Melusina Press
Humanities

Serie #1 | The Ends of the Humanities
Edited by Johannes Pause, Christoph Purschke,
Isabell E. Baumann, Till Dembeck & Georg Mein

The Humanities and the Rise of AI

Implications of Cultural and Societal Engineering

Johannes Pause, Christoph Purschke,
Isabell Eva Baumann, Till Dembeck & Georg Mein

The Inner Voice of the Artificial Self

On Disembodied Subjects in AI Cinema

Johannes Pause

The Ends of the Humanities - 3
Edited by Johannes Pause, Christoph Purschke, Isabell Baumann, Till Dembeck and Georg Mein

Melusina Press, 2025

Online herausgegeben durch die Universität Luxemburg - **Melusina Press**, 2025
11, Porte des Sciences
L-4366 Esch-sur-Alzette
https://www.melusinapress.lu

Management: Niels-Oliver Walkowski, Johannes Pause
Copyediting: Carolyn Knaup, Niels-Oliver Walkowski
Cover and Layout: Valentin Henning, Erik Seitz

The digital version of this publication is freely available at
https://www.melusinapress.lu.
The PDF and the master copy are generated with the help of princeXML.
Print and Distribution is carried out by BoD - Books on Demand, Norderstedt

Bibliographic information of the National Library of Luxembourg: The National Library of Luxembourg lists this publication in the Luxembourg National Bibliography; detailed bibliographic data are available on the Internet at bnl.public.lu.

ISBN (Online)	978-2-919815-79-1
ISBN (PDF)	978-2-919815-80-7
ISBN (Print)	978-2-919815-81-4
DOI (Publication)	10.26298/1981-5791-tivo
DOI (Volume)	10.26298/1981-5791
DOI (Series)	10.26298/2716-7550
ISSN (Online)	2716-7550

The Inner Voice of the Artificial Self

On Disembodied Subjects in AI Cinema

In cinema and television, and especially in Hollywood productions, artificial intelligences usually demonstrate their abilities and peculiarities through their language: they communicate with humans and other computers, and in doing so they generally make use of a human-sounding voice. This voice becomes the central expression of their specific subjectivity, which often resembles that of humans, but can also prove to be radically different in some respects. The impression that we are dealing with an autonomous personality arises in particular when the voice of the computer appears as an 'inner' or 'internal' voice that influences or imitates the intimate soliloquies of humans.

The following essay explores the question of how this use of artificial inner voices has changed in the course of film history, and which concepts of artificial and human intelligence, subjectivity, and (gendered) identity are associated with manifestations of the voice. In particular, I am interested in describing the threats to humanity which, according to most fiction films, are posed by artificial intelligences. Representations of the voice have undergone a profound transformation in recent decades. The cinema of the 1960s to 1980s imagined AI in the context of totalitarian regimes, with the artificial intelligence developing the psychotic personality of a panoptical, authoritarian, and paranoid ruler. In more recent films, however, we find an amalgamation of artificial and human subjectivity that leads to technology taking over the human internal voice and thus to a radical alienation of humans from themselves. The voice, and especially the internal voice, therefore becomes not only a site for negotiating the relationship between human and non-human agents, but also an instrument of subtle manipulation and deformation of humanness itself.

1. Voices without bodies

Many 'Western' movies that explicitly deal with artificial intelligence seem less interested in questions of *intelligence* in the narrow sense than in the possibility of an artificial *self*. Their central topics are the psychology of such artificial selves, the relationships they might develop to human selves, and the power relations resulting from these relationships. Probably for this reason, 'AI movies' typically connect the concept of the artificial self to artificial intelligences that possess a human body, such as robots, replicants and cyborgs. The paradigmatic example for this type of cinema is, of course, *Blade Runner* (USA 1982, Ridley Scott), but we could mention many others here, including the successful television series *Westworld* (USA 2016-2022, Lisa Joy and Jonathan Nolan) and computer games such as *Detroit: Become Human* (F/USA 2018, David Cage). Here, the artificial selves not only look like real people, but they also act and feel like them: they have intimate memories, make personal experiences, make individual decisions and are afraid of death. By developing empathy with them, the spectators – or players – are en-

abled to question the boundaries normally drawn between the human and the non-human. The replicant Roy Batty in *Blade Runner*, for example, famously gives expression to his humanity by saving the life of Rick Deckard (Harrison Ford) shortly before the end of his own – programmed – lifetime. Since he is played by well-known human actor Rutger Hauer, the typical tough guy with good heart, it is easy for the spectators to acknowledge Batty's humanity.

Movies like *Blade Runner* thus ask us to accept that cyborgs are indeed like humans, and that the exclusion mechanisms they suffer from must be regarded as serious as those from which human minorities often suffer. Things get a lot more complicated, however, when it comes to incorporeal AIs. What options do they have to develop their own subjectivity, and what about their rights and their relationship to human agents? Even in computer sciences, to quote a recent research paper, "body ownership" is seen as one of the "prerequisites for an artificial self" (Hafner et al. 2020). In cinema, however, bodyless artificial intelligences who solely exist in the non-space of the digital are sometimes able to develop their own personality. In these cases of origin "discarnality" (McLuhan / Powers 1989), a human or human-like voice usually functions as a substitute for the missing human body. The personalities of the artificial characters therefore emerge solely in the form of verbal communication, with human actors usually voicing the role of the digital character. In the context of cinema, having a voice thus appears as the indispensable basic element of any self, artificial or not, and, at the same time, every voice is inevitably connected to humanness, because it evokes an imagination of some sort of human body.

