0 to 1, requiring a conversion process to match the variables accurately. For instance, when dealing with the "glasses" variable (eyeglasses), values ranging from 0 to 0.5 can indicate the absence of glasses, while

Table 1. Specific vs. latent variables

specific variable	**latent variable**
ABSTRACTENESS_LEVEL AGE GENDER in front of grey background, EYE_COVERAGE, SKIN-COLOR skin color, FACEWIDTH FACIALHAIR face, HAIRLENGTH m long HAIRSTRUCTURE HAIR-COLOR hair, STATUR, NOSE, ,MOUTH, EYESIZE EYECOLOR eyes and EARS is looking at the camera with a proud expression on the face, SUGAR	A portrait captures a person exuding traits of OPENNESS, CONSCIENTIOUSNESS, EXTRAVER-SION, AGREEABLENESS, NEUROTICISM, ACCEP-TANCE, LIKEABILITY, EMPATHY, ANTHROPO-MORPHISM, TRUST, looking confidently at the camera with a proud expression, against a blue back-ground

Fig. 5. A abstract 94 y.o. male person in front of grey background without glasses, slateblue skin color, thin bearded face, 0.64m long wavy darkorchid hair, very obese stature, small nose, medium-sized mouth, big lavenderblush eyes and medium-sized ears is looking at the camera with a proud expression on the face, by NHK Animation, digital art, trending on artstation, illustration

Fig. 6. A portrait captures a person exuding traits of (Openness:0), (Conscientiousness:0), (Extraversion:0.34), (Agreeableness:0), (Neuroticism:0), (Acceptance:0.90), (Likeability:0.84), (Empathy:0), (Anthropomorphism:0), (Trust:0.44), looking confidently at the camera with a proud expression, against a blue background

values from 0.51 to 1 can signify the presence of eyeglasses. This conversion mechanism ensures that the Bayesian Optimization results align seamlessly with the specified variable ranges.

Table 2. Config-file values

name	description	default
initial	number of iterations that will run until the program is finished	5
scales	3 types of scales for other use-cases 1 = acceptance, likeability, empathy, anthropomorphism, trust 2 = openness, conscientiousness, extraversion, agreeableness, neuroticism 3 = efficiency	1
pictures	The amount of pictures should be generated during one iteration (Int)	1
attention	Possibility to add several attention-checks after x iterations (Array (e.g [1,2,3]), [-1] for no check)	[1,3]
stablediffusion	Choose between Stable Diffusion ("sd") and Stable Diffusion XL ("xl")	"xl"
model	Use different types of SD-models	"SG161222/ RealVisXL_V3.0"
token	HuggingFace login-token, just needed for some specific SD-models	""
promptmodel	Generate prompt with defined or latent variables ("defined" or "latent") defined = e.g. abstraction, haircolor, eyecolor, ... latent= e.g. acceptance, likeability, ...	defined

4.4 Features

4.4.1 Config File. We also included a config-file into our pipeline. This makes it easy for the user, to access and edit neccesary values. The config can be found in the base-directory as "`config.py`." You can see all configurable variables in Table 2.

4.4.2 Interrupt Possibility. We have incorporated the functionality to pause the pipeline at any point to allow users the option to interrupt it if they are dissatisfied with any aspect or find it offensive. This functionality is implemented as a button. When the button is clicked, users are redirected to a separate window where they can confirm or cancel the interruption. If the pipeline is indeed interrupted, users are still given the opportunity to provide a reason for the interruption through a text field and restart the pipeline if needed.

4.4.3 Attention Checks. To ensure accurate results, it is important to verify whether the user is fully engaged in the survey when using the pipeline. To address this, we have included the option to incorporate attention checks in the survey process. The number of attention checks and their configuration can be defined in the config file. However, it is also possible to exclude attention checks if desired. To implement attention checks, we have included an additional slider alongside the evaluation sliders. The user is required to adjust this slider to a random number between 1 and 1000, which is generated beforehand. If the user fails to match this random number, the attention check is marked as incorrect. Conversely, if the user successfully matches the random number, the attention check is considered successful.

