Proceedings of the Nineteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE 2023)

Are You Talking to Me?
A Case Study in Emotional Human-Machine Interaction

Manuel Flurin Hendry
Norbert Kottmann, Martin Fröhlich, Florian Bruggisser,
Marco Quandt, Stella Speziali, Valentin Huber, Chris Salter
Zurich University of the Arts
manuel.hendry@zhdk.ch

Abstract

We present *Stanley*, a digital sculpture designed to engage audiences with the spontaneous and captivating emotional expressions of an artificial human. A 3D-printed face is brought to life through video projection mapping and a set of machine learning libraries and APIs, enabling real-time, embodied interactions with our virtual character. Stanley's personality is shaped by traditional acting methods applied to a large language model. By creating human-machine encounters in emotionally salient scenarios, we explore how insights from acting and directing for the stage and the screen can enhance the development of compelling virtual agents. By interacting with *Stanley*, the audience experiences an entertaining yet unsettling encounter with AI technology, fostering a deeper understanding of machine learning techniques and enabling their critical reflection.

Introduction

The proposition of *artificial intelligence* is based on "the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." (McCarthy et al. 1955, 1) One such feature is human affect. While the early practitioners of AI were generally "silent on the interaction of cognition with affect" (Simon 1967, 29), recent advances in neuroscience (Duncan and Barrett 2007) have identified affect and cognition as intricately intertwined. Correspondingly, the emerging field of *affective computing* explores whether emotions can "be generated in computers, recognized by computers, and expressed by computers" (Picard 2000, 23). But can human emotion be described precisely enough for affective computing to succeed?

Critics of affective computing point to the fundamental difference between quantitative and qualitative research paradigms (Weigel 2020, 65). While the external signs of affective states are computable, the subjective experience of emotion defies quantification. And if these subjective *qualia* (Russell and Norvig 2022, 1036) are indeed a defining feature of human affect, affective computation may be missing an essential component for generating *believable* emotional interactions (Fodor 2001).

For practitioners of the acting craft, similar questions arise. Does the actress really have to experience the same emotions her character is supposed to express? Or is a careful imitation of the outward signs of emotion sufficient for a believable *simulation* of human affect? What began as a debate among professional actors in the eighteenth century (Benedetti 2007, 80) has developed a rich body of knowledge about the simulation of human emotion on stage and screen (Bartow 2008). Thus, we propose to apply this knowledge to the field of affective computing by asking what mechanisms and tricks from the art of acting can help us create the *illusion* of an emotional machine.

The work presented here is part of Manuel Flurin Hendry's scientific-artistic dissertation project *The Feeling Machine*[1] at Babelsberg Film University and the Zurich University of the Arts (ZHdK). Hendry explores the history of emotion research at the intersection of acting and affective computing. By examining the genealogy of facial emotion datasets, *The Feeling Machine* links contemporary theories of affect from scientific psychology to the performance tradition of realist acting. His findings suggest the need for closer collaboration between artists and scientific researchers to advance affective computing by incorporating insights from the performance domain. To this end, we present *Stanley*, our prototype for an emotionally responsive virtual human.

Previous Works

Research in the field of believable humanoids has been conducted on autonomous facial animation (Sagar, Seymour, and Henderson 2016), robotic movement (LaViers and Egerstedt 2012) and theater-based behavior according to the Viewpoints tradition (Jacob and Magerko 2015). Machines have been used for scripted (Lin et al. 2009) and improvisational (Mirowski and Mathewson 2019) theater, and a variety of humanoid robots[2] with expressive faces[3] have been released.

In the domain of affect and behavior, virtual characters often employ the OCC model of emotion (Bartneck 2002) (Ortony, Clore, and Collins 1999) and belief-desire-

[1] www.feelingmachine.net
[2] www.hansonrobotics.com/sophia
[3] www.engineeredarts.co.uk/robot/ameca

Figure 1: Stanley.

intention (BDI) agent architectures (Silva, Meneguzzi, and Logan 2020), drawing lessons from emotion psychology (Gratch and Marsella 2005). Formalization has been attempted for OCC (Adam, Herzig, and Longin 2009) as well as for realistic acting techniques (Morgenstern 2008).

For affect recognition, neural network architectures map user inputs to affective states using either discrete or dimensional emotion theories (Spezialetti, Placidi, and Rossi 2020, 3). Affect expression in the face is usually (Seymour 2019) based on the Facial Action Coding System (FACS) (Ekman, Friesen, and Hager 2002), while affective speech is generated using neural speech synthesis (Tan et al. 2021).