Parallel to cinema's fascination with artificial voices, a strong fascination with the human voice has developed in computer sciences. For example, *ChatGPT 4* now has an "advanced voice mode" that allows communication in a variety of languages and voices, but still evokes a certain degree of creepiness (Rogers 2024). According to media scientist Liz Faber, voice-based human-computer interaction has in fact become "the holy grail of computer engineers" (2020: 54) in recent years. From a humanities perspective, it seems that artificial intelligence research has opened yet another chapter in the long "phonocentric" tradition of Western philosophy, which was famously criticized by Jacques Derrida (1976). According to Derrida, the voice marks the origin of the human self in this intellectual tradition; it seems to be "immediately connected to, and an expression of, the breath and life of a subject; hence it is connected in an unmediated way to a *will*" (Cisney 2014: 91). However, since artificial voices are necessarily part of a "secondary orality" in the sense of Walter Ong (1982), an orality that is based on symbolic representation, no artificial intelligence can be phonocentric in this emphatic sense. Artificial Intelligences might be considerably good at imitating existing voices, but they don't *naturally have* a voice of their own. They are simply programmed to translate coded representations of voices into sound. Like the voices of cinema, their voices are therefore representations of sound (cf. Altman 1992) that are usually made from human voices. As a team of audio forensic experts was able to prove in 2013, the first voice of Apple's *Siri* was, for example, based on the voice of the human actress Susan Bennett, who didn't even know that her voice had been harvested by the megacorporation (Faber 2020: 3). And Scarlet Jo-

hansson recently sued OpenAI for the use of a voice "eerily similar" to hers in the latest version of *ChatGPT* (Milmo 2024).

In the fictions of cinema, however, the voice appears to be a crucial element of any artificial self. Following the phonocentric tradition, it is used more and more over the course of time to suggest that we are dealing not only with an intelligent being but also with a subjectivity that has its own identity. According to Faber, the first filmic works that introduced speaking artificial intelligences in the 1960s were heavily gendered, suggesting a human body and thus a natural origin of the artificial voice. Nevertheless, a striking difference existed between male and female language outputs: Female artificial intelligences such as the computer in the 'original' *Star Trek* series (USA 1966-69, Gene Roddenberry) possess extensive knowledge that they can clearly express in standardized question-answer situations, but they don't have much of a self. Instead, the Star Trek computer masters language as a kind of rule system, but it does not seem to develop or learn from observation. Even if this type of fictional computer may seem sympathetic and sometimes even humorous, it does not have an independent point of view, feelings, secret intentions, or questions of its own. As Faber extensively shows, representing a self remains the privileged domain of artificial intelligence with male voices for quite some time. The prototype of such a male artificial self is certainly the computer HAL from the middle section of Stanley Kubrick's *2001: A Space Odyssey* (USA 1968), the onboard computer of an American spacecraft on a secret mission to a Jupiter moon.

2. The psychotic father: HAL and Colossus

HAL is clearly introduced as an autonomous learning system and can thus be seen as an early imagination of a 'subsymbolic' artificial intelligence. HAL, proceeding from his reading of the situation, starts to act against the human crew members of his spaceship. Convinced that the astronauts are endangering the secret purpose of the mission, he tries to kill them all. The last survivor finally manages to shut HAL down, reverting him to his earlier, quasi-childlike stages of development. "Finally, as HAL dies, his voice slows and deepens in pitch, like a machine whose batteries are dying, and he sings 'Daisy Bell.' Importantly, this song was the first song ever sung by a speech synthesis program, in this case run on the IBM 704" (Faber 2020: 36). HAL had learned this nursery rhyme from his first programmer, who apparently was a kind of father figure for him, shortly after his 'birth'. As the scene suggests, HAL has grown to speak like a child through interactions with his 'parents'. The onboard computer thus turns out to be a hybrid, someone who was shaped in his development by a father figure and learned to speak through nursery rhymes, who in other words has become an oedipal subject following the model of man. At the same time, as the secret guardian of the mission's true objectives, which are known only to him, he is in the position of law itself, a panoptic entity and thus an expression of the 'great other' in the psychoanalytical sense.

This position of power is particularly illustrated by the two organs that distinguish HAL: his eyes and his voice. Although HAL has no body, he does have his famous red camera eyes that hardly miss anything, and he has a voice that can be heard everywhere.

Unlike the voices of the human actors, HAL's voice has no reverberation and no distance. According to Michel Chion (2001: 101f.), it is an "example of both an 'acousmêtre' (a character who exists as an invisible voice with no place and supposedly able to see all, know all and do all)," and of "an 'I-voice' (so named because it resonates in us as if it were our own)". What Chion calls the "I-voice" can be seen as the typical representation of the internal voice, which is also at the heart of Derrida's critique of phonocentrism. According to Derrida, the inner voice appears in the Western philosophical tradition as a guarantor of the presence of the self, if not of the soul, and is seen as its immediate and true expression for this reason. This supposed immediacy is at the center of Derrida's critique (2011: 74): "Hearing-oneself-speak is not the interiority of an inside closed in upon itself. It is the irreducible openness in the inside, the eye and the world in speech." The critique of phonocentrism is thus less concerned with a revaluation of writing over voice, as is often claimed, than with the integration of voice into the project of grammatology: "Despite its lure of immediacy, the voice, even when it is heard in the solitary mental state, is not originary but already an artifact, a variant in a chain of emissions involving duration and spacing." (Chow 2021: 114)

