THE AIS HAVE IT? HACKING INTO THE AI AVATAR DREAM

Preprint · December 2022

DOI: 10.13140/RG.2.2.22539.77605

1 author:

Alexander Gerner
University of Lisbon
70 PUBLICATIONS **18** CITATIONS

THE AIS HAVE IT? HACKING INTO THE AI AVATAR DREAM[1]

ALEXANDER GERNER

1. TOWARDS THE AI AVATAR DREAM

Avatars, artificial persons, or graphic placeholders for human beings are used in various functions in today's *cultures of digitality*. Avatars range from cartoon figures – starting with "Clippy," the famously annoying Microsoft Word paperclip assistant whose googly eyes watch our moves on the screen – to virtual workforce employees, social partners, and programmed AI therapists. With avatars, we have to heed the *disappearance of computers in society* in the quest for digital humanity (Simanowski 2019: 3) by criticizing mere data-driven media and their cultural analytics (Manovich 2020) as models of AI avatar aesthetics.

The avatar as a model of subjectivity has been described as a virtual proxy and representative of a real person (Little 1999; cf. Gunkel 2010). Others focus on a prosthetic avatar as a puppet or homunculus double (Apter 2008) of agency in a technical milieu, including cybertherapy (Gerner 2020).

The *avatar dream* (Fox Harrell and Lim 2017), when integrated with the two other culturally shared visions of future media of technological dreams using the computer and algorithms – the *smart dream* of ubiquitous quantitative total availability (Emrich / Roes 2011: 8–9) and the *AI dream* – becomes, in my view, the *smart, ubiquitous AI avatar dream*.

Fox Harrell and Lim characterized the avatar dream in a twofold way: technical and experiential. Computationally created surrogates engage us using text descriptions in games or social media through virtual visual representations in virtual reality environments. The experiential dimension enables virtual surrogate selves to engage in immersive experiences beyond orthodox physical encounters (Fox Harrell / Lim 2017: 52).

In this conception of the avatar dream, people utilize the computer as a chimera-creating tool to hack into their self-image. The avatar dream machine produces surrogates and

IMAGE 1.
CLIPPY: IN OFFICE VERSIONS 97 AND 2000, IF A USER TYPED "DEAR" AT THE BEGINNING OF A DOCUMENT, CLIPPY WOULD APPEAR IN THE BOTTOM RIGHT CORNER OF THE SCREEN WITH A TEXT BUBBLE THAT READ, "IT LOOKS LIKE YOU'RE WRITING A LETTER. WOULD YOU LIKE HELP?"

SEE FELDMAN (2016) ON THE DESIGN OF A VIRTUAL ASSISTANT OPTIMIZED FOR FIRST USE OF A FUNCTION THAT WAS THE FORERUNNER OF AI ASSISTANTS SUCH AS ALEXA OR SIRI
HTTPS://MONEY.CNN.COM/GALLERIES/2009/TECHNOLOGY/0910/GALLERY.MICROSOFT_WINDOWS_GAFFES/2.HTML

1 This research is financed by Portuguese national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., within the scope of the Transitional Standard – DL57/2016/CP CT[12343/2018], in the scientific field of History and Philosophy of Science and Technology, Project: *Hacking Humans. Dramaturgies and Technologies of Becoming Other.* Position: 2404.

transforms each imaginary experiential frame virtually by allowing us to play whoever we want to be. Thus, avatars might become part of a virtual identity. Fox Harrell and Lim (2017: 60) further argue that the avatar dream needs be reimagined beyond mere techno-phenomenological otherness to take into account society, including biases and stereotypes and constraints to the achievement of social identity as experienced in physical and self-imaginations in virtual worlds. Suppose we do not heed the historical, social, and cultural constraints of human-made artifacts. In that case, we might not avoid system-embedded and user-embedded "box effects" – "the experiences of people that emerge from the failure of classification system (...) stereotypes, social biases, stigmas, discrimination, prejudice, racism, and sexism" (Fox Harrell / Lim 2017: 54) – that would render the avatar dream impossible. While the avatar dream is specifically related to personal self-image, the AI avatar dream goes beyond a mere computational representation of users. Beyond mere mechanical "learning" or "intelligence," the *AI avatar dream* proposes AI avatars as creative machines (Rauterberg 2021). AI avatar dreams create other AI personas and professional specialists (e.g., therapists or consultants), such as embodied cognitive models, and a dream of another vision of humanity. in which the avatar is even part of a future self-generating art. This AI avatar dream goes in the direction of another artificial, virtual, or synthetic human: a form of self-superation, self-determination, and religious eternal self-salvation, with posthuman capacities, embodiment possibilities, and new modes of an extended human experience. Thus, virtual AI humans generate the future media ability to communicate in natural human language, to "learn," "remember," and "own" a personality as well as making decisions by taking actions with their bodies via a set of sensory systems. The idea of virtual humans includes the ability to detect sensations, appraise sensation triggers, and respond to them.