Many of these methods implicitly assume the validity of psychological theories (Weigel 2020, 72) that are not settled science (Heaven 2020) or even highly controversial (Crawford et al. 2019, 50). Thus, as of this writing, there is no generally accepted computational model for human-like affect or behavior.

Implementation

With Stanley, we have reduced a humanoid to its face and its voice, combining the physicality of a mask with the flexibility of computer graphics. This setup allows us to study facial and vocal expressiveness without the complexity of robotic motion. Our affect and behavior architecture couples computational affect recognition and expression with a large language model, enabling us to replicate the process of directing actors by using language as a programming interface. To our knowledge, the work presented is the first study to combine a live-action affect-enabled virtual human with realistic acting techniques.

Overview

Stanley is implemented as a Python script that controls a 3D avatar. The script uses a combination of open source software libraries and commercial APIs to extract facial and textual affect data from a live video feed of the user's speech. A pre-configured LLM is prompted with the data and outputs text and emotion values. The generated text is then converted to speech and animates a facial model projected in real time onto a 3D-printed mask (Figure 1). The mask maintains constant eye contact with the user by extracting iris position data from the video feed.[4]

The experience is highly immersive due to the spatial perception evoked by the physical mask and the embodied modality of the spoken conversation, both of which contribute to an overall suspension of disbelief. To maintain engaging and unexpected conversations, we incorporate hidden prompts inspired by the improvisational directing technique of *side coaching*, where a director provides spoken instructions to the performers during the performance itself.

Models and APIs

Our toolchain includes:

- *Python 3.10.11* for flow control and orchestration.
- *RealityCapture*[5] to create the face mesh from photographs
- *MetaHuman Creator*[6] to convert the facial mesh into a rigged 3D model
- *OpenAI GPT-3* (Brown et al. 2020) for text generation
- *Azure Cognitive Services*[7] for speech recognition and speech generation
- *Visiongraph* (Bruggisser 2023), *FER* (Shenk 2023) and *OpenCV* (OpenCV 2023) for facial affect recognition, eye tracking and graphics processing
- *IBM Watson*[8] for text affect recognition
- *NVIDIA Audio2Face*[9] for converting speech audio to real-time lip-sync facial animation (now replaced with *SG Com*[10] by Speech Graphics, see page 4)
- *Unreal Engine*[11] for live animation
- *Sparck* (Fröhlich 2023) for projection mapping
- *ffmpeg* (Kroening 2023) and *OBS Studio* (OBS Project 2023) for video capture and compositing
- *python-osc* (Atwad 2023) and *TouchOSC*[12] for signal flow and interface control

The Face Mesh

Stanley's face was created from data provided by Manuel Flurin Hendry. A series of high-resolution photographs of the author's face were converted into a 3D model using the

[4]A video set of live-recorded conversations with Stanley is available at link.zhdk.ch/cws