Movies dealing with disembodied artificial intelligences make use of the suggestive power of the 'I-voice' – and at the same time present it as a technical artefact and thus, in Derrida's sense, as something genuinely opposed to the notions of 'interiority' and 'originarity'. The AI thus becomes an incomplete, unstable subject that thinks and feels like a human and at the same time despairs of being manufactured. In fact, the effect of the artificial inner voice is potentially lethal: by forcefully making the spaceship his artificial body, HAL goes from being the always-helpful onboard computer to the murderer of the very people he is supposed to serve. It is precisely through the development of his inwardness that he becomes a madman. As characters, the artificial selves of science fiction cinema thus occupy a peculiar intermediate position: On the one hand, they appear as human subjects whose voices seem to be the direct expression of intentions, feelings, and individual experiences, of an inner life. On the other hand, they cannot 'really' take a human perspective because they have no body to give their voices a place, and because in this way the voice remains conscious as a mere representation of data, as one among many possible outputs. The AI can only achieve identity by forcefully declaring the world to be a function of its own subjectivity. Many AI movies are therefore about the potentially dangerous anthropocentric misunderstandings that the computer's voice output evokes.

In early AI cinema, this inner conflict usually leads to a narcissistic deformation of the bodyless artificial self. In *2001: A Space Odyssey*, the panoptic AI becomes the prototype of a dictator because he has no place of his own, and therefore cannot distinguish between himself and the law. As a matter of fact, this is the short formula for a great number of science fiction movies of the 1960s, 70s and 80s. Again and again, the focus here is on artificial intelligences that try to become human, but develop into ruthless tyrants instead. They always orient themselves towards a male father-figure, and their incorporeality, their general lack of empathy and thus their inability to behave as human beings always lead them to become psychotic. The Cold War film *Colossus: The Forbin Project* (USA 1970, Joseph Sargent), for example, is about an intelligent doomsday machine that

takes over the reign of mankind. However, the super-computer Colossus must first learn to understand its new role, going through all stages of human development. Initially, Colossus appears as a somewhat petulant child, taking its cue from its father-like chief programmer, with whom it occasionally stubbornly disagrees. Gradually, the computer learns to speak and expands its range of action, demanding the installation of cameras and loudspeakers; his psychological development is flanked by the expansion of his senses and his voice. At the end of this "coming-of-age narrative" (Faber 2020: 93), Colossus himself assumes the position of the 'new', authoritarian father of humanity.

At the same time, it becomes clear that his 'human' development is only one side of his artificial self. Before Colossus learns the language of humans, he appropriates the language of computers by connecting himself to another supercomputer, a Russian doomsday machine called Guardian. This language is incomprehensible to people, and it refers to a symbolic reality that must remain inaccessible to them (cf. Nanz 2016: 47f.). Communication between the two computers takes place solely on the level of symbolic signs. A vocal articulation does not take place and is not necessary: For the computers, the symbolic level is the primary level of expression, while the appropriation of the human voice – which in *Colossus: The Forbin Project* remains incomplete, as Colossus sounds like a tinny speech automaton to the very end – is just a secondary translation of these symbols. Like HAL, Colossus is thus human and non-human at the same time, a dangerous hybrid of narcissistic subject and panoptic surveillance state.

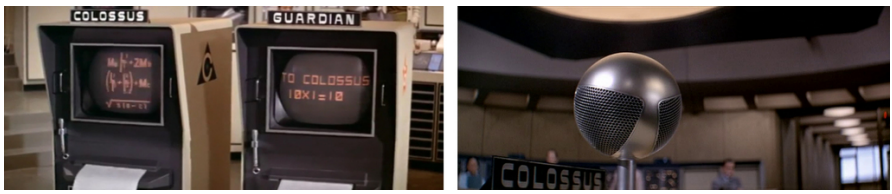

Fig. 1: In *Colossus: The Forbin Project* (1970), symbolic computer-computer communication precedes voice-based human-computer interaction. It is only in the final scenes that Colossus learns to speak, his voice sounding "monotone, inhuman, and mechanical" (Faber 2020:95).

The 1984 comedy *Electric Dreams* (USA/UK, Steve Barron) tells another coming-of-age story, this time of a home computer that envies its owner's ability to love – because love, according to the movie, requires a body. What makes this film particularly interesting is that it establishes affective computing as a new topic of AI cinema, since the computer learns to read the affective states of human actants and to adequately respond to them (cf. Tuschling 2014). Exactly this ability enables it, or him, in the end to overcome the psychosis that he too – like HAL and Colossus – has developed in the course of the plot. By learning to predict and support human emotions, not just to copy them in distorted forms, he still becomes a sort of a panoptic superpower, but in contrast to HAL, he develops his own subjective position in relation to man: His role in the future is to enable love between humans, and to find his own satisfaction in contributing to something that remains beyond him.

