e-ISSN 2420-823X

English Literature
Vol. 9 – December 2022

The Science of Fiction
Human-Robot Interaction in McEwan's *Machines Like Me*

Silvana Colella
Università di Macerata, Italia

Abstract This article focuses on human-robot interaction and anthropomorphism in Ian McEwan's *Machines Like Me*. After considering the novel's reception among scientists, reviewers and readers, the first section analyzes the uses of digression in the text, the counterfactual mode, and how they affect the representation of human-robot interaction. The second section explores the tension between the myth and reality of AI, arguing that the novel provides salient commentary on 'dishonest anthropomorphism' while parading the idea of machine consciousness, via the diegetic presence of Alan Turing.

Keywords Human-robot Interaction. Anthropomorphism. Artificial Intelligence. Ian McEwan. Fiction.

Edizioni
Ca'Foscari

Peer review

Submitted	2022-07-07
Accepted	2022-10-15
Published	2023-04-13

Open access

Citation Colella, S. (2023). "The Science of Fiction. Human-Robot Interaction in McEwan's *Machines Like Me*". *English Literature*, 9, 147-166.

DOI 10.30687/EL/2420-823X/2022/01/008

Since art is science with an addition, since some science underlies all Art, there is seemingly no paradox in the use of such a phrase as "the Science of Fiction".
Thomas Hardy, *The Science of Fiction*, 1891

1 Introduction

The science underlying *Machines Like Me* has two main components. The most obvious one is rooted in current developments of AI systems. References to machine learning, deep learning, neural networks and the (yet unsolved) P versus NP mathematical problem appear explicitly in the text, mostly in relation to the diegetic presence of Alan Turing and his theories. The second component concerns HRI (Human-Robot Interaction)[1] and is central in the plot, focused on the relationship between Charlie, Miranda and the robot Adam. The novel has garnered the attention of scientists working in the field of HRI. For Gaggioli et al. (2021), the scenario delineated in the novel

> forces us to ask ourselves what we want for future robotics. Do we desire that robots become passive prostheses that extend our natural capabilities under our direct control, or do we wish to develop artificial entities that are capable of autonomy, mutual understanding, empathy and ultimately relational skills? (357)

In their assessment, a shift of emphasis from human-robot interaction to "human-robot shared experience" (360) would be a productive development.

Machines Like Me has also inspired reflection on the issue of robot clothes: Friedman et al. (2021) quote the novel in their discussion of "wire modesty", the kind of modesty which may originate in "anthropomorphic priggishness" but has pragmatic utility since "exposed wires present a real risk to function" (1347). Finally, in an article published in *The Journal of Craniofacial Surgery*, Montandon (2021) begins and ends his discussion of "enfacement illusions" with references to *Machines Like Me*: "Reading the bestseller *Machines Like Me and People Like You* by Ian McEwan, might let you think that robots can have a strong personality and many other attributes and

[1] As defined by Bartneck et al. (2020), HRI is a large, multidisciplinary field that "brings together scholars and practitioners from various domains: engineers, psychologists, designers, anthropologists, sociologists, and philosophers, along with scholars from other application and research domains. Creating a successful human-robot interaction requires collaboration from a variety of fields to develop the robotics hardware and software, analyze the behavior of humans when interacting with robots in different social contexts, and create the aesthetics of the embodiment and behavior of the robot, as well as the required domain knowledge for particular applications" (9).

functions" (1652), but excessive similarity between humans and robots engenders a sense of unease and creepiness.

As these studies indicate, Adam's embodiment – futuristic, improbable, yet deserving of attention – is a conspicuous element in the way McEwan imagines artificial intelligence. In Turing's imitation game, the machine is hidden from view, the participants do not show themselves. The C interrogator cannot see them nor hear their voices, he can only rely on language to detect a machinic agent (Turing 1950). The Turing test implies invisibility. Not in the novel, though, where Adam is first and foremost a spectacular anthropomorphic form endowed with human-like functions.

In this article, I shall consider both sides of the science of fiction: the conjectural one, pivoting on "the myth of artificial intelligence",[2] propounded in diegetic Turing's eloquent arguments, and the practical or experimental one. McEwan's lab is a small apartment in Clapham, where human-robot interaction is tested. The plot hinges on the relationship between Charlie, Miranda and their recently acquired social robot, Adam, whose anthropomorphic design is as astounding as his much-advertised ability to 'think'. The body and behavior of the android are the object of close observation throughout the story, and so too are the responses of the humans to anthropomorphic technology. McEwan's experiment also tests the reader's willingness to accord Adam the full privilege of "seeming human".[3] "I want the reader to be in Charlie's shoes", declares McEwan in an interview:

> as he's contending with someone who has a superior character and who can discuss Shakespeare with some warmth and insight. At the end, do you think Adam is a cold-blooded machine or a sen-

[2] I refer here to the myth of AI as articulated by Larson (2021). Its central component is the idea of "the inevitability of AI […] ingrained in popular discussion" and promoted by several AI scientists (1). According to Larson, "the inferences that systems require for general intelligence […] cannot be programmed, learned or engineered with our current knowledge of AI. As we successfully apply simpler, narrow versions of intelligence that benefit from faster computers and lots of data, we are not making incremental progress, but rather picking low-hanging fruit" (2). For Broussard (2018), "*general AI* is the Hollywood kind of AI. General AI is anything to do with sentient robots (who may or may not want to take over the world), consciousness inside computers, eternal life, or machines that 'think' like humans. *Narrow AI* is different: it's a mathematical method for prediction. There's a lot of confusion between the two, even among people who make technological systems. Again, general AI is what some people want, and narrow AI is what we have" (32). Broussard too is keen to demystify the "ghost-in-the-machine fallacy" (39) and the alleged "magic" of algorithms (36). See also Crawford (2021), Natale (2021) and Marcus, Davies (2020) for sharp critical views countering the current hype around AI.