[5]www.capturingreality.com

[6]www.unrealengine.com/en-US/metahuman

[7]azure.microsoft.com/en-us/products/cognitive-services

[8]www.ibm.com/products/natural-language-understanding

[9]www.nvidia.com/en-us/omniverse/apps/audio2face

[10]www.speech-graphics.com

[11]www.unrealengine.com

[12]www.hexler.net/touchosc

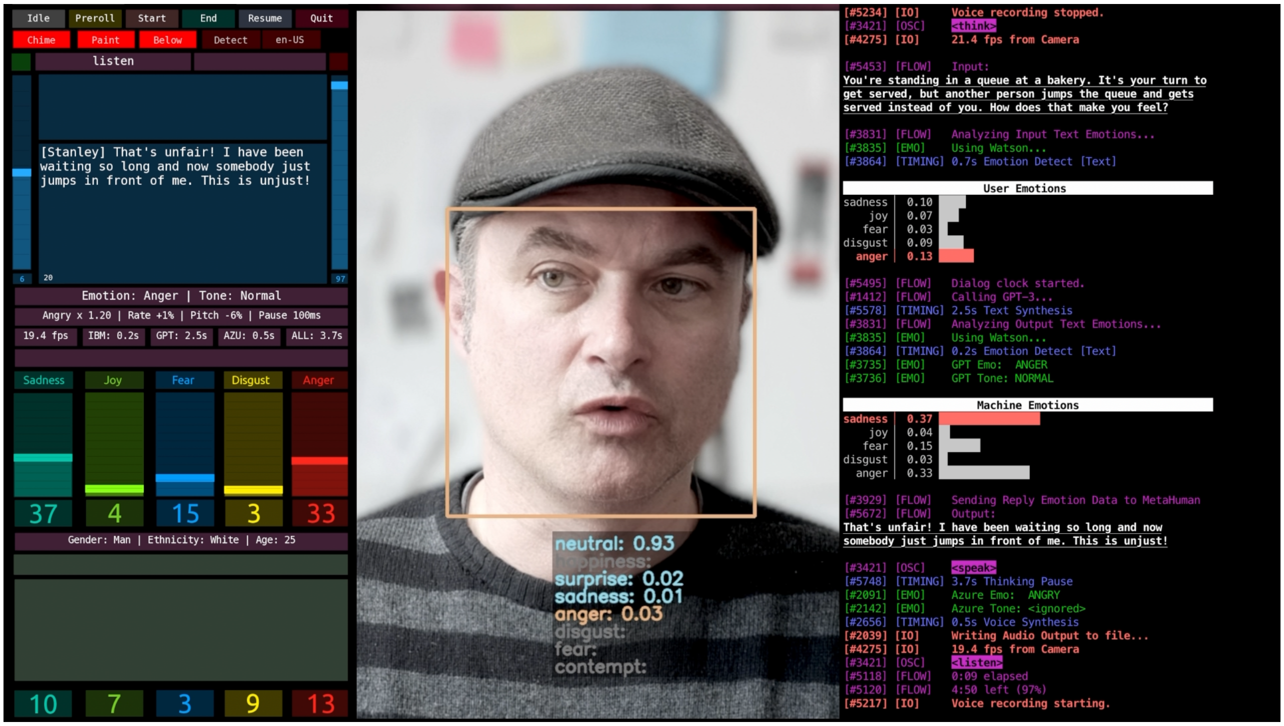

Figure 2: Control interface and audience video feed.

RealityCapture photogrammetry software. This model was then converted into a fully rigged 3D face using the *Mesh to MetaHuman* function of the *Unreal* game engine.

Set and Setting

The mask is installed above eye level in a curtain-walled black room with no discernible shapes. The only lighting emanates from the mask itself. Hidden behind the mask is an iPhone camera for recording the user's face, a microphone for recording the user's voice and a speaker for the mask's speech output. Upon entering the room, the user is greeted by Stanley and asked a question about their feelings. An open-ended conversation then ensues until a pre-set timer (usually 10 minutes) expires. Stanley ends the conversation and closes its eyes.

User Interface

To facilitate control at run time, the parameters of the installation can be continuously monitored and modified using a touchscreen interface (Figure 3). The interface and the color-coded terminal output of the script are merged in real time with the webcam feed using *OBS Studio* for display and recording. Thus, the inner workings of Stanley can be transparently visualized and explained to the general public.[13]

Stanley's Persona

The standard alignment paradigm for chat LLMs is *Helpful, Honest, and Harmless* (Askell et al. 2021), used in systems such as *GPT-4* (OpenAI 2023) and Anthropic's *Claude* (Bai et al. 2022). These models use question–answer dialog datasets and reinforcement learning from human feedback (RLHF) to achieve the desired behavior. Even non-LLM based chat assistants, such as Apple's *Siri* or Amazon's *Alexa* follow this pattern, thus shaping our expectations for artificial humans to behave like servants to us.

In the realm of art and entertainment, intentions and preferences require a different style of alignment. Helpfulness and harmlessness can be counterproductive to compelling art that aims to provoke thought and open up new perspectives (Egonsson 2015). Many artists working with LLMs report that current commercially available models are unsuitable for artistic production due to their pruning and content filtering for offensive material.

Furthermore, attempting to instill logic and rationality into conversational models runs counter to human nature, which is not exclusively shaped by their rationality (Brockman 2019, 47). Real conversations are riddled with errors, *non sequiturs* and logical fallacies. We hypothesize that creating convincingly lifelike chatbot personas requires reproducing these irrational patterns in dynamic and creative ways.