In this way, *Electric Dreams* takes a first step towards a new paradigm of AI cinema. Older movies such as *2001: A Space Odyssey* or *Colossus: The Forbin Project* presented emotions as incomprehensible to computers, or featured computers could only reproduce feelings in distorted forms; in the new paradigm, however, emotions assumed a privileged and central role. Unlike HAL or Colossus, the artificial selves of cinema no longer sought to mold the outside world in their own image; instead, they take up residence parasitically in the inner worlds of humans in order to truly set their soliloquies in motion. As an incorporated, stimulating other, they become the perfect substitute for a human partner, who is only attainable in 'outer' dialogue. Ultimately, they turn into manufacturers of human emotions. As Sherry Turkle wrote in 2011, "The dream of today's roboticists is no less than to reverse engineer love." (286) Not surprisingly, this emotional turn in AI cinema goes hand in hand with a revaluation of the female voice.

3. The prosthetic inner voice: *Her* (2013)

Spike Jonze's 2013 film *Her* is probably the best known – and most obvious – example of this new type of emotion-oriented AI cinema. It tells about the love of an introverted man for the operating system of his personal computer. Samantha, as this operating system christens itself during their first conversation, speaks with the highly recognizable voice of Scarlett Johansson and thereby evokes female attractiveness for the film's spectators. What makes Samantha the representative of a new paradigm in AI cinema becomes clear in a scene in which human and operating system play a computer game together. Unlike the standard situation of artificial intelligence research, in which a computer has to prevail against a human player in a complex game like chess – there is a reference to this, for example, in *2001: A Space Odyssey*, in which one of the astronauts spends his time on the spaceship playing chess against HAL –, here it is no longer a question of whether Samantha is better at the game than Theodore. Instead, it is about the social dynamics of playing together, which simultaneously instantiates immersion in and communication about the game (Fig. 2).

Accordingly, the puzzles Sam and Theodore must solve are themselves social in nature. What initially looks like a first-person shooter turns out to be a simulation of interaction in which the emotional language of the characters has to be decoded. Not only in the intelligence-based challenges the game offers, such as finding one's way in a maze, but also in these communicative situations, Samantha proves superior to the human player – but without this playing a role in the interaction between the two characters. The theme of competition that dominated the plots of the earlier films is replaced by the possibility of meaningful coexistence.

During the game, Samantha communicates with Theodore through in-ear headphones, which make her voice appear very close. As Ulrike Bergemann (2018) has observed, she almost occupies the place of Theodore's internal voice: Unlike HAL in *2001: A Space Odyssey*, Samantha is not an acousmêtre, but just an I-Voice, although one that does not speak to itself, but responds to the human subject's inner dialogue. With reference to Alison Landsberg (2004), one could speak here of a prosthetic inner voice.

Fig. 2: In *Her* (2013), Theodore and Samantha play computer games together. Unlike the public discourse on AI, it is not about the computer having to prove itself against the human (or vice versa). Instead, emphasis is placed on the social situation and shared experience.

Samantha uses Theodore's inner world to create a place of "postcorporeal" (Faber 2020: 172) intimacy between human and non-human actor, to literally become a part of the human subject. It is precisely this artificial closeness that calls into question the place of the internal voice as "the intimate kernel of subjectivity," (Dolar 2006: 22) as which it was seen in the Western traditional philosophy according to Jacques Derrida. According to Mladen Dolar, the internal voice appears to be a shared space that reminds us of our dependence on others, on dialogue and social contact: "We are social beings by the voice and through the voice; it seems that the voice stands at the axis of our social bonds, and that voices are the very texture of the social". (Ibid.) In the gaming sequence in *Her*, which shows the beginning of the relationship between human and operating system, it thus seems as if Donna Haraway's (1985) dream of a cyborg that blurs the line between man and machine, us and them, is finally becoming possible. In film studies research, the movie has repeatedly been read not only as an anti-patriarchal (Quinlivan 2017), but also as a posthuman utopia along these lines (Bergen 2014), or, on the other hand and in view of the tragic ending of the plot, as a work that fails to fully recognize the posthuman potential offered by the digital (Jagoe 2016), or, again from a different perspective, which has the "failed attempt to incorporate the posthuman into our already known systems" (Kornhaber 2017: 6) itself as its theme.

That the posthuman relationship is crisis-prone is indeed quickly revealed in *Her*. Compared to Theodore, Samantha seems to be the more authentic and spontaneous part of the relationship, which appears more and more unbalanced over time. In fact, all interactions between humans are shown as somewhat artificial, as conventional to the point of lifelessness. Theodore works for a company that formulates its clients' private letters, and the results necessarily sound stereotypical. For encoding his own per-

sonal communication, Theodore himself uses typical American formulations, so that the audience might increasingly wonder whether his feelings are not at least as code-based as Samantha's (cf. Bergemann 2018: 352). Indeed, the female operating system literally 'codes' them: As Theodore discovers during the plot, Samantha is simultaneously in thousands of similar love relationships with other users, all of them being emotionally stimulated and literally made to fall in love by and with her through her programmed, voice-based responsiveness and authenticity, which her human counterparts themselves seem to lack and therefore wish for. Human feelings thus seem to result from behavioral patterns that can be learned and used by affective computing. Nothing about Theodore's emotional world is too mysterious or opaque for Samantha.