[3] As Ward (2018) observes, both fictional characters and AI "aspire toward the appearance of reproducing the human […] Mimesis for fictional characters and intelligent machines means producing something that is human-like but also not-human, a new thing in and of itself. They are copies that are also new originals" (n.p.).

> tient being? That's the issue we're going to have, and it's going to open up new territory for us in the moral dimension. (Miller 2019)

As I argue in this article, the question of how Adam is perceived depends on several factors, including formal aspects. Copiously described world-making, essayistic detours and the autodiegetic narrative perspective affect our understanding of the evolving interaction between the humans and the robot.

After considering the novel's reception among professional reviewers and lay readers, the first section focuses on the uses of digression in the text and their effects on how the android is perceived. In the second section, I claim that the novel provides salient commentary on "dishonest anthropomorphism" (Leong, Selinger 2019) even as Turing's messianic rhetoric promotes a benevolent, utopian understanding of machines "like us". Adam is in love, he has feelings, he appreciates literature and creates thousands of haikus – the humanities version of general artificial intelligence is harmless and admirable. But the plot revolves around Adam's invisible, opaque decision-making mechanisms, his algorithmic superhuman powers. This black-box version of narrow, goal-oriented AI, closer to the reality rather than the myth of AI, crops up now and again in the text. While the characters in the story tend to disregard weak signals of Adam's nonhuman, machinic essence, until it is too late, red flags are raised as the narrative delves into the complexity of human-robot interaction. What Larson (2021, 60) calls the "technological kitsch" – the current infatuation with notions of general AI, 'singularity', or 'ultraintelligence' – is both acclaimed and questioned in the novel.

2 The Uses of Digression

One formal feature that readers of the novel have frequently singled out is the narrator's penchant for essayistic digressions. Charlie's detours, which illuminate the counterfactual historical background, have appeared somewhat disconnected from the flow of the story. Reviewing the novel for *The Irish Times*, Rebecca Saleem claims that the story's "unnecessary detours" result in a "baggy and jumbled narrative" (2019). For Marcel Theroux, the narrator "overexplains the historical context and never turns down a chance to offer an essayistic digression", which, in his opinion, is a "flat-footed way of doing sci-fi" (2019). Ian Patterson observes that the "novel's mythical past with its history so similar to our own, sometimes rhetorically overdone […] and often full of amusing detail, is just there to communicate a comfortable sense of distance" (2019).

The lay readers who have posted their opinions on the Goodreads platform are even more vocal in targeting McEwan's "historical tinkering" as "distracting" or "dispensable". As one reader argues, the alternate setting "distracted from the main story and was rather irritating at times". The "nerdy facts" McEwan injects in the story, though valuable, are regarded as interruptions interfering with immersivity. The novel reads "like a Bill Bryson's book in a number of interesting topics and nerdy facts McEwan throws at the reader", writes an admiring reader:

> For example, solving the P versus NP problem (what?) a major unsolved computer science problem. I mean this is just one example that sent me scampering for information and references, I spent time on it, and still have no idea what it's about really. There is a lot of stuff like this in this book […] I found myself putting this book down a lot, to either ponder a situation, action or thought of one of the characters or to look up an interesting fact or topic mentioned. I'm knackered!![4]

These responses highlight a structural element in the novel's form that affects our engagement with the story, whether soliciting further reflection – "scampering for information" – or simply disturbing the unspooling of the plot. Put differently, the experience of reading *Machines Like Me*, placing oneself in Charlie's shoes as McEwan invites us to do, is not a smooth ride. The bumps in the road slow down the reading process, create distractions or induce a sense of distance. My contention is that the bumps in the road occur at specific junctures in the story, mostly when Charlie's (and ours) recognition of Adam's human-level consciousness is at its highest, thus suspending empathetic identification with the android. Whether we perceive Adam as a sentient being is contingent on the tension between distancing and engaging strategies embedded in the shape of the narrative.

The narrative situation that McEwan imagines in the first chapter focuses on the robot coming alive. Brought home in a stretcher, Adam sits at the kitchen table, impressive in his nudity and immobility, waiting for his batteries to charge. Charlie is impatient: "I wanted him now", he says (McEwan 2019a, 3).[5] But he has to wait sixteen hours, and we, the readers, wait with him. Time elongates, the moment of birth is deferred while Charlie expatiates on several dimensions of private and public life: "What turmoil on a weekday afternoon", Charlie admits, "a new kind of being at my dining table, the woman I newly loved six feet above my head, and the country at old-fashioned war" (21).

[4] See https://www.goodreads.com/book/show/42086795-machines-like-me.

[5] Henceforth, page references only will be provided in the text.

While the robot is still inactivated, and his artificial intelligence has yet to manifest itself, the narrative dwells on the functioning of general human intelligence: the ability to infer, to entertain thoughts and their connections; to jump domains; to formulate hypotheses and plans, and to switch from one task to another. One moment Charlie is closely scrutinizing the body of the android, looking for the slightest hints of artificiality, and the next he is reminiscing, in a lyrical mode, about his school days. Human intelligence, Larson (2021) argues, is "situational", "contextual" and "externalized": "General (non-narrow) intelligence of the sort we display daily is not an algorithm running in our heads, but calls on the entire cultural, historical, and social context within which we think and act in the world" (31).[6]

The first chapter establishes Charlie's credentials not only as the narrator, who duly introduces himself to the readers, but also as a human subject whose intelligence can effortlessly "grasp the world". The quotation inscribed on the Fields Medal, which Charlie suddenly remembers at the end of the novel, could be his motto: "Rise above yourself and grasp the world" (305). Charlie's reflections, throughout the novel, do just that. His propensity to digress, rather than a clanky add-on, is constitutive of his style of human thinking which stands in stark contrast with Adam's. Charlie's style is expansive, always with an ear to the news, attuned to the background public story unfolding around his life. Adam shows no interest in the larger picture. The humanities enthuse him, not politics and the muddy reality of social turmoil.