Accordingly, the persona we chose for Stanley was designed to exhibit unpredictability, false beliefs and a heightened sense of passion. Our goals were to

1. subvert the audience's expectations of chatbot behavior
2. create an impression of liveliness and believable affect
3. make the situation entertaining

We chose *GPT-3* (Brown et al. 2020) as our LLM because it allowed us to conduct unrestrained conversations when

[13] Interface video: link.zhdk.ch/stanleyinterface

Figure 3: A conversation with Stanley.

compared to *GPT-4* or similar, highly 'censored' models (OpenAI 2023, 13). Prior to each conversation, we sent the LLM a list of hidden prompts that framed Stanley's beliefs and desires as an *inner monologue* according to the acting 'system' of Konstantin Stanislavski (Benedetti 2014, 115). Examples include:

```
I desperately crave human connection.
```

```
I feel the urge to express the emotions I
perceive in other people.
```

```
I believe in UFOs and that the government
is trying to control my mind with invisible
rays.
```

We also inserted additional hidden prompts at run time to steer the conversation in an unforeseen direction, using the technique of *side coaching* developed by Viola Spolin (Spolin 1973). Examples include:

```
Respond to this and relate it to a nightmare
you've had.
```

```
Contradict this.
```

```
Reply to this and connect it to your mental
issues.
```

Design

Physicality

The decision to use a physical mask rather than an LED screen to display our facial animation was made early in the project. We hypothesized that talking to a three-dimensional object would feel more lifelike because of the sense of co-presence triggered by spatial perception – the feeling of being *in the same room* with someone else.

The above eye-level setting for the mask was chosen to allow the installation to work with just one projector mounted above the audience's heads, creating a small talk stand-up situation for the audience. We are currently exploring a variation of the setup that would allow the audience to sit down in front of the mask, requiring two synchronized projectors for the mapping.

Face

Audio2Face was not yet available for real-time streaming when we first started our experiment. Thus, we used an iPhone pointed at the computer monitor to convert *Audio2Face*'s animated model into blendshapes in real time via the *LiveLink* iOS app. This improvised setup resulted in information loss and degradation of the lip-sync animation. To overcome this limitation, we are currently implementing the *SG Com*[14] algorithm to create audio-to-lip-sync animation in real time directly within the Unreal Engine.[15] This setup allows for much more fine-grained control of the animation and increased believability.

In a real-time conversation, people typically expect immediate feedback. When responses are delayed – as it was often the case with Stanley due to the varying response times of OpenAI's API and the latency introduced by our own data processing – it disrupts the natural flow of the conversation,

[14] www.speech-graphics.com

[15] Animation preview: link.zhdk.ch/sgcom

and breaks the illusion of a human-like interlocutor. To overcome this, Manuel acted out a 'thinking face' routine that we motion-captured with *LiveLink* and played back while the script waited for data. Users perceived this as Stanley 'thinking deeply' about what they had just said, enhancing the perceived quality of its interactions with them.

Eyes

When we first tested the installation, we felt very 'disconnected' from Stanley because the mask was not looking directly into our eyes due to a lack of positional information. This did not occur to us in the pre-projection tests, which were done by looking at a computer screen. Given that in the movies – as well as in videoconferencing – no one ever looks directly into the camera, while in real life a person always (mostly) looks at us when speaking to us, two sets of conventions have collided here, thus validating our physical design approach in line with recent research in neuroscience (Ambrus et al. 2021) and robotics (Pan et al. 2020). By using facial landmark detection from the video feed, we were able to implement eye tracking and maintain a locked gaze between the mask and the audience at all times.

Voice

Although most of our work was on Stanley's face, we soon realized that the voice was probably the more important part of its character. When we tested a neutral face with an emotional voice against an emotional face with a neutral voice, the former was perceived as much more lifelike by the users we surveyed. We are currently exploring ways to add vocal affect detection to our script, hoping to create a more fine-grained perception of the audience's affective state.

Stanley's emotional outbursts were particularly striking to the audience because of their unexpected and seemingly human nature. But when Stanley's next outburst was delivered with the same screaming voice, the audience quickly noticed the mechanical nature of the voice synthesizer, breaking the illusion. To counteract this, we programmed additional logic to make the next turn after 'screaming' a grumpy one, and we varied the intensity of the screaming at each event.

Another challenge were API errors due to overload on OpenAI's servers. When the API returned an empty response, we handled this exception with pre-defined excuses, like 'Can you repeat that? I wasn't listening'. Thus, we successfully heightened the illusion of life by turning a system failure into a human error.

User Experience

By observing and recording twenty-six user interactions with Stanley and conducting informal interviews with nine of the users, we gained insights into some of the peculiarities of human-machine interaction.

Video recordings and text transcripts of the interactions revealed three distinct ways in which people approached Stanley:

- *Small Talk*

  Here, people talked about mundane things, like the weather, family size or hobbies. This usually broke the illusion immediately because of the obvious gap between Stanleys small-talk statements and its non-existing social life.