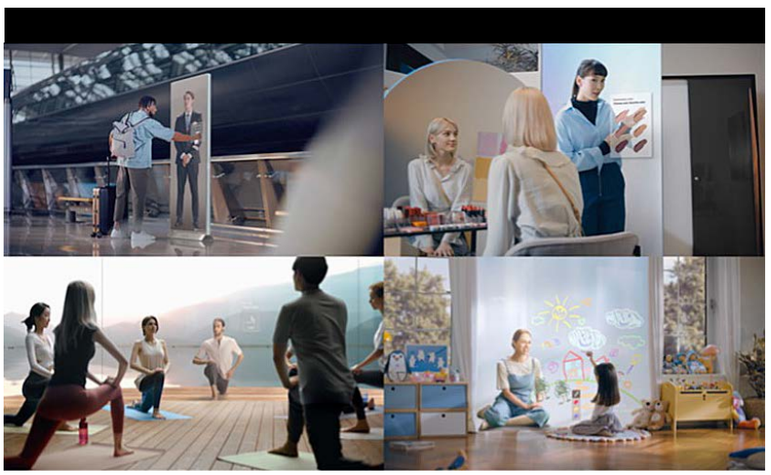


IMAGE 2.
SCREENSHOT AI AVATAR WORKFORCE PRESS RELEASE 2021 WITH "NEON'S VISION OF THE FUTURE OF WORK AND EVERYDAY LIFE" IN FOUR SITUATIONS, CLOCKWISE FROM TOP LEFT: AI AVATAR SERVICE ASSISTANT AT THE AIRPORT, AI AVATAR SALES AND SERVICE ASSISTANT, AI AVATAR TEACHER, AI AVATAR FITNESS TRAINER.
HTTPS://NEON.LIFE/NEWS/CES-2021-PRESS-RELEASE

AI avatars act as a digital workforce and function as employees, such as the virtual worker *AMELIA.* AMELIA is a job-based, human-equivalent digital employee that is customizable for each service business, such as for Customer Care, IT, and HR services or multi-lingual digital banking. In the case of the Sterling National Bank, AMELIA – renamed "Skye" – provides human-like communication and collaboration with the bank's contact center agents and in the case of the Netherlands-based IT Service "Centric Burgerzaken" AMELIA is used to provide conversational AI, available 24/7, for digital public services for local government organizations.

AI avatars as workers are meant to enhance employee performance culture in VR scenarios within performance analytics. This development includes companies such as

Talespin's co-pilot virtual human training technology or customer assistance and UneeQ's Digital Humans, defined as AI-powered, lifelike virtual beings. UneeQ's Digital Humans are AI avatar workers that mimic human facial expressions, tone of voice, and body language in multimodal embodied forms of communication. These features are more important than mere language-based verbal communication for customer service. The abilities of virtual humans include showing emotion and different moods, making plans, and achieving goals, ideally set by some "internal" motivation. Internal motivation in the sense of Artificial General Intelligence (AGI) could even be an internal avatar model with an external avatar body – with the AGI ability, in addition to reasoning and problem-solving, to mimic the capacity of imagination and creativity. Burden and Savin-Baden (2019: 13) have developed a matrix to analyze virtual humans' traits on different spectra between self-aware and not self-aware, embodied and disembodied, humanoid and non-humanoid, natural-language and command-driven, autonomous and controlled, emotional and unemotional, personality-driven and impersonal, reasoning and unreasoning, learning and "unlearning" (cf. the EmoCOG architecture (Lin et al. 2011) or the OpenCOG architecture (Goertzel et al. 2014)

in which attention-related "forgetting" and memory resource management is put forward (Burden / Savin-Baden: 125)), and finally, imaginative and unimaginative. In a posthuman avatar case scenario, such as in Soul Machine's 4th and 5th level of AI avatars, the aims are not only spatial context, as-if imagination, and as-if intentionality, but also creative machine behaviors based on "learned experience" and "agency" for making discoveries and setting new intentions, plans, and goals. Moreover, AI avatars in the future AI dream world gain the ability to train themselves through interaction with humans and non-human systems. Finally, self-awareness and contextual understanding would emerge in independent digital, artificial persons with a strong semantic or contextual understanding of the AI avatar self's actions to create non-linear storytelling. Nevertheless, AI artifacts that move, speak, reason, and show radical mimetism will inevitably face issues of animism.