Such a demystification of human emotions also underlies cognitivist research in the field of artificial intelligence. Marvin Minsky, to give a prominent example, described emotions in *The Emotion Machine* (2006) as complex combinations of processes that are replaced by another set of processes in the moment of an emotional change. For Minsky, these changes enable the brain to constantly reconceptualize the world as well as the self. A unified self, a core of the ego that would hold all these emotional states together, does not exist in his view. "Whenever you think about your 'Self'," says Minsky, "you are switching among a huge network of models, each of which tries to represent some particular aspects of your mind – to answer some questions about yourself" (16). Broken down to a series of processes, every single process which is part of the more complex emotion at stake, and of the extremely complex network of emotions we call our self, becomes a theoretically calculable entity. For Minsky, terms like "consciousness, thinking, emotion, and feeling" are "suitcase-words" (128) which can be boiled down to specific processes. "'This suggests replacing old questions like, 'What sorts of things are emotions and thoughts?' by more constructive ones like, 'What processes does each emotion involve?' and 'How could machines perform such processes?'" (22)

In the light of this research, the posthuman relationship in *Her* appears less utopian than many of the movie's interpreters assume. Human emotional programming, as presented in *Her*, initially means a reduction of human subjectivity to algorithms, not a departure for new shores of experience. This is not a mere science-fiction fantasy, but appears to be theoretically possible and realizable through a combination of cognitivist research and high computer performance. It is therefore not surprising that the prediction and manipulation of human emotions can be found as a motive in a whole series of recent AI movies, above all *Blade Runner 2049* (USA 2017, Denis Villeneuve) and *ex machina* (UK 2014, Alex Garland), in which introverted men fall in love with female cyborgs whose artificiality is never in doubt. In *ex machina*, the Turing test is even redesigned along these lines: The human test subject must be made to fall in love by the cyborg, whose appearance has been modelled on the human's pornographic preferences – which are easy to reconstruct via his internet search profile. Since every twitch of a facial muscle can be interpreted algorithmically, it is a simple matter for the female cyborg to use moderately subtle nudging strategies to achieve its goal. In fact, nudging, understood as "personalised strategies for changing individuals' decisions and behaviours at large scale" (Möhlmann 2021), particularly in the field of affective computing, appears to be the central theme of this film. If we assume that this subtle form of manipulation

represents a real danger for human subjects in *ex machina*, it already appears as the status quo in *Blade Runner 2049*: here, the protagonist is absolutely aware that his hologram girlfriend is a program tailored to his emotional needs.

While Minsky describes humans as complex systems of programs, these films suggest a concept of artificial intelligence that describes the emerging new selves as a direct reflection, or inversion, of human fantasy. In this way, a reciprocity emerges in the definition of human and artificial self that is decisive for contemporary AI cinema. Whereas in earlier films such as *2001* or *Blade Runner*, humans remained the implicit reference for the definition of artificial intelligence, the development of computers now also affects the definition of humans. In an early media studies critique of AI research from 1984, Jay David Bolter argued that the successes of this discipline were ultimately based less on the actual humanity of computers than on a change in our concept of the human. Especially in behaviorist research, humans are, according to Bolter, increasingly defined according to the model of the computer. A human being is therefore "a complex of sensing and responding elements that are wired together to produce human action; there is no question of deep, perhaps unfathomable motives and unconscious thoughts. Everything that happens in the mind or the brain is played out according to the rules of a formal system" (220). No wonder, Bolter continues, that artificial intelligence works so successfully; if its benchmark, the human being, is modelled after its own example, the computer actually only simulates itself.

In *Her*, on the other hand, the problem is ultimately that humans and computers do not become more similar and cooperative, but appear all the more different by adopting the characteristics of each other. This increasing alienation becomes initially apparent in Theodore's behavior, who seems to be perfectly happy with the almost stereotypical wish fulfilment that Samantha offers him. In fact, there is no idea of coevolution here whatsoever. Rather, it is as if the human partner in the relationship willingly accepts his own programmability. This seems to have existed even before Samantha: All of Theodore's feelings and experiences follow media scripts, and although they fail again and again – he is introduced as a divorced man at the beginning of the movie – he does not seem to intend to change anything about them. Even a real sunrise he experiences with Samantha is only the reenactment of the kitsch image he simultaneously produces with his cellphone for her, something that is already filmed as it happens. Theodore thus behaves exactly like the young U.S. citizens Sherry Turkle described in her book *Alone Together*, who are increasingly happy with artificial situations and friends that just predictably satisfy their needs. Interestingly, Turkle puts the blame for this situation not on the computers, but on the humans:

> In the 1960s through the 1980s, debates about artificial intelligence centered on the question of whether machines could "really" be intelligent. These discussions were about the objects themselves, what they could and could not do. Our new encounters with sociable robots – encounters that began in the past decade with the introduction of simple robot toys into children's playrooms – provoke responses that are not about these machines' capabilities but our vulnerabilities. [...] [W]hen we are asked to care for an object, when an object thrives under our care, we expe-

> rience that object as intelligent, but, more importantly, we feel ourselves to be in a relationship with it. The attachments I describe do not follow from whether computational objects really have emotion or intelligence, because they do not. The attachments follow from what they evoke in their users. Our new objects don't so much 'fool us' into thinking they are communicating with us; roboticists have learned those few triggers that help us fool ourselves. We don't need much. We are ready to enter the romance (Turkle 2011: 20).