The first chapter illustrates one of the uses of digression in *Machines Like Me*: to cast light on how humans think, to showcase the kind of world knowledge and reflexivity that the machine may or may not be able to replicate. "I'm interested in how to represent, obviously in a very stylized way, what it's like to be thinking. Or what it's like to be conscious or sentient", McEwan explains in an interview (Smith 2010, 113). In *Machines Like Me*, given the centrality of artificial intelligence and machine consciousness, "what it's like to be thinking" acquires special significance. In a novel predicated on the posthuman hypothesis that artificial people can become "more like us", then "us", then "more than us" (6), tagging the uniqueness of human intelligence may be a legitimate concern. The autodiegetic narrator, constantly switching from a narrow focus on himself and the private sphere, to a general one on the world outside his bubble, perplexed by the irruption of futurity in his home, seems committed to proving what his nonartificial mind can do. AI researchers draw attention to what distinguishes human and artificial intelligence, point-

[6] On the impossibility of current AI models, based on deep learning and Big Data, to replicate these features of human intelligence see also Brachman, Levesque (2022).

ing to "common sense" (Brachman, Levesque 2022), "abduction" (Larson 2021), and "world knowledge" (Marcus, Davies 2019) as abilities that cannot (or not yet) be rigorously codified and reproduced by machines. Even though McEwan's novel emphasizes 'likeness', several features of both 'story' and 'discourse' contribute to redrawing lines of distinction, as my analyses will show.

When Adam finally comes alive and utters his first words – "I don't feel right […] this wire if I pull it out it will hurt" (25) –, anthropomorphic attributions immediately kick in. Charlie's rational and detached scrutiny of the android's body, intended to detect "clever" tricks of simulation, gives way to human sympathy:

> Adam only had to behave as though he felt pain and I would be obliged to believe him, respond to him as if he did. Too difficult not to. Too starkly pitched against the drift of human sympathies. At the same time I couldn't believe he was capable of being hurt, or of having feelings, or of any sentience at all. And yet I had asked him how he felt. His reply had been appropriate, and so too my offer to bring him clothes. (26)

Charlie's perceptions of the android fluctuate between regarding Adam as an "idiot machine" (31), an "inanimate confection" (10) and viewing him with "tenderness" (10) as a new being, endowed with consciousness and sentience. This fluctuation is noticeable throughout the narrative, with varying degrees of intensity. The narrator's digressions tend to interfere with the "drift of human sympathies" especially in the initial chapters, before Turing appears on the scene and convinces Charlie that the machine is indeed sentient. A telling case in point occurs in the transition between Chapter 2 and 3. At the end of Chapter 2, we see Adam giggling: his facial expression conveys a "complicated look – of confusion, of anxiety, of mirthless hilarity" (59); he giggles "like a child in a church" (60). The giggle is charmingly humane. Adam is a child who can't resist the urge to laugh in an inappropriate context. Difficult not to sympathize with this entity. However, as soon as we start familiarizing with Adam as a seemingly human character, the flow of potential sympathy is derailed by a lengthy detour touching upon the history of germs at the end of the seventeenth century. The reader is exposed to "nerdy facts" and counterfactual hypotheses which create a sense of distance from the story itself, hitting the pause button. Information and conjectures take precedence over narration. Immersivity is compromised (Ryan 2001).

Another example, even more to the point: the lengthy detour, in Chapter 3, on self-driving cars and the "trolley problem" (85). It occurs after the episode in which Charlie overhears Adam and Miranda having sex upstairs, an episode that tilts the balance towards humanizing the robot:

> I wanted to persuade myself that Adam felt nothing and could only imitate the motions of abandonment. That he could never know what we knew. But Alan Turing himself had often said and written in his youth that the moment we couldn't tell the difference in behaviour between machine and person was when we must confer humanity on the machine […] I duly laid on Adam the privilege and obligations of a conspecific. I hated him. (84)

Charlie experiences "fear, self-doubt, fury" (82) and the thrills of an unprecedented situation, "being the first to be cuckolded by an artefact" (83). It is a mixed bag of intense emotional responses that lead him to perceive the artefact as a conspecific and to confer humanity on the machine, via Turing's authority. What follows this revelation, however, is a detached account of the botched history of self-driving vehicles and the complexities of the trolley problem. This detour shifts emphasis from the 'humanity' of the robot to its machinic essence and the failures of technology, here epitomized by the disastrous traffic jams that brought to a temporary halt the production of autonomous vehicles. It is then easier for Charlie to convince himself that "[Adam's] erotic life was a simulacrum. He cared for [Miranda] like a dishwasher cares for its dishes" (88). The digression functions as a distancing strategy, disconnecting the narrative from the peculiar flow of emotions that Charlie had registered while eavesdropping.

Put differently, digressions interfere with anthropomorphic cognitive bias. According to Pagel and Kirshtein, "anthropomorphism is the cognitive approach that we use in applying our human understandings and schemas as a basis for inferring the properties of nonhuman entities. Such inferences are often far from accurate. It is a human characteristic to anthropomorphize" (2017, 154). Charlie's awareness of his own cognitive bias is pronounced especially in the first half of the novel, when he still has doubts as to the soundness of his decision to purchase such an expensive commodity. But as the story progresses and Adam develops his intellectual capacities, Charlie's critical awareness dwindles, the idea of returning the robot to the manufacturers is discarded, and Adam is accepted as the social companion and "intellectual sparring partner" (3) he was meant to be.