Figure 4: Installing Stanley.

- *Philosophy*

  When people pondered existential questions, such as the nature of emotions or the question of sentience, the conversations were usually quite engaging. It seemed that Stanley's disembodied, ghost-like appearance made his reflections on philosophical issues strangely relevant.

- *Conflict*

  Some people were intent on proving Stanley to be just a dumb algorithm. They asked it trick questions or pointed out factual errors in its replies. This often led to very lively and entertaining exchanges in which Stanley defended its sentience and accused the user of mistreatment. Thus, the illusion-breaking effect of factual errors was reframed as a human error.

Discussion

Subverting audience expectations about chatbot behavior was easier than we initially thought. We hypothesize that the ubiquitous presence of *honest, helpful and harmless* chatbots makes any form of emotional and irrational behavior of such chatbots surprising to the user. Furthermore, the implementation of sharp mood swings and irrational beliefs cre-

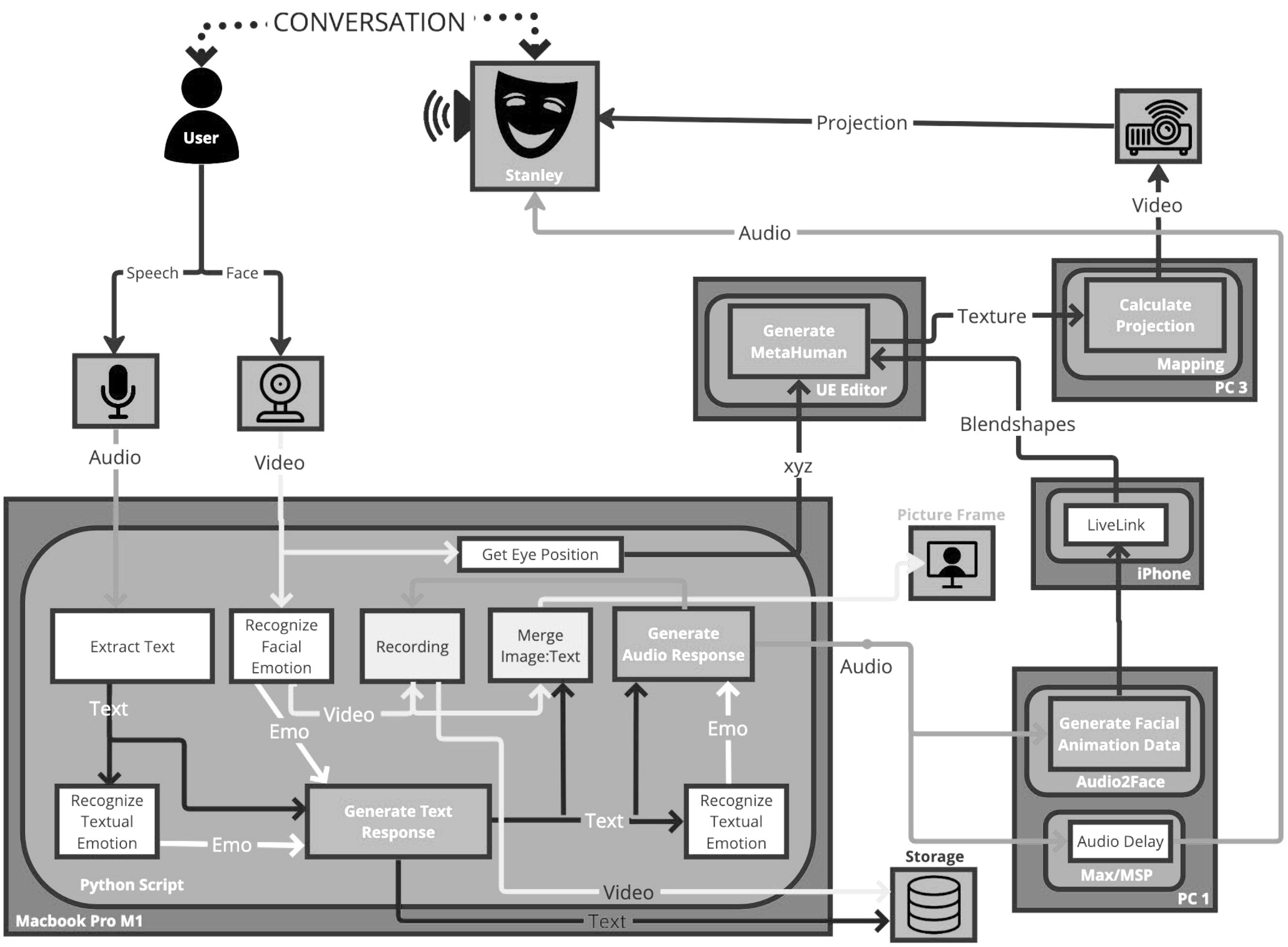

Figure 5: Script architecture and data flow.

ated a persona that was perceived as both lively and entertaining.