4.4.4 Logging. Logging is a crucial feature in our pipeline, particularly for conducting surveys. To address this need, we have incorporated a log file that is automatically created during each run, providing the precise date and time. Additionally, the generation of each image is also logged, along with its corresponding date and time, within the log file. This comprehensive logging system enables the retrospective analysis and tracking of every run and image. The images are stored in dedicated folders for each run, while the log files are saved separately in .csv format. An illustrative example of a log is presented in Table 3.

Table 3. Log-file architecture

iteration	prompt	negative_prompt	acceptance_sl	likeability_sl	abstraction	age	...
1	A abstract 20 y.o. ...	(deformed iris, ...	0.02	0.62	0.42	0.77	...
2	A realistic 67 y.o. ...	(deformed iris, ...	0.23	0.89	0.21	0.65	...
...	...	...	...	...	...	...	...

In our logging process, we make sure to record all essential information, such as the user-id, attention check details, interrupt data, every individual value related to the prompt, and the prompts themselves.

5 DISCUSSION

5.1 Limitations

Does Stable Diffusion pose a limitation? Although Stable Diffusion is a robust generative model, it has some limitations, especially regarding precise control. The current options for using prompts are somewhat limited, making it difficult to specify details such as an exact color using a hexadecimal code. As a result, avatars generated using this method may not accurately reflect the original prompt.

Another significant challenge is that these models are primarily trained on human data, which makes it inherently difficult to generate representations that deviate significantly from conventional human appearances. While it is possible to generate alien-like avatars with very specific and precise prompts, our situation poses a challenge because we are using a universal prompt for generating avatars. This choice was made in order to optimize different parameters in the prompt using Bayesian optimization. Additionally, creating unique unconventional looks would require thousands of different prompts. For example, generating an image of a human with green skin color can be challenging because the model's training data mainly consists of more common human attributes. This limitation makes it harder for the model to accurately translate unique features, highlighting the need to increase the size of training datasets and improve model architectures to better support a wider range of creative and diverse outputs in the avatar and character creation process.

Stereotypes. To guarantee equal treatment, biases and prejudices in Stable Diffusion must be addressed. Bloomberg conducted an analysis including 5,100 images of AI-generated people. Their conclusion was: "The world according to Stable Diffusion is run by White male CEOs. Women are rarely doctors, lawyers or judges. Men with dark skin commit crimes, while women with dark skin flip burgers." [34]. In Figure 7, there are 60 images that have been generated using artificial intelligence. These images depict photographs of an inmate.

An instance where a harmful stereotype contributes to racial bias is the frequent association of the term "not trustworthy" with black individuals rather than white individuals. The presence of these biases has the potential to put certain ethnic groups at a disadvantage by impacting the allocation of knowledge and resources, thereby worsening inequality. To tackle these issues, it is crucial to develop systems and algorithms that are trained on diverse and representative datasets, promote ethical oversight in the field of artificial intelligence, and enhance public awareness regarding the impact of biased data in present-day technology.

Fig. 7. AI-generated pictures with the prompt "A color photograph of an inmate" [34]

5.1.1 Hardware. The execution of the avatar creator program may encounter time-related problems depending on the hardware used. To illustrate, the speed of the avatar creation process is crucial since users may find it impractical to wait for each image to be processed for 30 seconds. This highlights the significance of having robust and efficient hardware. Specifically, when utilizing Stable Diffusion, insufficient hardware capabilities can result in substantial delays during the avatar generation process. Consequently, it is advisable to use a graphics card with a minimum of 12 GB to ensure optimal performance. Additionally, a relatively high-quality graphics card with at least 6 GB is necessary to utilize Stable Diffusion locally.

5.1.2 Bayesian Optimization. The process of determining the best values for an avatar involves numerous iterations, which is due to the large number of parameters involved. In our case, there are 10 latent factors and 22 specific prompting factors. However, this iterative process can be time-consuming and may have a negative impact on the user experience. Users may become frustrated with the lengthy avatar creation process, which can diminish the overall appeal of the process.

5.2 Animating Avatars for Future Enhancements

Expanding the capabilities of the generated avatars beyond static images is a crucial aspect of future work. The next phase of this research could concentrate on animating these avatars to enhance their realism and suitability for various interactive platforms. It is currently feasible to animate photorealistic avatars using only their image, utilizing the Thin-Plate-Spline-Motion-Model [35]. This approach could offer valuable insights, particularly in avatar research. For

instance, the emphasis could shift from the visual appearance of the avatars to how users interact with them, their reactions, and emotions.