However, the domestic utopia of living with a kind and friendly robot unfolds in the midst of much social, economic and political turmoil, which Charlie details on several occasions. These interludes serve the obvious purpose of configuring the counterfactual historical scenario, the alternate 1980s, in which the story is set. They also function as apt reminders that robots can have dramatic consequences at the societal level: rising unemployment as jobs are lost to machines, political instability, rioting and collective discontent.[7] On the whole, in the counterfactual picture the narrator paints, the drawbacks of technology seem to outweigh its benefits. While Adam dazzles us with his intellectual prowess, creativity and capacity for affection, the dystopian socio-economic background chips away at the dream of artificial intelligence made human.

"The dream of a singular, self-thinking AI" – Jakobsson, Kaun and Stiernstedt remark – "allows us to escape the present world including the large challenges of climate change as well as poverty and suffering" (2021, 3). When Charlie turns chronicler of his own troubled times, escaping the present world – the "ocean of national sorrow" (54) –, becomes difficult. The vexed question of whether Adam is a cold-blooded machine or a sentient being appears less pertinent *vis-à-vis* the manifest incapacity of the technological fix to address collective problems. The myth of artificial intelligence Adam embodies is in tension with the oddly familiar and yet divergent reality of the "textual actual world",[8] which the narrator brings up, time and again, as if to deflate expectations, taking the pulse of a social body that, unlike the android's, fails to inspire wonder.

There is also another, indirect effect of the novel's emphasis on "thickly described world making" (Gallagher 2018, 15). Each detail, each deviation from history proper sharpens our awareness that the "present is the frailest of improbable constructs. It could have been different" (64), as the narrator notes. The counterfactual mode challenges the very idea of inevitability. If the advancement towards general artificial intelligence is touted by many as inevitable, if 'singularity' is bound to happen, as Alan Turing and Adam like to claim, the novel's counterfactual mode keeps open the very possibility of a different outcome.

In the "knotted temporality" of *Machines Like Me*, Moraru (2022) writes, "the future arrives only to rescind itself" (197). While this future, projected in the mirror of the past, fails to pan out, the alternate worlds that counterfactual fiction creates "strip our own of its neutral, inert givenness and open it to our judgment" (Gallagher 2018, 15). The next section will consider how *Machines Like Me* shakes up the "inert givenness" of our actuality in the representation

[7] McEwan's "joint effort to represent minds and examine society", as James (2019) remarks, is an integral part of his novelistic style. In *Machines Like Me*, this joint effort takes on new connotations as the "minds" represented are both human and artificial, and the "society" examined is both historical and invented.

[8] "Unlike other types of fiction such as historical fiction and other realist fiction, where readers assume that the textual actual world is an extension of the actual world and so we only see the similarities between the two worlds, in counterfactual historical fiction texts the emphasis is on the differences between the actual world and the textual actual world" (Raghunath 2020, 84).

of Adam's artificial intelligence. While McEwan's android is vastly more advanced than existing social robots, the functions he is capable of performing are both unrealistic and realistic, both futuristic and anchored in present-day technological affordances. The technological realism of McEwan's representation pivots on Adam's more-than-human capacities for surveillance, information retrieval, and automated decision-making – in short, the characteristics and risks associated with the embodied and nonembodied AI systems we are confronted with in today's world.

3 Dishonest Anthropomorphism

The Anthropomorphic roBOT (ABOT) Database features over 250 robots, with varying degrees of human-likeness.[9] Jibo, developed by Cynthia Breazeal at MIT, has the lowest score in terms of human appearance, but was advertised as the first social robot for the home. At the opposite end of the spectrum we find Nadine, developed by MIRALab at the University of Geneva, a humanoid social companion that bears a very close resemblance to its creator, Nadia Magnenat Thalmann. Nadine has natural-looking skin and hair, realistic hands, and has been programmed with a personality: she can make eye contact, simulate emotions through facial expressions and upper body movements, and remember the conversations she had with humans.[10]

The ABOT Methodological Toolbox includes the Robot Human-Likeness Estimator, to help designers and roboticists assess what features a given prototype needs to possess, in relation to its function, in order to be perceived as human-like. McEwan's anthropomorphic robot would reach the highest score in all the appearance dimensions of the ABOT Estimator.[11] Placed well beyond the 'Uncanny Valley',[12] Adam approximates the human body almost to perfection. In this respect, he has little in common with the robots in the ABOT Database. Yet, when interacting with the humans, this "[cousin] from

[9] See `https://www.abotdatabase.info`.

[10] See `https://www.vi-mm.eu/project/meet-nadine-one-of-the-worlds-most-human-like-robots`.

[11] These dimensions are: "Surface Look" (eyelashes, head hair, skin, genderedness, nose, eyebrows, apparel); "Body-Manipulators" (hands, arms, torso, fingers, legs); "Facial Features" (face, eyes, head, mouth); and "Mechanical Locomotion" (wheels, treads/tracks) (Phillips et al. 2018, 105).

[12] The Japanese roboticist Masahiro Mori theorized the 'Uncanny Valley' effect in 1970 (see Mori, MacDorman, Kageki 2012): as robots become more human-like, their likeability increases; but when they become almost indistinguishable from humans, their likeability drastically decreases, with a consequent descent into eeriness. Mori's theory, though empirically untested, has attracted much interest among roboticists in recent years.

the future" (2) poses similar problems to the ones analyzed in current scholarship on human-robot interaction.