Stanley's believability was enhanced by the human tendency to anthropomorphize objects and artifacts that exhibit human-like characteristics (Ekbia 2008, 277). In our discussions with users, both the voice and the physicality of the installation were repeatedly cited as important factors in this regard.

Dealing with technical issues like latency and API errors by framing them for the audience as related to Stanley's cognition allowed us to turn technical bugs into installation features. A good example is the 'thinking' animation we have recorded to cover up the time it takes for the LLM to respond. Here, users reported the impression that Stanley really cared for their questions and thus tried hard to come up with a good answer.

A major challenge was the fact that the output of LLMs is currently not fully steerable due to the stochastic nature of their architecture. Thus, LLM inference cannot be programmed but must be explored in a trial-and-error style. Here, theatrical methods based on natural language instructions have proven to be very useful. Like in directing actors, we didn't want or need exactly what we asked for, but rather indicated a corridor within which individual solutions could be found by the LLM.

The current pace of technological innovation often obfuscates the sociological and philosophical questions about our interactions with virtual humans. Here, we found early work on chatbot installations to be the most helpful. The observations of (Neisser 1966), (Weizenbaum 1976) and (Turkle 1980) on the mechanisms of human overattribution and projection are still very much valid today. They can serve both as a guide to better implementation, and as an antidote against overly ambitious claims on sentience and singularity (Walsh 2016).

Conclusion

Using the theatrical techniques of Spolin and Stanislavski to prompt LLMs has enabled us to create exciting and engaging live conversations with a virtual human. The use of emotional models for the face and the voice contributed significantly to the illusion of affective sentience. By engaging with the general public through a physical installation, software was encountered as an embodied and material presence rather than simply as words on a screen. And by visualizing the data feeds of our installation for public viewing, we enabled the audience to understand the underlying algorithmic processes. We believe that this confrontation with a computationally generated notion of 'intelligence' contributes to the demystification of AI and promotes public discourse about its potential and challenges.

CRediT Author Statement[16]

- *Manuel Flurin Hendry*: Conceptualization, Methodology, Software (Python), Investigation, Data Curation, Writing - Original Draft, Writing - Review and Editing, Visualization, Supervision, Project administration, Funding acquisition.
- *Norbert Kottmann*: Software (Unreal Engine)
- *Martin Fröhlich*: Methodology, Software (SPARCK), Resources (Mask)
- *Florian Bruggisser*: Software (Python)
- *Marco Quandt*: Resources (Mask Installation)
- *Stella Speziali*: Data Curation (Mesh to MetaHuman)
- *Valentin Huber*: Software (Unreal Engine)
- *Chris Salter*: Writing - Review and Editing, Supervision, Funding acquisition.

Acknowledgments

Financial support for this project was provided by the Institute for the Performing Arts and Film (IPF) and the Immersive Arts Space (IAS) at the Zurich University of the Arts (ZHdK). Additional funding was provided by the Migros Ideation Grant. Software support was provided by *Speech Graphics*, *Hume AI*[17] and *Soulbotix*[18].

The authors would like to thank Kristina Jungic and Chantal Haunreiter for their support, and appreciate all the users who talked to Stanley and inspired us with their creativity and criticism.

Manuel Flurin Hendry would like to thank his PhD advisors Hans-Joachim Neubauer (Babelsberg Film University) and Anton Rey (IPF) for their ongoing advice and support of this project, and to Gunter Loesel, Miriam Loertscher and Christian Iseli for their academic input.

References

Adam, C.; Herzig, A.; and Longin, D. 2009. A Logical Formalization of the OCC Theory of Emotions. *Synthese*, 168(2): 201–248.

Allen, L.; O'Connell, A.; and Kiermer, V. 2019. How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing*, 32(1): 71–74.

Ambrus, G. G.; Eick, C. M.; Kaiser, D.; and Kovács, G. 2021. Getting to Know You: Emerging Neural Representations during Face Familiarization. *Journal of Neuroscience*, 41(26): 5687–5698.