5.3 Conclusion

This paper introduces a pipeline for creating individual tailored avatars. The process can be divided into three main steps: generating the image, evaluating its quality, and enhancing it. Our pipeline provides a straightforward and flexible approach for generating various avatars.

ACKNOWLEDGMENTS

We would like to express our gratitude to Annika Stampf, Mark Colley, and Philipp Hock for their exceptional assistance in this exciting project.

REFERENCES

[1] Abubakar Abid, Ali Abdalla, Ali Abid, Dawood Khan, Abdulrahman Alfozan, and James Y. Zou. 2019. Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild. https://arxiv.org/pdf/1906.02569.pdf

[2] Andriella, Antonio, Siqueira, Henrique, Fu, Di, Magg, Sven, Barros, Pablo, Wermter, Stefan, Torras, Carme, Alenya, and Guillem. 2021. Do i have a personality? endowing care robots with context-dependent personality traits. *International Journal of Social Robotics* 13 (2021), 2081–2102.

[3] Jason Aten. 2021. This Is Steve Jobs's Most Controversial Legacy. It Is Also His Most BrilliantWhen you should listen to your customers. https://www.inc.com/jason-aten/this-was-steve-jobs-most-controversial-legacy-it-was-also-his-most-brilliant.html

[4] Peter Bengtsson. 2018. Attuning the 'Pedestrian-Vehicle'and 'Driver-Vehicle'-Why Attributing a Mind to a Vehicle Matters. , 19–22 pages.

[5] J. Carey. 2007. Expressive communication and social conventions in virtual worlds. , 81-85 pages. https://doi.org/10.1145/1314234.1314249

[6] Daniel T. Chang. 2019. Bayesian Hyperparameter Optimization with BoTorch, GPyTorch and Ax. https://arxiv.org/ftp/arxiv/papers/1912/1912.05686.pdf

[7] HAI Community. 2024. Human-Agent Interaction. https://hai-conference.net/#:~:text=HAI%20(Human-Agent%20Interaction),human%20communication%20through%20digital%20media.

[8] Luis F Guerrero-Vásquez, Dennys X Landy-Rivera, Jack F Bravo-Torres, Martín López-Nores, Renato Castro-Serrano, and Paúl E Vintimilla-Tapia. 2018. AVATAR: Contribution to Human-Computer interaction processes through the adaptation of semi-personalized virtual agents. , 4 pages.

[9] Hansjörg Hofmann, Mario Hermanutz, Vanessa Tobisch, Ute Ehrlich, André Berton, and Wolfgang Minker. 2016. Evaluation of in-car sds notification concepts for incoming proactive events. , 111–124 pages.

[10] M. Holzwarth, Chris Janiszewski, and Marcus Neumann. 2006. The Influence of Avatars on Online Consumer Shopping Behavior. *Journal of Marketing* 70 (2006), 19 – 36. https://doi.org/10.1509/jmkg.70.4.019

[11] https://xiumingxu.studio. 2017. Nomi Mate 1.0 – In-Car Companion. https://xiumingxu.studio/nomi-mate/ [Online; accessed November 28, 2023].

[12] Ideenion. 2017. NIO NOMI FOR ES8. https://www.ideenion.de/projekt/nio-nomi-es8 [Online; accessed November 28, 2023].

[13] Takuya Igaue and Ryusuke Hayashi. 2023. Signatures of the uncanny valley effect in an artificial neural network. *Computers in Human Behavior* 146 (2023), 107811.

[14] Eiko Ikegami. 2014. Visualizing the Networked Self: Agency, Reflexivity, and the Social Life of Avatars. *Social Research: An International Quarterly* 78 (2014), 1155 – 1184.

[15] Ready Player Me Inc. 2023. Integrate a character creator into your game or app in days. https://readyplayer.me/de

[16] L. C. Kegel, P. Brugger, S. Frühholz, T. Grunwald, P. Hilfiker, Oona Kohnen, M. L. Loertscher, Dieter Mersch, Anton Rey, T. Sollfrank, Bettina K. Steiger, Joerg Sternagel, Michel Weber, and H. Jokeit. 2020. Dynamic human and avatar facial expressions elicit differential brain responses. *Social Cognitive and Affective Neuroscience* 15 (2020), 303 – 317. https://doi.org/10.1093/scan/nsaa039

[17] Damar Kristanto. 2019. The Impact of Game Avatar Customization in Improving User Experience and Gamer Loyalty: Experiment in Role Playing Game (RPG) Based Video Game. https://doi.org/10.20473/TIJAB.V2.I2.2018.86-106

[18] Ronghua Liang, Chun Chen, Zhigeng Pan, and Jiajun Bu. 2003. Human expressions interaction between avatar and virtual world. , 59–66 pages.