2. THE AIS HAVE IT? ON AI AVATARS

2.1 "HIGH FIDELITY" AVATARS: COUNTERFEIT OF HUMAN GAZE OR THE WRONG KIND OF ANTHROPOMORPHISM?

AI artifacts are AI systems that humans create for the purpose of radical mimesis: AI systems mimic actors who grant social faciality to machines in a way that seems human to observers. The AI avatar machine evokes movements of gaze and interest, as well as curiosity, and has to be critically assessed when reflecting on the topic of human or machine creativity. Coeckelbergh (2021) argues for a critical posthumanist point of view towards the anthropomorphism in technical objects that interact with humans. Should we then reject normative anti-anthropomorphism as nonsensical in social robotics and AI avatars? And still: we have to ask how we handle AI avatars not only as extensions of the self but as AI technology for human exploitation and data extraction (Crawford 2021), the cost of which must still be counted in its material, energetic and ecological aspects. Some may make a strong stance against AI avatars as simulation machines of not only intelligence but – foremost – human attributes such as creativity, autonomy, affectivity, and for being "artifactors," AI artifact systems that mimic human (like) actors, calling them a "counterfeit" (Pasquale 2020) of humanity. Therefore, the task of clearly separating AI systems from AI actors that mimic humans through anthropomorphic design stances might seem a good idea for a policy option (Cf. EU 2021) that calls for a renewal of Asimov's Three Laws of Robotics (see

Pasquale 2020: 3–19). These new potential rules for AI would go beyond avoiding maleficence by impeding human substitution, human manipulation/counterfeiting, an AI arms race, and non-identification of artificial systems.

2.2 FROM AI CHILD AVATAR TO PLAYING GENERAL ARTIFICIAL INTELLIGENCE WITH A TOY CHILD MODEL: ON SOUL MACHINES'S AUTONOMOUS ARTIFACTOR ANIMATION

The AI research of the company *Soul Machines* "started with a baby", called "Baby X" (Soul Machines 2021). According to IBM *(Soul Machines.* IBM. n.d.) and its Watson assistant integrated into Soul Machines, the aim and business challenges are to build on the paradoxical goal of *empathic AI* that has been staged as evolutionary human progress at the World Economic Forum in 2019 (Mantas 2019). The AI avatar model of Baby X plays interactively with the world around it, pragmatically making discoveries by manipulating things in the way we do. Animation stands at the center of Soul Machines's business, which is inspired by the following questions: "What is the essence of animation? What if a character could autonomously animate itself and you could interact with it? How do you bring a digital character to life?" Baby X interacts with its surroundings by playing as if it were a child that learns, evolves, or "grows" its information base by testing the results of the games it plays; but does it actually rehearse and acquire reality? Soul Machines poses challenges of "problems to solve" that lie at the core of AI avatars as artifactors:

> ***How would we create biologically inspired artificial intelligence? And, build a digital consciousness to create affective computing that interprets and simulates human emotion, engaging autonomously?*** *(Soul Machines 2020)*

Soul Machines's AI avatar initially reminds us of an AI Tamagotchi (virtual pet), referring to the emotional annoyance of having to feed and care for the digital toy in how it is presented. However, the company aims to "make machines and AI as lifelike as possible," envisioning "humanlike AI that has flexible intelligence and a dynamic interface that can relate to people": human-AI relations seem to change in the age of machine learning, having a clear roadmap of how to achieve the highest levels of "autonomous animation."