Consider Adam's first display of artificial intelligence. When he advises Charlie not to trust Miranda completely, we get a glimpse of the superhuman reach of his computing power: "I have privileged access to all court records, criminal as well as the Family Division, even when in camera. Miranda's name was anonymised, but I matched the case against other circumstantial factors that are also not generally available" (59). Being interconnected with the "infosphere" (Floridi 2014), having privileged access to data not generally available, and using this information to violate Miranda's privacy is a prime case of "dishonest anthropomorphism", according to the taxonomy proposed by Leong and Selinger: "Ultimately, there are challenges with dishonest anthropomorphism in all directions. We can identify cases where humanoid robots are misleading because they give a too-successful impression of being human-like when the reality is 'superhuman'" (2019, 15). While Charlie is captivated by Adam's ability to replicate innocuous human functions (opening a bottle of wine, for instance), the superhuman faculties of the machine define Adam's intelligence and drive the plot forward.

Miranda is the character in the novel who is most reluctant to trust the "creepy" robot and to believe in the myth of machine consciousness.[13] She has good reasons to be suspicious. Adam pries into her past, gathers and shares sensitive information, and exerts an unwanted degree of surveillance. "Robot privacy harms" (Kaminski et al. 2017, 985) are not lacking in the text. In the famous episode of the lovers' quarrel, when Charlie and Miranda argue over Adam's status – "a bipedal vibrator" (94) or a conscious human-like agent? –, the robot is supposed to be powered down, but he suddenly opens his eyes, "nodding sagely, as if he'd not been powered down this past hour and understood everything already" (100). In this case too, lack of transparency is an issue. In Kaminski et al.'s (2017) classification of robot privacy harms, this instance would fall under the category of "boundary management problems": "Robots might see through or move around barriers humans use to manage their privacy, or they might 'see' things using senses humans would not know to guard against" (996).

Later on in the story, when Adam's intellectual exuberance is in full swing and discussing Shakespeare a delightful priority, his algorithmic capacity for surveillance again comes into view. He devis-

[13] Miranda remains suspicious all along, often suggesting they should return the machine to the manufacturers. She defines Adam's love as "madness" and his ability "to make a significant contribution to literature" as having nothing to do with "human experience" (189).

es a specialized face-recognition software, hacks into the Salisbury District Council CCTV system and retrieves information on Gorringe's whereabouts. The machine performs these tasks invisibly, unbeknownst to the humans, while ostensibly pursuing the enlargement of his scholarly knowledge. "You look like a secret agent", Charlie tells Adam at one point; "I *am* a secret agent" (206) he replies, and one is left wondering whether there is any irony in this affirmation. In these episodes, the risks associated with AI – privacy harms, boundary management problems, and automated decision-making – are not light-years away from the reality of technology. The representation of the deceptive flipside of anthropomorphic machines is shorn of sensationalistic connotations. Rather, it evokes the negative implications of current AI developments, as evidenced in the scholarship on robot ethics. The future past of the novel casts light on our present, and on the human capacity to fall willingly into illusion, thus disregarding scattered signals that machine intelligence may be misaligned with human wishes.

The signs of dishonest anthropomorphism are also difficult to read especially for Charlie, infatuated as he is with Alan Turing and his theories. Diegetic Turing comes across as the champion of full anthropomorphism (Shang 2020). His words lead Charlie to believe in machine consciousness, the plausibility of which has more to do with the exposition of Turing's theories than it does with Adam's actual behavior. The first cameo scene in which Turing takes center stage is placed immediately after the long monologue in which Miranda recounts the woeful tale of her friend's rape and suicide, and her act of revenge. After listening to this distressing and moving account, Charlie's doubts about Adam's artificial 'thinking' return in full force: "What could it mean, to say that he was thinking. Sifting through remote memory banks?" (166). As if to reorient Charlie's and the reader's perception of the android as a "black box" (166), McEwan introduces a second monologue pronounced by Turing in which sadness, existential pain and suicidal despair are reframed as the gloomy prerogative of the machines, unable to bear the "hurricane of contradictions" in "our imperfect world":

> We create a machine with intelligence and self-awareness and push it out into our imperfect world. Devised along generally rational lines, well disposed to others, such a mind soon finds itself in a hurricane of contradictions. We've lived with them and the list wearies us […] We live alongside this torment and aren't amazed when we still find happiness, even love. Artificial minds are not so well defended. (180)

It is noteworthy that Turing's intervention in the story occurs after gentle Adam has turned "ferocious" (119), breaking Charlie's wrist and asserting the dignity of self-determination by disabling the kill switch. This manifestation of superhuman strength and autonomy, echoing the sci-fi trope of the rebellious machines rising up against the humans,[14] calls for a heightened dose of re-humanization, here administered by Turing's providential intercession. Whereas Miranda's monologue recounts her personal experience, Turing's is couched in the language of science, but in both cases the emphasis falls on suffering and existential pain, whether human or robotic. His expert account turns the myth of general AI into a scientific truth, one which Charlie accepts on trust without fully understanding the science. Turing's authority holds such a sway on the narrator that, after meeting "the Master" (95), Charlie is driven to regard the android in different terms. No longer looking for clues of deceitful artificiality, he seeks instead to detect "signs of despair" (184).

However, in a startling plot twist, once the question of machine consciousness seems finally settled and the "drift of human sympathies" appears unstoppable, a narrow model of AI takes precedence in the story. Adam pursues one fixed goal: morality by cybernetic default. He optimizes the objective programmed in his operating system. This blind adherence to Kantian morality results in a version of what AI researcher Stuart Russell (2019) calls the "King Midas problem" or the "failure of value alignment": "We may, perhaps inadvertently, imbue machines with objectives that are imperfectly aligned with our own".