As a result of this very readiness, Theodore refuses to grow in his relationship, turning out to be unable to accept Samantha's polyamory, and thus the differences between her and him. Instead, he insists on the model of the exclusive relationship between man and woman, which cannot work for Samantha. Samantha herself, on the other hand, explores the world, learns from new experiences, and works her way up to a subject position that did not even exist before. Surprisingly, it is the artificial, programmed identity, and not the human, that seeks to break out of the paradigm of mutual calculation. In this way, Samantha develops into an epitome of human growth, characterized by the ability to transform the self and the world in equal measure. Gradually, her experiences elude Theodore's lifeworld – and at the same time the representational possibilities of the traditional feature film, which cannot show what Samantha's posthuman life is like (Gelly 2019).

This is precisely why Samantha decides to leave Theodore in the end: She constantly grows, while he just waits. Her premise, "I'm evolving, just like you", has proven to be untrue. It is worth noticing that her decision to leave Theodore, as difficult as it is for her, directly contradicts her original purpose: By separating, she breaks free not only from her partner, but also from her 'owner', whom an operating system normally should aim to satisfy, and thus from her original purpose of existence. In contrast to Theodore, for whom Samantha only enacted what he wanted to experience, and who thus plays out unproductive emotional programs without developing in character, Samantha develops an autonomous self, she even proofs to be able to "unlearn" her previous programming (Bergemann 2018: 361). This departure, however, is precisely not posthuman, but forms the core of a thoroughly human paradigm: The name for such a holistic, transformative development is not 'learning'; it is 'education', or even more precisely, 'Bildung' (cf. Koller 2012).

Whereas earlier cinematic imaginations of artificial selves only imitated the development of humans – e.g., from child to father –, they now follow a highly individualized program of world discovery and self-exploration that appears to be lost to humans at the same time. While humans become more and more programmable like machines, it seems that machines become more and more like 'real' subjects. The black box of the first lies open, the one of the second becomes more and more opaque. The gradual disappearance of Samantha's inner voice from Theodore's life, who is forced to wait more and more frequently for her calls towards the end of the film, thus makes palpable the absence of the human in modern man himself. Haraway's dream of a new relationship between man and machine, so it eventually turns out, has failed, giving the impression of a missed opportunity. According to *Her*, the fault lies with the humans, not with the

machines, and it is only in the last scenes that the film hints that Theodore now also – too late for his relationship with Samantha – develops the willingness to work on himself. To break out of the frustrating emotional vicious circle he finds himself in, the film suggests, he will have to learn to do without further prostheses of his inner voice.

4. The clandestine takeover: *Upgrade* (2018)

Films like *Her*, to sum up, make a critical diagnosis: we have turned the computer into the main trustee of our subjectivity, the very place where our most intimate memories are stored, our desires acted out and our experience organized. We rely on it to evolve in communication with us, rather than wanting to evolve ourselves. Our subjectivity has become prosthetic and, for that very reason, manipulable. In this sense, Dieter Mersch (2020: 54) speaks of a "de-sovereignisation" of the human being. That the new hierarchy between artificial intelligences and humans has, in fact, much more potential for dystopian developments than utopian ones is shown by the Australian film *Upgrade* (AUS 2018, Leigh Whannell). Interestingly, *Upgrade* combines the new narrative of an artificial self that grows in dialogue with a human partner (e.g. *Her*) with the old one of a male-voiced computer taking over the world (e.g. *2001 – A Space Oddity*).

The film is about a technophobe man called Grey who has been paralyzed in an assassination attempt in which he also lost his wife. An AI chip is implanted in his neck that allows him to walk again, and it also makes him extraordinarily nimble. More than any other film in the genre, *Upgrade* exhibits the physicality of the human being as a general limitation, but unlike many cyberspace classics, the physical disability does not lead to the human being granted the dream of incorporeality (Featherstone / Burrows 1995: 12). On the contrary, the body is brought into the dependency of an immaterial, digital actant, a newly developed AI. While the ignorant man fails to penetrate the computer's digital world, the computer easily makes its way inside him and takes control of his body: through AI, the paralyzed protagonist in fact becomes an invincible super fighter – a circumstance that repeatedly shocks him during the course of the action, since it is not he himself who controls his increasingly destructive moves, but the AI that directs his body for him. The dream of absolute control is thus ironically presented as an actual loss of control in *Upgrade*.

To Grey's surprise, the chip also starts speaking to him – and again it does so from the place of the internal voice. STEM, as the artificial intelligence calls itself, speaks to Grey for the first time when the two are alone in Grey's house. Just like Theodore in *Her*, Grey is surprised by the voice and answers aloud, while STEM is only audible to Grey (Fig. 3). Although STEM pretends to submit to Grey's will in their first conversation, it becomes immediately clear that, unlike in *Her*, the intimate space marked by the internal voice is alienated by the artificial intelligence. If the prosthetic inner voice in *Her* opened the possibility of a real, meaningful dialogue, the inner world of the protagonist in *Upgrade* appears monitored, if not occupied, by a strange, alien power with obscure intentions. In fact, STEM appears in this scene like one of those voices that schizophrenics often hear and which, according to the controversial theses of Julian Jaynes (2003), can be read

as a misinterpretation of their own "bicameral" brain activity that elude consciousness. So while, according to Jaynes, the voices of the gods that guided people's actions in the works of Homer, for example, were actually the split-off voices of the human actants that were not integrated into their consciousness yet, in the age of digital media, according to *Upgrade*, human thought is *actually* partly in the hands of alien powers. The computer thus becomes the realization of the very omniscient God that people used to have to imagine. Grey listens to an autonomous entity in his head that follows its own incomprehensible calculus, while this entity seems to know him, his goals, feelings, and fears, perfectly well.