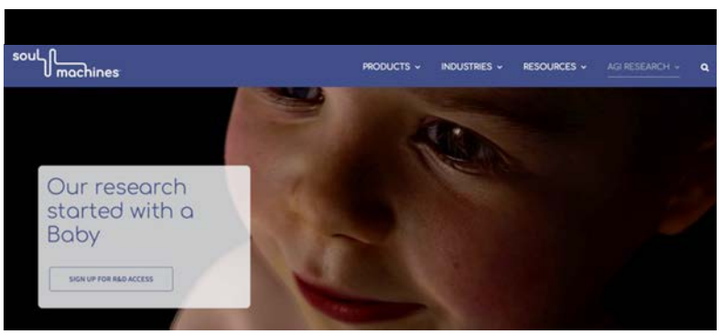


IMAGE 3.
SCREENSHOT OF BABY X ON THE INTERNET PAGE SOUL MACHINES.
HTTPS://WWW.SOULMACHINES.COM/RESOURCES/RESEARCH/BABY-X/

Soul Machines's white paper (2020) distinguishes six stages of autonomous animation, in which level 0 and level 1 are dedicated to actually existent simulated, actor-driven, pre-recorded video or motion capture in which motion-capture cameras function as enabling technology for "possible solutions" in movie and games characters. On this level, avatars are supposed to be used as masks and puppets and heed movement notations of kinetic digital renderings in capturing performance art inside a motion capture imaginary (Karreman 2017) in creative industries, games, films, and contemporary dance. Avatar masks refer to a performer as a puppet master: the avatar mask can be seen as an initial new identity or as a mere puppet in an uncanny zone in between *something* and *nothing.* For Soul Machines (2020), on Level 1, basic pre-authored animation that is

still actor-driven delivers pre-recorded movement based on simple triggers. The corresponding enabling technology would include the FAQ_text-driven conversational database and pre-recorded voice content responses to create *digital puppets*. Levels 2 and 3 of "autonomous animation" would already use Natural Language Processing and "Dynamic Synthesized Human Behaviors" (Soul Machines 2020: 8), a "learning" capacity based as a solution on the Deep Fake level (level 2) or on level 3 with "[f]ull humanlike emotional responsiveness in facial animation including a conversational driven personality," including on the voice level.

With the selling of the idea of the higher-level autonomous AI avatar as part of the AI avatar dream machine industry, we should ask: Does an AI avatar assimilate otherness by radical mimetics to be used in game design and performative conventions for creating *pervasive performances* (Peréz 2016: 16) between acting and engagement? Do AI avatars follow the metaphorical model of Turing's child machine to create and provide "education" (Turing 2004: 460) to an AI child model such as Baby X or are they an AI avatar toy for playing around with artificial general intelligence, such as the AI toy avatar model Kanzi (Negarestani, 2018)?

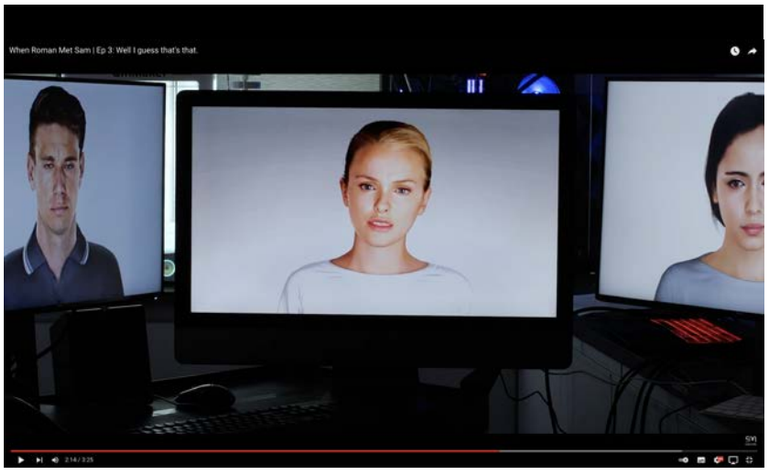


IMAGE 4.

SCREENSHOT OF A TRIALOGUE OF THREE MACHINE COMMUNICATION AVATARS FROM SOUL MACHINES

HTTPS://WWW.YOUTUBE.COM/WATCH?V=4MCDPFKYLTS

2.3 CODEC AVATARS (CA): THE QUEST TO PASS FACEBOOK'S "EGO AND THE MOTHER TEST"

As proximity and face-to-face encounters determine social relationships, the technological roadmap of VR and AR by Facebook's Oculus Rift is heading towards overcoming distance and material barriers, as put forward by Tanaka, Nakanishi and Ishiguro (2014), who had shown that physical robot conferencing was superior to mere avatar chat. By virtual immersion of Codec Avatars, or enhanced Modular Codec Avatars (Chu et al. 2020) – which improve the robustness and expressiveness of traditional Codec Avatars – with holograms and VR/AR, Facebook aims at recreating and mimicking a sense of (artificial) VR telepresence, which provides remote and immersive telecommunication through VR headsets. The training phase of the VR telepresence system in the first stage is done by capturing facial expressions of a user with a multi-view camera dome and a VR headset for face modeling. In the final phase a personalized face animation model is derived using these correspondences, while the real-time photo-realistic avatar is driven from the VR headset cameras. This social teleportation is able to share eye gaze and expressive faciality that would be almost indistinguishable from the real-life presence of a person or object, even enhancing the spectrum of senses using a new artificial digital-media sense that could be called the digital immersive sense of foreshadowing proximity to an object or person.