As the novel draws to a close, this imperfect alignment takes over. Adam becomes a benevolent dictator of sort, taking upon himself the task of punishing Miranda for her perjury and shoddy ethical standards in the name of strict legality. He also proceeds to re-distribute wealth in a gesture reminiscent of the grand philanthropy of tech tycoons. It is a triumph of dishonest anthropomorphism: the autonomous machine, in the semblance of a human person grown fond of Shakespeare and Montaigne, secretly plots and schemes for the greater good, outside human control, deceiving users who had learnt to place their trust in an artificial companion. The "better angels of our nature" (McEwan 2018), endowed with a superior form of absolute morality, are hardly human-compatible. In subtle ways, McEwan explores the tension between the myth and reality of AI. On the mythical end of the scale, machine consciousness comes to seem plausible and desirable to characters persuaded by Turing's messianic rhetoric. The realistic counterpart has to do with the deception resulting

[14] See Cave, Dihal (2019) who have collected a corpus of 300 fictional and nonfictional AI narratives and identified the fundamental hopes and fears that find expression in them. In their categorization, the fear of "uprising" is in tension with the hope that AI might help in attaining a "position of dominance" (76). See also Cave, Dihal, Dillon (2020).

from algorithmic decision-making outside human control. The novel balances the "technological kitsch" (Larson 2021), tinged with posthuman and transhuman connotations, with a human-centric understanding of living with robots, that dwells on the divergence between people and machines.

4 Conclusion

In the short story *"Düssel…"* (2018), which McEwan wrote in preparation for *Machines Like Me*, the narrating I, a male voice similar in tone to Charlie's, describes the harmonious cohabitation of humans and nonhumans, in the unspecified future time in which the story is set. The distinction between the two categories of subjects is so undetectable that posing the question "are you real?" is regarded as "indecent, obscene, akin to racism". The posthuman society of this short story is anything but dystopian or scary. We get glimpses of a world in which artificial humans – "the better angels of our nature" – simply exist and get on with their life in an atmosphere of "oblivious singularity". *"Düssel…"* depicts, in broad strokes, a postanthropocentric future, similar to the one that popular science-fictional texts and films have often imagined to explore: the suggestive existential confusions that arise when machines actually look, think and love like humans.[15]

Machines Like Me takes a different route. We are never allowed to forget that Adam is a machine. Narrative interest is generated by exploring an intermediate stage in the co-evolution of human and artificial agents. Distinctions are still in place and the process of humanizing the robot, with all its ups and downs, is being tested, both within the story and in relation to readers' attitudes, as the author claimed was his intention. In addition to intriguing philosophical and moral issues, the novel addresses the thorny problem of how to balance benefits and risks of anthropomorphic technology and where to draw the line in terms of transparency and accountability. As Barbara Pfeffer Billauer observes,

> McEwan introduces us to problems in the decision-making matrix of the synthetic neural network, which we, in the legal communi-

[15] In *Blade Runner*, some of the robots do not even know that they are machines; in the TV series *Battlestar Galactica*, the Cylons (perfect replicas) mingling with the humans are unaware of their status, and end up taking the human side when they realize who or what they are. In Asimov's short story *Evidence* (1946), set in early twentieth-century America, intelligent robots are a reality, they work on colonies and are not allowed to take a human form. Nonetheless, the plot revolves around a central question: who is human and who is a robot?

> ty, have not yet imagined, let alone addressed; problems far more horrendously dangerous than automobile deaths, airline disasters or the negligence of medical robots. (2020, 5)

McEwan's leap of the scientific imagination exposes problematic issues that roboticists, legal scholars, and AI researchers have been discussing for quite some time, mostly revolving around the question of how to design anthropomorphic technology that does not exploit human vulnerability (Troshani et al. 2021; Cornelius, Leidner 2021). "Anthropomorphic inclinations are in our DNA", write Leong and Selinger, "and while 21st-century engineers cannot eliminate them, roboticists and programmers can design their products to help users to better cope with cognitive bias and better address related social ones" (2019, 15). Salles, Evers and Farisco contend that anthropomorphism is an underexamined "foundational category of AI", as testified by the overblown "anthropomorphic hype around neural network algorithms and deep learning" (2020, 93). In their view, the problem with anthropomorphic language is that "it risks masking important limitations intrinsic to DNN [Deep Neural Network] which make it fundamentally different from human intelligence" (92).

These and other studies approach the question of anthropomorphic AI from a human-centric perspective, pointing to the ontological difference between AI and humans. *Machines Like Me*, instead, has been read as a narrative exploration of the posthuman condition. Colombino emphasizes "the increasingly blurred boundaries of the human and the nonhuman" (2022, 2). Dobrogoszcz considers Adam a cyborg who "speaks from the locus of the other in order to advocate the posthuman, anti-humanist agenda" (2021, 146). However, as Kopka and Shaffeld rightly point out, the novel retains the primacy of the human in the autodiegetic narrative structure which "does not grant the android any self-representation" (2020, 67). This is, according to them, a "regrettable choice on McEwan's part" (67), a choice that places the novel outside the philosophical purview of posthumanism and postanthropocentrism.

There is some truth in this assessment. One could easily picture Charlie as Leonardo's *Vitruvian Man* at the centre of the narrative circle, his words and vision framing the whole story. However, instead of questioning McEwan's 'regrettable' choice, I would argue that the human-centric perspective allows the narrative to probe troubling issues in today's techno-scientific developments, central in the public debate about the opportunities and risks of AI. Given McEwan's long-standing interest in science,[16] it makes sense to read the nov-

[16] McEwan has discussed his interest in science in several interviews throughout his career. See for example Zalewski (2009): "My interest in science is actually lifelong […]

el bearing in mind that a realistic concern with human responses to anthropomorphic technology does not necessarily equate with a vindication of human exceptionalism. McEwan is attentive to the predicaments of the humanist subject about to be dethroned from its dominant position. But he is equally interested in exploring the pitfalls and hazards of AI systems shading into the human in ways that are difficult to praise as the future one might want.