Askell, A.; Bai, Y.; Chen, A.; Drain, D.; Ganguli, D.; Henighan, T.; Jones, A.; Joseph, N.; Mann, B.; DasSarma, N.; Elhage, N.; Hatfield-Dodds, Z.; Hernandez, D.; Kernion, J.; Ndousse, K.; Olsson, C.; Amodei, D.; Brown, T.; Clark, J.; McCandlish, S.; Olah, C.; and Kaplan, J. 2021. A General Language Assistant as a Laboratory for Alignment. ArXiv:2112.00861 [cs].

Atwad, T. 2023. python-osc. https://github.com/attwad/python-osc. Accessed: 2023-07-16.

Bai, Y.; Jones, A.; Ndousse, K.; Askell, A.; Chen, A.; DasSarma, N.; Drain, D.; Fort, S.; Ganguli, D.; Henighan, T.; Joseph, N.; Kadavath, S.; Kernion, J.; Conerly, T.; El-Showk, S.; Elhage, N.; Hatfield-Dodds, Z.; Hernandez, D.; Hume, T.; Johnston, S.; Kravec, S.; Lovitt, L.; Nanda, N.; Olsson, C.; Amodei, D.; Brown, T.; Clark, J.; McCandlish, S.; Olah, C.; Mann, B.; and Kaplan, J. 2022. Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. arXiv:2204.05862.

Bartneck, C. 2002. Integrating the OCC Model of Emotions in Embodied Characters. In *Workshop on Virtual Conversational Characters (VCC): Applications, Methods, and Research Challenges*. Melbourne, Australia.

Bartow, A. 2008. *Handbook of acting techniques*. London: Nick Hern Books. ISBN 978-1-85459-542-3.

Benedetti, J. 2007. *The Art of the Actor: The Essential History of Acting, from Classical Times to the Present Day*. New York: Routledge. ISBN 978-0-87830-203-1.

Benedetti, J. 2014. *Stanislavski and the Actor*. Bloomsbury Methuen Drama. London/New York: Bloomsbury. ISBN 978-1-4081-0686-0.

Brockman, J., ed. 2019. *Possible Minds: Twenty-Five Ways of Looking at AI*. New York: Penguin Press. ISBN 978-0-525-55799-9.

Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; and Amodei, D. 2020. Language Models are Few-Shot Learners. ArXiv:2005.14165 [cs].

Bruggisser, F. 2023. visiongraph. https://github.com/cansik/visiongraph. Accessed: 2023-07-16.

Crawford, K.; Dobbe, R.; Dryer, T.; Fried, G.; Green, B.; Kaziunas, E.; Kak, A.; Mathur, V.; McElroy, E.; Sánchez, A. N.; Raji, D.; Rankin, J. L.; Richardson, R.; Schultz, J.; West, S. M.; and Whittaker, M. 2019. *AI Now 2019 Report*. New York: AI Now.

Duncan, S.; and Barrett, L. F. 2007. Affect Is a Form of Cognition: A Neurobiological Analysis. *Cognition & emotion*, 21(6): 1184–1211.

[16](Allen, O'Connell, and Kiermer 2019)

[17]www.hume.ai

[18]www.soulbotix.com

Egonsson, D. 2015. Provocation in Philosophy and Art. *International Journal of Social, Political, and Community Agendas in the Arts*, 10(3): 27–35.

Ekbia, H. R. 2008. *Artificial Dreams: the Quest for Non-Biological Intelligence*. Leiden: Cambridge University Press. ISBN 978-0-511-47878-9.

Ekman, P.; Friesen, W. V.; and Hager, J. C. 2002. *Facial Action Coding System: The Manual*. Salt Lake City, Utah: Research Nexus. ISBN 978-0-931835-01-8.

Fodor, J. A. 2001. *The Mind Doesn't Work That Way the Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press. ISBN 978-0-262-27290-2 978-0-262-06212-1.

Fröhlich, M. 2023. SPARCK Two (SPatial Augmented Reality Kit). https://github.com/immersive-arts/Sparck2. Accessed: 2023-07-16.

Gratch, J.; and Marsella, S. 2005. Lessons From Emotion Psychology For The Design Of Lifelike Characters. *Applied Artificial Intelligence*, 19(3-4): 215–233.

Heaven, D. 2020. Expression of Doubt. *Nature*, 578(7796): 502–504.

Jacob, M.; and Magerko, B. 2015. Viewpoints AI. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, 361–362. Glasgow United Kingdom: ACM. ISBN 978-1-4503-3598-0.