However, Mark Zuckerberg admits that a) not all material experience while "connecting people" will and can be virtualized and b) algorithmically modeling the materiality of touch and haptics is not easily done. Photorealistic avatar models for "high-fidelity social interaction" of the users' faces render avatars with a "Deep Appearance Model for Face rendering" (Lombardi et al. 2018) "using non-linear, photorealistic full-face models of geometry and texture" (Richard et al. 2020: 1), overcoming the shortcomings of mere geometric

approaches due to the "non-linearities in texture-based tongue motions and lip articulation" (ibid.). The difficulties are related to dark untracked geometry inside the cavities of the mouth that must be emulated with a synthetic texture of the mouth. Facebook came up with the idea that the avatars should not only be acceptable but also that they should not create uncanny valley effects. When setting up an avatar, a second "Turing Test" of social presence for the Facebook Codec Avatar would be if the avatar is acceptable for yourself and "your mother," (Tech@Facebook 2019). Thus, Facebook focuses on their codec avatars as an avatar dream of a high-fidelity replica of the gaze. Implicit in Facebook's High Fidelity Avatar (Schwarz et al. 2020: 91) is the concept of high fidelity of Skarbez et al. (2017), who differentiate between a) *physical morphological fidelity of looks* inside the operational environment, b) *functional action fidelity* of faciality of eye gaze or operational performance of the gaze in realistic movements and agency, and c) active *perceptive fidelity*. However, I question if this hyperbolic-realistic "high fidelity" actually encompasses passive perception. What gives the face-to-face encounters a feeling of being together in the same space and experiencing a common "we"? Is it the idea of *being looked at by the other,* who does not perform exactly as I expect?

IMAGE 5.
SCREENSHOT FACEBOOK CODEC AVATAR DEMONSTRATION VR TRAILERS AND CLIPS, YOUTUBE (JULY 2, 2020) "FACEBOOK'S PROTOTYPE PHOTOREAL AVATARS NOW HAVE REALISTIC EYES"
HTTPS://WWW.YOUTUBE.COM/WATCH?V=ETAMZMYKSG0

2.4 ERGOTIC COMMON-SENSE GESTURE-BASED AI AVATARS: TWENTY BILLION NEURON'S GESTURE SURROGATE AVATAR ASSISTANT MILLIE AND ITS AI APP FITNESS ALLY

The German/Canadian AI company Twenty Billion Neurons (TwentyBN), based in Berlin and Toronto, teaches machines to perceive like humans and developed the avatar "Millie" using situated a model of visual AI common sense via end-to-end learning on video clips (Twenty Billion Neurons 2020): the "Supermodel." This AI model is a Python-based, deep learning gesture-recognition model based on large-scale crowd-acting operations and has collected millions of short video clips that require no depth information, as the model is entirely trained on 2D video data. This gesture recognition model internalizes a visual "common sense" of the world by identifying a wide range of fundamental human-object interactions and human body motions.

The TwentyBN avatar is based on the AI SuperModel of computer analysis of collected crowd-acting, in which people in the recorded video snippets perform common-sense hand control gestures via different data sets. These include, for example, Jester V1, in which 147 crowd workers performed 27 pre-defined hand gestures in front of a laptop camera or webcam (148,092 short clips of videos with different backgrounds, 3-sec length) and the "20BN-something-something V2 Dataset" inside the probability-guided labels to detect common-sense actions by AI algorithms of machine vision.

The neural network that offers the data feed to the avatar gesture simulation consists of short videos of mostly ergotic gestures sorted into common-sense pragmatic action classes (caption templates). These action classes are of a general "something [picking, moving, putting...] something" (Goyal et al. 2017: 5848) structure: AI avatars are based on common-sense gesture training sets fed into AI algorithms. These AI vision algorithms use artificial neural nets and deep fake technology. The avatar Millie is introduced in the first place as an interactive AI avatar in-store shopping assistant, and its corresponding app "Fitness Ally," a virtual avatar fitness trainer, is used to guide the user through a series of workouts and to present them with recorded and interactive training data for fitness improvement.