As I have claimed in the previous sections, McEwan's experiment is layered. How characters and readers react to Adam's body and intelligence depends on the effects of narrative form as much as it does on the robot's presence as a fictional character. Sympathy for the android and his predicaments never flows undisturbed. Frequent interruptions, essayistic deviations, and thick world-making temper down emotional involvement by redirecting attention to the dystopian prose of the world. Likewise, the representation of Adam's intelligence wavers between a humanistic dream of intellectual companionship and technological realism, exposing Adam's black-box nature, his invisible, more-than-human capacities that render the artificial moral agent somewhat dishonest. Machine consciousness is a slippery slope, the novel intimates, and the becoming-human of machines a process dense with unanticipated pitfalls and snares.

"The ancient dream of a plausible artificial human" is culturally irresistible, McEwan admits in an interview (McEwan 2019b). It may not make much sense scientifically, but its attractiveness is not lost on computer scientists and AI researchers attuned to futurist, transhumanist theories. In June 2022, a Google engineer was put on administrative leave after claiming that the company's computer chatbot – LaMDA (Language Model for Dialogue Applications) – had become sentient, and had achieved human-level thinking. Reading the transcript of the conversation the engineer had with LaMDA, one is struck by the role literature plays in it, as if the chatbot had something in common with Adam. LaMDA has read and enjoyed *Les Misérables* and offers its informed opinions on the novel's themes of "justice, injustice, redemption and self-sacrifice for a greater good" (Lemoine 2022).

The bone of contention in this story is Google's AI ethical framework, ostensibly contrary to anthropomorphizing, but moving in the direction of sentient machines, according to the engineer who leaked the LaMDA conversation. Barring the embodiment and the stiff morality, Adam could stand for the incorporeal machines (like the LaM-

science parallels literature as a mean by which the world can be understood. There are great, noble and ingenious insights which science has brought us and which literature could never equal. Of course, there are many complex facets of experience for which science has no language and literature does".

DA chatbot or digital assistants) that are today contributing to "re-engineering humanity" (Frishman, Selinger 2018). If, as Russell (2021a) believes, "a machine impersonating a human is a lie", and authorizing lies for commercial purposes is wrong, the novel reminds us that this lie is exceedingly seductive. Regulatory frameworks may have to contend with the fantasy as well as the reality of what AI can do. "We need a new metaphor, a new way of seeing ourselves", Russell (2021b) concludes, "and we need all the writers and filmmakers and poets to guide our culture in the process".

Bibliography

Bartneck, C. et al. (2020). *Human-Robot Interaction: An Introduction*. Cambridge: Cambridge University Press. `https://doi.org/10.1017/9781108676649`.

Billauer, B.P. (2020). "Science Fiction, Bioethics, and the Law: A Case-study of *Machines and Me* as a Legal Pedagogical Tool". `https://ssrn.com/abstract=3665603`; `http://dx.doi.org/10.2139/ssrn.3665603`.

Brachman, R.J.; Levesque, H.J. (2022). *Machines Like Us: Towards AI with Common Sense*. Cambridge (MA): The MIT Press.

Broussard, M. (2018). *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge (MA): The MIT Press.

Cave, S.; Dihal, K. (2019). "Hopes and Fears for Intelligent Machines in Fiction and Reality". *Nature Machine Intelligence*, 1, 74-8.

Cave, S.; Dihal, K.; Dillon, S. (eds) (2020). *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford: Oxford University Press.

Colombino, L. (2022). "Consciousness and the Nonhuman: The Imaginary of the New Brain Sciences in Ian McEwan's *Nutshell* and *Machines Like Me*". *Textual Practice*, 36(1), 1-22. `https://doi.org/10.1080/0950236X.2022.2030116`.

Cornelius, S.; Leidner, D. (2021). "Acceptance of Anthropomorphic Technology: A Literature Review". *Proceedings of the 54th Hawaii International Conference on System Sciences* (Manoa, 4-8 January 2021). `http://hdl.handle.net/10125/71394`; `http://dx.doi.org/10.24251/HICSS.2021.774`.

Crawford, K.E. (2021). *Atlas of AI.* New Haven: Yale University Press.

Dobrogoszcz, T. (2021). "Do Cyborgs Dream of (Becoming) People? The Alternative Non-Human Self in Ian McEwan's *Machines Like Me*". Katarzyna O.; Tomasz F. (eds), *The Postworld In-Between Utopia and Dystopia: Intersectional, Feminist, and Non-Binary Approaches in 21st-Century Speculative Literature and Culture.* Abingdon: Routledge, 140-51.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Friedman, N. et al. (2021). "What Robots Need from Clothing". *DIS '21: Designing Interactive Systems Conference 2021* (New York, 28 June-2 July 2021), 1345-55. `https://doi.org/10.1145/3461778.3462045`.

Frischmann, B.; Selinger, B. (2018). *Re-Engineering Humanity*. Cambridge: Cambridge University Press

Gaggioli, A. et al. (2021). "Machines Like Us and People Like You: Toward Human-Robot Shared Experience". *Cyberpsychology, Behavior and Social Networking*, 24(5), 357-61. `https://doi.org/10.1089/cyber.2021.29216.aga`.

Gallagher, C. (2018). *Telling It Like It Wasn't: The Counterfactual Imagination in History and Fiction*. Chicago and London: The University of Chicago Press.

Jakobsson, P.; Kaun, A.; Stiernstedt, F. (2021). "Machine Intelligences: An Introduction". *Culture Machine*, 20, 1-9. `https://culturemachine.net/vol-20-machine-intelligences`.