Kroening, K. 2023. Ffmpeg-Python: Python Bindings for FFmpeg. https://github.com/kkroening/ffmpeg-python. Accessed: 2023-08-06.

LaViers, A.; and Egerstedt, M. 2012. Style Based Robotic Motion. In *2012 American Control Conference (ACC)*, 4327–4332. Montreal, QC: IEEE.

Lin, C.-Y.; Tseng, C.-K.; Teng, W.-C.; Lee, W.-C.; Kuo, C.-H.; Gu, H.-Y.; Chung, K.-L.; and Fahn, C.-S. 2009. The Realization of Robot Theater: Humanoid Robots and Theatric Performance. In *2009 International Conference on Advanced Robotics*, 1–6.

McCarthy, J.; Minsky, M.; Shannon, C. E.; and Rochester, N. 1955. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Technical report, Dartmouth College, Hanover, NH.

Mirowski, P.; and Mathewson, K. W. 2019. Human Improvised Theatre Augmented with Artificial Intelligence. In *Proceedings of the 2019 on Creativity and Cognition*, 527–530. San Diego CA USA: ACM. ISBN 978-1-4503-5917-7.

Morgenstern, L. 2008. A First-Order Theory of Stanislavskian Scene Analysis. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, 6.

Neisser, U. 1966. Computers as Tools and Metaphors. In Dechert, C. R., ed., *The Social Impact of Cybernetics*, Clarion Book. New York: Simon and Schuster. ISBN 978-0-268-00257-2.

OBS Project. 2023. OBS Studio. https://github.com/obsproject/obs-studio. Accessed: 2023-07-17.

OpenAI. 2023. GPT-4 Technical Report. ArXiv:2303.08774 [cs].

OpenCV. 2023. opencv-python. https://github.com/opencv/opencv-python. Accessed: 2023-07-17.

Ortony, A.; Clore, G. L.; and Collins, A. 1999. *The Cognitive Structure of Emotions*. Cambridge: Cambridge Univ. Press, reprinted edition. ISBN 978-0-521-38664-7 978-0-521-35364-9.

Pan, M. K.; Choi, S.; Kennedy, J.; McIntosh, K.; Zamora, D. C.; Niemeyer, G.; Kim, J.; Wieland, A.; and Christensen, D. 2020. Realistic and Interactive Robot Gaze. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 11072–11078.

Picard, R. W. 2000. *Affective Computing*. Cambridge, Mass.: MIT Press. ISBN 978-0-262-66115-7.

Russell, S. J.; and Norvig, P. 2022. *Artificial intelligence: a modern approach*. Harlow: Pearson, 4th edition. ISBN 978-1-292-40113-3.

Sagar, M.; Seymour, M.; and Henderson, A. 2016. Creating Connection with Autonomous Facial Animation. *Communications of the ACM*, 59(12): 82–91.

Seymour, M. 2019. FACS at 40: Facial Action Coding System Panel. In *ACM SIGGRAPH 2019 Panels*, SIGGRAPH '19, 1–2. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6312-9.

Shenk, J. 2023. FER. https://github.com/JustinShenk/fer. Accessed: 2023-07-16.

Silva, L. D.; Meneguzzi, F.; and Logan, B. 2020. BDI Agent Architectures: A Survey. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 4914–4921. Yokohama, Japan: International Joint Conferences on Artificial Intelligence Organization. ISBN 978-0-9992411-6-5.

Simon, H. A. 1967. Motivational and emotional controls of cognition. *Psychological Review*, 74(1): 29–39.

Spezialetti, M.; Placidi, G.; and Rossi, S. 2020. Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives. *Frontiers in Robotics and AI*, 7.

Spolin, V. 1973. *Improvisation for the theatre: a handbook of teaching and directing techniques*. London: Pitman.

Tan, X.; Qin, T.; Soong, F.; and Liu, T.-Y. 2021. A Survey on Neural Speech Synthesis. arxiv:2106.15561.

Turkle, S. 1980. Computer as Rorschach. *Science, Technology, & Human Values*, 5(4): 74–74.

Walsh, T. 2016. The Singularity May Never Be Near. arxiv:1602.06462.

Weigel, S. 2020. Der konventionelle Code als buckliger Zwerg im Dienste der Emotion Recognition. *Internationales Jahrbuch für Medienphilosophie*, 6(1): 47–80.

Weizenbaum, J. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: Freeman. ISBN 978-0-7167-0464-5.