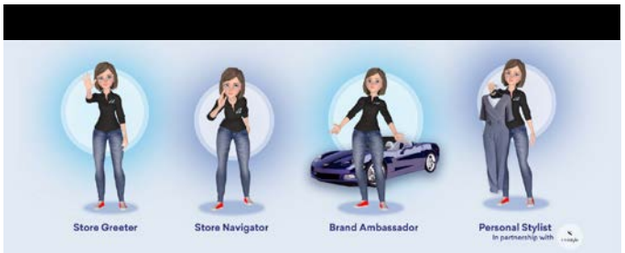


IMAGE 6.

SCREENSHOT TWENTYBN MILLIE'S FUNCTIONAL APPLICATION AS "DIGITAL IN-STORE EXPERT."

SEE: HTTPS://MEDIUM.COM/TWENTYBN/YOUR-DIGITAL-IN-STORE-EXPERT-FOR-EVERYTHING-D0865B82E27A.

THE SLOGAN OF THE COMPANY IS "BREATHING LIFE INTO VIRTUAL BEINGS/ OUR HUMAN-CENTRIC AI TECHNOLOGY BRINGS SEEING AND SOCIABLE DIGITAL ASSISTANTS TO LIFE." THE DATABASE IN 2017 CONSISTED OF MORE THAN 100,000 VIDEOS ACROSS 174 CLASSES; BY 2021 THE DATABASE HAD GROWN MORE THAN TENFOLD.

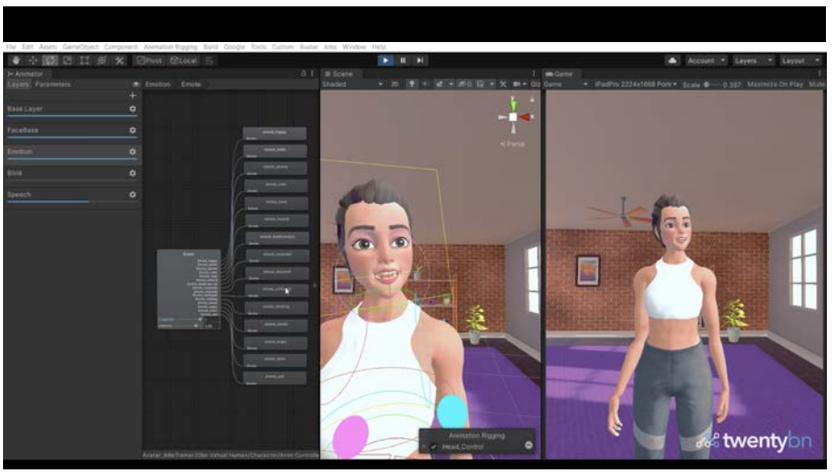
IMAGE 7.

SMARTPHONE APP FITNESS ALLY "REAL-TIME INSTRUCTION AND MOTIVATION."

HTTPS://FITNESSALLYAPP.COM

IMAGE 8.

SCREENSHOT OF THE AI FITNESS TRAINER ANIMATION FACE STACK DESIGN MADE FOR ALLY FITNESS BY FERGUI.