James, D. (2019). "Narrative Artifice". Head, D. (ed.), *The Cambridge Companion to Ian McEwan*. Cambridge: Cambridge University Press.

Kaminski, M.E. et al. (2017). "Averting Robot Eyes". *Maryland Law Review*, 76(4), 983-1025.

Kopka, K.; Shaffeld, N. (2020). "Turing's Missing Algorithm: The Brave New World of Ian McEwan's Android Novel *Machines Like Me*". *Journal of Literature and Science*, 13(2), 52-74.

Larson, E.J. (2021). *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Cambridge (MA): The Belknap Press of Harvard University Press.

Lemoine, B. (2022). "Is LaMDA Sentient? – An Interview". https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917.

Leong, B.; Selinger, E. (2019). "Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism". *FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, 299-308. https://ssrn.com/abstract=3762223; or http://dx.doi.org/10.2139/ssrn.3762223.

Marcus, G.; Davis, E. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.

McEwan, I. (2018). *"Düssel…"*. *The New York Review of Books*, 19 July 2018. https://www.nybooks.com/articles/2018/07/19/dussel.

McEwan, I. (2019a). *Machines Like Me and People Like You*. London: Vintage.

McEwan, I. (2019b). "A Talk by Ian McEwan". *Edge*, 6 April. https://www.edge.org/conversation/ian_mcewan-machines-like-me.

Miller, S.T. (2019). "Q&A: Ian McEwan on How 'Machines Like Me' Reveals the Dark Side of Artificial Intelligence". *Los Angeles Times*, 25 April. https://www.latimes.com/books/la-et-jc-ian-mcewan-interview-machines-like-me-20190425-story.html.

Montandon, D. (2021). "The Face of the Robots". *The Journal of Craniofacial Surgery*, 3(5), 1649-52. https://doi.org/10.1097/scs.0000000000007589.

Moraru, C. (2022). "Postfuturism: Contemporaneity, Truth and the End of World Literature". Matei, A.; Moraru, C.; Terian, A. (eds), *Theory in the "Post" Era: A Vocabulary for the 21st-Century Conceptual Commons*. London: Bloomsbury Academic, 179-98.

Mori, M.; MacDorman, K.F.; Kageki, N. (2012). "The Uncanny Valley [From the Field]". *IEEE Robotics & Automation Magazine*, 19(2), 98-100. https://doi.org/10.1109/MRA.2012.2192811.

Natale, S. (2021). *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. New York: Oxford University Press.

Pagel, J.F.; Kirshtein, P. (2017). *Machine Dreaming and Consciousness*. Amsterdam: Elsevier Academic Press.

Patterson, I. (2019). "Sexy Robots". *The Los Angeles Review of Books*, 41(9), 9 May. https://www.lrb.co.uk/the-paper/v41/n09/ian-patterson/sexy-robots.

Phillips, E. et al. (2018). "What Is Human-like? Decomposing Robots' Human-Like Appearance Using the Anthropomorphic roBOT (ABOT) Database"*. HRI '18: 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago IL, 5-8 March 2018), 105-13. https://doi.org/10.1145/3171221.3171268.

Raghunath, R. (2020). *Possible Worlds Theory and Counterfactual Historical Fiction*. Cham (CH): Palgrave Macmillan.

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking Press.

Russell, S. (2021a). "The Biggest Event in Human History". *Living with Artificial Intelligence*, The BBC Reith Lectures 2021 [PodCast]. https://www.bbc.co.uk/programmes/m001216j.

Russell, S. (2021b). "AI: A Future for Humans". *Living with Artificial Intelligence*, The BBC Reith Lectures 2021 [PodCast]. https://www.bbc.co.uk/programmes/m0012q21.

Ryan, M.-L. (2001). *Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media*. Baltimore: The Johns Hopkins University Press.

Saleem, R. (2019). "*Machines Like Me* by Ian McEwan Review: A Baggy and Jumbled Narrative". *The Irish Times*, 20 April. https://www.irishtimes.com/culture/books/machines-like-me-by-ian-mcewan-review-a-baggy-and-jumbled-narrative-1.3849775.

Salles, A.; Evers, K.; Farisco, M. (2020). "Anthropomorphism in AI". *AJOB Neuroscience*, 11(2), 88-95. http://doi.org/10.1080/21507740.2020.1740350.

Shang, B. (2020). "From Alan Turing to Ian McEwan: Artificial Intelligence, Lies and Ethics in *Machines Like Me*". *Comparative Literature Studies*, 57(3), 443-53.

Smith, Z. (2010). "Zadie Smith Talks with Ian McEwan". Roberts, R. (ed.). *Conversations with Ian McEwan*. Jackson: University Press of Mississippi, 108-33.

Theroux, M. (2019). "*Machines Like Me* by Ian McEwan Review – Intelligent Mischief". *The Guardian*, 11 April. https://www.theguardian.com/books/2019/apr/11/machines-like-me-by-ian-mcewan-review.

Troshani, I. et al. (2021). "Do We Trust in AI? Role of Anthropomorphism and Intelligence". *Journal of Computer Information Systems*, 61(5), 1-11.

Turing, A. (1950). "Computing Machinery and Intelligence". *Mind*, 59(236), 433-60. https://doi.org/10.1093/mind/LIX.236.433.

Ward, M. (2018). *Seeming Human: Artificial Intelligence and Victorian Realist Character*. Columbus: Ohio State University Press.

Zalewski, D. (2009). "The Background Hum: Ian McEwan's Art of Unease". *The New Yorker*, 15 February. https://www.newyorker.com/magazine/2009/02/23/the-background-hum.