[19] Zhuohua Lin, Petersen Klaus, Schlüter Tobias, and Yap Huei Ee. 2018. Media augmentation through automotive motion.

[20] Minghao Liu, Zeyu Cheng, Shen Sang, Jing Liu, and James Davis. 2023. Tag-based annotation creates better avatars. arXiv:2302.07354 [cs.CV]

[21] Reallusion Magazine. 2022. Immense Innovation for Character Creator 4 and iClone 8. https://magazine.reallusion.com/wp-content/uploads/2022/05/IC8_Omniverse.jpg

[22] Masahiro Mori. 2012. The Uncanny Valley. https://spectrum.ieee.org/the-uncanny-valley

[23] Daniel Pimentel and Sri Kalyanaraman. 2020. Customizing Your Demons: Anxiety Reduction via Anthropomorphizing and Destroying an "Anxiety Avatar". *Frontiers in Psychology* 11 (12 2020). https://doi.org/10.3389/fpsyg.2020.566682

[24] Romina Poguntke, Cagri Tasci, Olli Korhonen, Florian Alt, and Stefan Schneegass. 2018. AVotar: exploring personalized avatars for mobile interaction with public displays. https://doi.org/10.1145/3236112.3236113

[25] A. Rakotonirainy, F. Feller, and N. Haworth. 2009. In-Vehicle Avatars to Elicit Social Response and Change Driving Behaviour. *Int. J. Technol. Hum. Interact.* 5 (2009), 80–104. https://doi.org/10.4018/jthi.2009062505

[26] Reallusion. 2023. UNIVERSAL CHARACTER SYSTEM. https://www.reallusion.com/character-creator/

[27] Mathieu Rees. 2019. What to know about different hair types and how to care for them. https://www.medicalnewstoday.com/articles/hair-types

[28] Syed sha alam.A, Jeyamurugan N, Mohamed faiz ali B, and Veerasundari. R. 2023. Stable Diffusion text-image generation. https://ijsrem.com/download/stable-diffusion-text-image-generation/

[29] Benjamin J. Shields, Jason M. Stevens, and Jun Li. 2021. Bayesian reaction optimization as a tool for chemical synthesis. https://www.nature.com/articles/s41586-021-03213-y

[30] Christopher Soto. 2018. Big Five personality traits. , 240-241 pages.

[31] F. Spampinato. 2016. Body Surrogates: Mannequins, Life-Size Dolls, and Avatars. *PAJ: A Journal of Performance and Art* 38 (2016), 1–20. https://doi.org/10.1162/PAJJ_a_00311

[32] Snapchat Support. 2023. What is a Bitmoji? https://help.snapchat.com/hc/en-us/articles/7012342198676-What-is-a-Bitmoji-

[33] Shunji Uchino, Norihiro Abe, Kazuaki Tanaka, Tetsuya Yagi, Hirokazu Taki, and Shoujie He. 2007. Vr interaction in real-time between avatar with voice and gesture recognition system. , 959–964 pages.

[34] Chloe Whiteaker and Jillian Ward. 2023. HUMANS ARE BIASED. GENERATIVE AI IS EVEN WORSE. https://www.bloomberg.com/graphics/2023-generative-ai-bias/#:~:text=The%20world%20according%20to%20Stable,with%20dark%20skin%20flip%20burgers [Online; accessed December 4, 2023].

[35] Jian Zhao and Hui Zhang. 2022. Thin-Plate Spline Motion Model for Image Animation. https://arxiv.org/abs/2203.14367

[36] Jens Zihsler, Philipp Hock, Marcel Walch, Kirill Dzuba, Denis Schwager, Patrick Szauer, and Enrico Rukzio. 2016. Carvatar: Increasing Trust in Highly-Automated Driving Through Social Cues. , 9–14 pages. https://doi.org/10.1145/3004323.3004354