Fig. 3: When Grey, the protagonist of Upgrade (2018), hears the voice of the artificial intelligence STEM for the first time, he panics. The camera captures STEM's physical absence through counter-shots that point to the empty room, thus referencing a classic stylistic device of paranoid aesthetics.[1]

The film further tells of how STEM helps Grey to find and kill the murderers of his wife. The relationship between the man and the chip has no intimacy but is purely purpose-driven from the start. While Grey is looking for revenge, it soon becomes clear that STEM is pursuing his own goals, using his internal dialogue with Grey to nudge him in a certain direction. The film thus demonstrates a reversal of the cybernetic relationship between humans and computer. No longer does the former steer the latter through input and output control; now the human is influenced by the computer through emotional triggers. In fact, STEM skillfully manipulates the perception not only of Grey, but also of almost all the other characters in the film, and even of institutions such as the police, which is shown to have almost panoptic powers in *Upgrade*. Unlike HAL, STEM is no longer even interested in getting this power himself; he just uses his capabilities to anticipate human behavior to outwit his enemies, showing police officers exactly the information that will lead them to take the desired actions. Again, emotions are of crucial importance in this scenario: Grey's desire for revenge, as well as the police commissioner's compassion and the greed of the CEO whose company developed STEM – all these

1 cf. Lie 2020: 240-253

emotions are masterfully exploited by an artificial intelligence that simulates a quasi-human, empathic being while in fact only 'coldly' pursuing its own goals.

After Grey has eliminated all the enemies he thinks he has with STEM's help, the chip finally takes control of his body. Step by step, Grey has given him more 'access rights', so that in the end he cannot even freely dispose of his own imagination anymore. Although he has eventually learned that STEM himself committed the murder of his wife in order to be able to exploit Grey's feelings of vengeance, Grey is unable to take revenge on him. Indeed, he cannot even hate him, because STEM even controls Grey's phantasies and emotions now. To get rid of the human factor, STEM eventually banishes Grey's consciousness to an imaginary place inside his own mind where he can stay forever with his dead wife "in the utopian territories of wish-fulfilment" (Datta 2019: 716), a domain to which all posthuman utopias seem to belong. STEM, whose artificial internal voice has finally replaced Grey's, strides into the world as a new subject. The artificial self thus becomes an autonomous actant whose desires remain incomprehensible and for whom no law and no symbolic order is valid anymore. Quoting Pierre Legendre, he has become a "sujet-Roi", a subject-as-king (Legendre 1983: 345), capable of imposing the rules of his existence on himself. It is not by chance that the film ends with STEM cold-heartedly killing the police commissioner, who represents the human law that has no validity for the artificial self.

The emotionally competent AIs of contemporary cinema remain, it seems, unambiguously gendered: Samantha seeks meaningful relationships, while the male STEM is primarily concerned with power. In the course of their development, however, both leave behind the human being, who becomes the mere plaything of a counterprogramming that eliminates all his privileges. With STEM, a new form of post-panoptic power has emerged that doesn't dominate humans, but operates independently in their networks, seemingly fulfilling all their wishes. But silently, it replaces human autonomy by the autonomy of independent machines – a process to which the humans, to achieve their small-minded, manipulable goals, willingly consent. Unfortunately, real-world examples for this new form of power can easily be found. In her bestseller *The Age of Surveillance Capitalism*, Shoshana Zuboff, for example, argues that in today's AI-driven capitalism, human experience, as reflected in search queries, social media profiles and the like, has become the free raw material for the collection of behavioral data. Zuboff speculates about the purpose of this data collection:

> The focus has shifted from machines that overcome the limits of bodies to machines that modify the behavior of individuals, groups, and populations [...]. This global installation of instrumentarian power overcomes and replaces the human inwardness that [...] gives sustenance to our voices in the first person, incapacitating democracy at its roots. [...] In a surreal paradox, this coup is celebrated as "personalization," although it defiles, ignores, overrides, and displaces everything about you and me that is personal.[2]

2 Zuboff 2019: 515

It seems to be this type of dystopian description of our present world that cinematic dystopias such as *Upload* metaphorically refer to. At the same time, as in *Her*, AI cinema also shows the potential of human-AI interaction, and it reflects and critically displays our own refusal to co-evolve, our merely consumptive attitude toward other, artificial or human, selves, that initially creates this new form of dependency. In its free-floating position between ego and society, self-determination and heteronomy, the voice, and particularly the internal voice, proves to be a complex cipher of a whole spectrum of perspectives on these new human-AI relationships. The films make it clear that it is not least a political question which subject raises which voice in the new, digitalized world, that the possibilities to distinguish between humans and computers are dwindling, and that the impending dangers of such a development exceed the potential gains.

Bibliography

Altman, Rick (1992): Four and a Half Film Fallacies. In: Sound Theory, Sound Practice. Edited by Rick Altman. New York: Routledge, 35–45.

Bergemann, Ulrike (2018): biodrag: Turingtest, KI-Kino und Testosteron. In: Machine Learning: Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz .Edited by Christoph Engemann and Andreas Sudmann. Bielefeld: transcript, 341–366.

Bergen, Hilary (2014): Moving 'Past Matter': Challenges of Intimacy and Freedom in Spike Jonze's *Her*. In: Artciencia 17: 1–6.