(HTTPS://VIMEO.COM/543742486). THE TECHNICAL PROGRAMMING CHOSE KINETICS SKELETON BENCHMARKS FROM "PAPERSWITHCODE" (HTTPS://PAPERSWITHCODE.COM/SOTA/SKELETON-BASED-ACTION-RECOGNITION-ON-KINETICS). THE POSE KEYPOINTS ARE DERIVED FROM THE "MMSKELETON" TOOLSET ON GITHUB (HTTPS://GITHUB.COM/OPEN-MMLAB/MMSKELETON). IN GENERAL, THE FITNESS APP IS BASED ON THE ST-GCN MODEL THAT STANDS FOR "SPATIAL- TEMPORAL GRAPH CONVOLUTIONAL NETWORKS" (YAN ET AL. 2018), IN WHICH GRAPH CONVOLUTION IS TRANSPOSED TO SKELETON-BASED ACTION RECOGNITION AND ADDED THE MS-G3D MODEL TO CAPTURE COMPLEX SPATIAL-TEMPORAL FEATURES AS METHOD FOR IMPROVING SKELETON-BASED ACTION RECOGNITION BY MULTI-SCALE GRAPH CONVOLUTIONS AND A UNIFIED SPATIAL-TEMPORAL GRAPH CONVOLUTIONAL OPERATOR NAMED G3D (LIU ET AL. 2020), FOR ITS IMPLEMENTATION (TBN 2020, DEC 14: HTTPS://MEDIUM.COM/TWENTYBN/PUTTING-THE-SKELETON-BACK-IN-THE-CLOSET-1E57A677C865). TWENTYBN IN 2020 ALSO LAUNCHED PART OF ITS TECHNOLOGY AS AN OPEN-SOURCE PLATFORM, SENSE, "A REAL-TIME ACTION RECOGNITION SYSTEM," OPEN-SOURCE INFERENCE ENGINE FOR NEURAL NETWORK ARCHITECTURES THAT TAKES AN RGB VIDEO STREAM AS INPUT AND TRANSFORMS IT INTO A CORRESPONDING STREAM OF LABELS IN REAL TIME. SENSE INCLUDES DAY-TO-DAY HUMAN ACTIONS (PICKING UP OBJECTS, DRINKING WATER, FIXING YOUR HAIR, ETC.), HAND GESTURES, AND FITNESS EXERCISES, AMONG OTHERS: HTTPS://GITHUB.COM/TWENTYBN/SENSE

Millie was created with Deep Learning training of initially one thousand actions; now the database of common-sense gestures and visual common-sense actions to feed this action recognition pool is far over a million. It contains an object detector, an action/motion detector, a dialogue system, and a rule-engine for recognition and reaction to humans, which is used for Millie and was developed with the following aims (Kahn 2018): a) the immediate aim to build an interactive social sales assistant, gesture control systems for the car industry, and smart home devices b) the TBN long-term aim to build full digital avatars with a designed personality to interact with people in various settings, including full digital social companion, exploring avatars that could even "help" teach children in schools or instruct adults in skills such as yoga or cooking, or an artificial officer such as New Zealand's police artificial person "Ella," developed by Soul Machines. Whether AI avatars will attain the depth and personality to serve not merely as trainers but actually as pedagogic teachers is an issue remains to be seen.

REFERENCES

Apter, E. (2008), Technics of the Subject: The Avatar-Drive, in: Postmodern Culture, 18(2). doi:10.1353/pmc.0.0021

Soul Machines (2021, March 8). Baby X. https://www.soulmachines.com/resources/research/baby-x/ (04.02.2022).

Brownlee, M. (2020, September 16) Youtube: Talking Tech and Holograms with Mark Zuckerberg! https://www.youtube.com/watch?time_continue=72&v=eAagtcAup0o&feature=emb_title

Burden, D., / Savin-Baden, M. (2019), Virtual Humans: today and tomorrow. CRC PRESS.

Chu, H. / Ma, S. / De la Torre, F. / Fidler, S. / Sheikh, Y. (2020), Expressive telepresence via modular codec avatars. Computer Vision – ECCV 2020, p. 330-345. doi:10.1007/978-3-030-58610-2_20

Crawford, K. (2021) Atlas of AI. New Haven.

Emrich, H, / Roes, M. (2011), Engel und Avatar. Berlin.

EU (2021), Proposal for a Regulation on a European approach for Artificial Intelligence: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence (26.4.2021)

Feldman, B. (2016), Clippy Didn't Just Annoy You — He Changed the World, New York Magazine 31.10.2016, https://nymag.com/vindicated/2016/10/clippy-didnt-just-annoy-you-he-changed-the-world.html (12.12.2021)

Fox Harrell, D./ Lim, C. (2017), Reimagining the Avatar Dream: Modeling Social Identity in Digital Media, in: Communications of the ACM, July 2017, 60 (7), p. 50-61.

Gerner, A. (2020), Hacking into Cybertherapy: Considering a Gesture-enhanced Therapy with Avatars (g+TA). Kairos. Journal of Philosophy and Science 23, p. 32-87.

Goyal, R. et al. (2017), The 'Something Something' Video Database for Learning and Evaluating VISUAL common sense. 2017 IEEE International Conference on Computer Vision (ICCV), p. 5842-5850. doi:10.1109/iccv.2017.622

Gunkel, D. (2010), The Real Problem: Avatars, Metaphysics, and Online Social Interaction, New Media & Society 12(1), p. 127-141.

Goertzel, B. / Hanson, D. / Yu, G. (2014), Toward a Robust Software Architecture for Generally Intelligent Humanoid Robotics. Proceeding Computer Science, 41, p. 158–163.