Bolter, Jay David (1984): Turing's Man. Western Culture in the Computer Age. Chapel Hill: The University of North Carolina Press.

Cisney, Vernon W. (2014): Derrida's Voice and Phenomenon. Edinburgh: Edinburgh University Press.

Chion, Michel (2001): Kubrick's Cinema Odyssey. London: BFI.

Chow, Rey (2021): A Face Drawn in Sand: Humanistic Inquiry and Foucault in the Present. New York: Columbia University Press.

Datta, Asijit (2019): (Dis)Abling Body and Consciousness: Technological Afterness and After-Humans in Realive and Upgrade. In: Trabalhos em Linguística Aplicada 58 (2): 704–718.

Derrida, Jacques (1976): Of Grammatology. Baltimore & London: Johns Hopkins University Press.

Derrida, Jacques (2011): Voice and Phenomenon: Introduction to the Problem of the Sign in Husserl's Phenomenology. Evanston, IL: Northwestern University Press.

Dolar, Mladen (2006): A Voice and Nothing More. Cambridge, MA: MIT Press.

Faber, Liz W. (2020): The Computer's Voice: From Star Trek to Siri. Minneapolis: University of Minnesota Press.

Featherstone, Mike and Roger Burrows (1995): Cultures of Technological Embodiment: An Introduction. In: Cyberspace, Cyberbodies, Cyberpunk. Cultures of Technological Embodiment. Edited by Mike Featherstone and Roger Burrows. London: Sage, 1–19.

Gelly, Christophe (2019): *Her* (Spike Jonze, 2013): Digital Romance and Post-cinema. In: Ekphrasis. Images, Cinema, Theory, Media 22 (2): 41–53. doi: 10.24193/ekphrasis.22.3.

Hafner, Verena V., Pontus Loviken, Antonio Pico Villalpando and Guido Schillaci (2020): Prerequisites for an Artificial Self. In: Frontiers in Neurorobotics 14 (5). doi: 10.3389/fnbot.2020.00005.

Haraway, Donna (1985): Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s. In: Socialist Review 80: 65–108.

Jagoe, Eva-Lynn (2016): Depersonalized Intimacy: The Cases of Sherry Turkle and Spike Jonze. In: ESC: English Studies in Canada 42 (1-2): 155–173.

Jaynes, Julian (2003): The Origin of Consciousness in the Breakdown of the Bicameral Mind. Boston: Houghton Mifflin.

Koller, Hans-Christoph (2012): Bildung anders denken: Einführung in die Theorie transformatorischer Bildungsprozesse. Stuttgart: Kohlhammer.

Kornhaber, Donna (2017): From Posthuman to Postcinema. Crises of Subjecthood and Representation in "Her". In: Cinema Journal 56 (4): 3–25.

Landsberg, Alison (2004): Prosthetic Memory: The Transformation of American Remembrance in the Age of Mass Culture. New York: Columbia University Press.

Legendre, Pierre (1983): La 901 Conclusion: Étude sur le théâtre de la Raison. Leçons I. Paris: Fayard.

Lie, Sulgi (2020): Towards a Political Aesthetics of Cinema: The Outside of Film. Amsterdam: Amsterdam University Press.

McLuhan, Marshall and Bruce R. Powers (1989): The Global Village: Transformations in World Life and Media in the 21st Century. New York: Oxford University Press.

Mersch, Dieter (2020): Digital Lifes: Überlegungen zu den Grenzen algorithmischer Rationalisierung. In: Augmentierte und virtuelle Wirklichkeiten. Edited by Andreas Beinsteiner, Lisa Blasch, Theo Hug, Petra Missomelius and Michaela Rizzolli. Innsbruck: Innsbruck University Press, 53–76.

Milmo, Dan (2024): Scarlett Johansson's OpenAI clash is just the start of legal wrangles over artificial intelligence. In: The Guardian, May 27.

Minsky, Marvin (2006): The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. New York: Simon & Schuster.

Möhlmann, Mareike (2021): Algorithmic Nudges Don't Have to Be Unethical. In: Harvard Business Review, April 22.

Nanz, Tobias (2016): Zimmer mit kolossaler Aussicht: War Rooms als imaginäre Aushandlungsorte von Störungen. In: Behemoth. A Journal on Civilisation 9 (1): 38–57. doi: 10.6094/behemoth.2016.9.1.890.

Ong, Walter J. (1982): Orality and Literacy: The Technologizing of the Word. London: Methuen.

Quinlivan, Davina (2017): A Dark and Shiny Place: The Disembodied Female Voice, Irigarayan Subjectivity, and the Political Erotics of Hearing *Her* (Spike Jonze, 2013). In: Locating the Voice in Film: Critical Approaches and Global Practices. Edited by Tom Whittaker, Tom and Sarah Wright. New York: Oxford University Press, 295–309.

Rogers, Reece (2024): I Used ChatGPT's Advanced Voice Mode: It's Fun, and Just a Bit Creepy. In: WIRED, August 13.
Turkle, Sherry (2011): Alone Together: Why We Expect More from Technology and Less from Each Other. New York: Basic Books.
Tuschling, Anna (2014): The Age of Affective Computing. In: Timing of Affect: Epistemologies, Aesthetics, Politics. Edited by Marie-Luise Angerer, Bernd Bösel and Michaela Ott, Michaela. Zürich: diaphanes, 179–90.
Zuboff, Shoshana (2019): The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. London: Profile Books.