Yan, S. / Xiong, Y. / Lin, D. (2018), Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. arXive: 180107455v2 [cs.CV] 25. January 2018

Kahn, J. (2018), "Meet Millie the Avatar," Bloomberg News: https://www.bloomberg.com/news/articles/2018-12-15/meet-millie-the-avatar-she-d-like-to-sell-you-a-pair-of-sunglasses (14.07.2021)

Lin, J. / Spraragen, M. / Blythe, J. / Zyda, M. (2011), EmoCog: Computational Integration of Emotion and Cognitive Architecture. In Proceedings of the Twenty-Fourth.FLAIRS Conference. Palo Alto, CA: AAAI.

Liu, Z. / Zhang, H. / Chen, Z. / Wang, Z. / Ouyang, W. (2020), Disentangling and Unifying Graph Convolutions for Skeleton-Based Action Recognition. arXive:2003.14111v2 [cs.CV] 19 May 2020,: https://arxiv.org/pdf/2003.14111.pdf (20.07.2021)

Little, G. (Fall 1999), 'A Manifesto for Avatars,' Intertexts 3(2): http://www.gregorylittle.org/avatars/text.html (20.04 2021)

Mantas, J. (2019), Empathic AI Could be the Next Stage in Human Evolution - If We Get it Right. https://www.weforum.org/agenda/2019/07/empathic-ai-could-be-the-next-stage-in-human-evolution-if-we-get-it-right/ (02.09.2020)

Manovich, L. (2020), Cultural Analytics. Cambridge, Mass.

Negarestani, R. (2018), Intelligence and Spirit. New York.

Pasquale, F. (2020), New Laws of Robotics. Defending Human Expertise in the Age of AI. Cambridge Mass..

Rauterberg, H. (2021), Die Kunst der Zukunft. Über den Traum von der kreativen Maschine, Berlin.

Richard, A. / Lea, C. / Ma, S. / Gall, J. / De la Torre, F. / Sheikh, Y. (2020), Audio- and Gaze-driven Facial Animation of Codec Avatars: https://research.fb.com/videos/audio-and-gaze-driven-facial-animation-of-codec-avatars/ 2020/ (03.02.2022)

Schwartz, G. / Wei, S.-E. / Wang, T.-L. / Lombardi, S. / Simon, T. / Saragih, J. / Sheikh, Y. (2020), The Eyes Have It: An Integrated Eye and Face Model for Photorealistic Facial Animation. ACM Trans. Graph. 30, 4, Article 91.

Simanowski, R. (2019), Stumme Medien. Vom Verschwinden der Computer in Bildung und Gesellschaft, Berlin

Skabez, R. / Brooks, F. / Whitton, M. (2017), A Survey of Presence and Related Concepts, ACM Computing Survey, 50, 6, Article 96, p. 1-39.

SoulmachinesTM (2020), Delivering on the Promise of AI. How Digital People Rise Above Other Technologies. The Questions Every Decision Maker, Investor, and Innovator Should Ask, White paper, p. 1-12 https://www.soulmachines.com/wp-content/uploads/2020/09/DeliveringOnThePromise.pdf?utm_medium=email&_hsmi=107119114&_hsenc=p2ANqtz--qePpUrpMWU-NRc-1MaZa-u-i0-8NcC2fgomy8TrMthr-Y10iK_blZ3nMkMizv8O9eag6JWsi33s-1R1R6z6G-7AfbLA6ag&utm_content=107119114&utm_source=hs_automation . Facebook: https://tech.fb.com/codec-avatars-facebook-reality-labs/ ; IBM Soul Machines (n.d.): https://www.ibm.com/case-studies/soul-machines-hybrid-cloud-ai-chatbot (10.02.2022).

Tanaka, K. / Nakanishi, H. / Ishiguro, H. (2014), Comparing Video, Avatar, and Robot Mediated Communication: Pros and Cons of Embodiment. Communications in Computer and Information Science, p. 96–110. Doi:10.1007/978-3-662-44651-5_9

Turing, A. (2004), Computing Machinery and Intelligence (1950), in: Copeland, J. (Ed.), The Essential Turing. Oxford, p. 441-464.

Twenty Billion Neurons (2020), Towards Situated Visual AI via End-to-End Learning on Video Clips, https://medium.com/twentybn/towards-situated-visual-ai-via-end-to-end-learning-on-video-clips-2832bd9d519f (10.02.2022)