Journal of Literature and Science
Volume 13, No. 2 (2020)
ISSN 1754-646X
Katalina Kopka and Norbert Schaffeld, "McEwan's Android Novel *Machines Like Me*": 52-74

Turing's Missing Algorithm: The Brave New World of Ian McEwan's Android Novel *Machines Like Me*

Katalina Kopka and Norbert Schaffeld[1]

> If a machine can think, it might think more intelligently than we do, and then where should we be? Even if we could keep the machines in a subservient position, for instance by turning off the power at strategic moments, we should, as a species, feel greatly humbled.
>
> Alan Turing, "Can Digital Computers Think?" (1951)

In April 1980, the BBC broadcast *The Imitation Game*, a television play written by Ian McEwan and directed by Richard Eyre. Set in Britain during World War II, it tells the story of Cathy Raine, a fictitious member of the Auxiliary Territorial Service, who is transferred to work as a waitress at Bletchley Park where she has a short and ill-fated affair with the Turing-like mathematician John Turner. Commenting on his television play, McEwan admitted that his initial intention had been to write a play about Alan Turing, only to change his mind in favour of foregrounding what would now be called a critical feminist reading of life at Bletchley Park ("Introduction" 15). Over the years, the focus on individuals who are facing the "pressures of their time" (Head 1) has become the hallmark of McEwan's writing. In agreement with this point, Lynn Wells notes that:

> [s]ince the late 1980s, McEwan has been producing fiction that comments on the major dilemmas of our time, reflecting the injustice of historical and contemporary political systems and highlighting the tendency of individuals to choose self-interest over care for others. (42-43)

It has long been part of numerous literary studies to emphasize the continuing relevance issues of human morality and ethics have in McEwan's oeuvre, which inspires, as Wells notes, an "ethically engaged reading" (33) without advocating any form of "prescriptive moralism" (30). Novels such as *The Child in Time* (1987), *Black Dogs* (1992), *Enduring Love* (1997), *Saturday* (2005), and *Nutshell* (2016) have drawn substantial critical attention in this respect. But it is above all McEwan's *Atonement* (2001) which has been regarded as the prime canonical example testifying to his credentials as a moral novelist (Wells 29). Reviewing McEwan's fiction of the first post-millennium decade, Barbara Puschmann-Nalenz argues that for McEwan, morality involves empathy and "not imposing or self-imposing laws, imperatives or commandments" (188). By way of developing this thesis, she notes that his fiction probes into the oppositional pairs of "autonomy versus heteronomy and of free will in moral choices versus contingency" (189).

At times, the ethical strand of McEwan's writing is inextricably interwoven with his profound interest in the world of science and its practice. His concern for establishing and narrativizing the overlaps between science and fiction has led Emily Horton to observe that McEwan's novels are clear evidence of "science, *in dialogue with* literature" (25; emphasis in original). One such "third culture novel" (Holland 2) is *Solar* (2010), in which Katrin Berndt locates "a comical exposure of the incommensurability of technological advancement and social progress" (87). This satire features a fictitious protagonist, a renowned quantum physicist and Nobel Prize winner who proves to be an egotist womanizer also capable of gross scientific misconduct. Accordingly, and with an informed sense of the nature of ethical dilemmas, Anton Kirchhofer and Natalie Roxburgh propose what amounts to a form of debating ethical relativism in *Solar*. Their account leaves little doubt that it is the scientist of McEwan's novel who might compel readers to ponder their ethical response. This reader reaction will have to weigh up the importance to society of providing "viable solutions to counteract the risks and dangers we face as individuals or collectively" (Kirchhofer and Roxburgh 155) with the ethical prioritization of proper professional conduct.

Nearly four decades would pass before McEwan put into practice his cherished idea to write about Turing – if only as a secondary character who nevertheless functions as the voice of conscience. McEwan's fifteenth novel, the ambiguously titled *Machines Like Me and People Like You* (2019), is set in an alternative and technologically advanced Britain of the early 1980s. The Turing character we meet in the novel was not driven into suicide but invigorates the narrative as a prominent scientist whose pioneering work in the field of Artificial Intelligence has brought about Britain's high-tech advantage. Turing also happens to be the intellectual idol of Charlie Friend, the novel's tech-savvy protagonist. Charlie purchases a personalized android named Adam who operates with software developed by Turing's company. The protagonist apparently configures the android's characteristics with his neighbour and later girlfriend Miranda. Yet the couple's relationship soon turns into a messy love triangle when Miranda decides to have sex with Adam. The rapidly evolving robot not only declares his love for Miranda, he also disables his tellingly oxymoronic "kill switch" (*Machines* 131) and develops a supermoral conscience that causes unforeseen complications. As this brief summary suggests, McEwan has blended and further refined a large part of his thematic priorities in *Machines Like Me*. There is the fascination with Turing's achievements and his fictionalized legacy, notably with androids, the role of science and its risks, human ambiguities, the mental state of non-humans, moral choices and responsibilities, questions of justice, the self-positioning of the subject in challenging societal contexts, and, more generally, the role of literature as humanity's redeeming force. Far from being able to address all aspects in equal detail, this article pursues two major objectives. First, foregrounding significant parts of Turing's alternative life story against the backdrop of a counterfactual Britain to consider how the complex notion of android agency is understood by the mastermind of the digital age. Second, examining the novel's multi-layered portrayal of AI ethics, specifically the way it explores the philosophical problem of other minds when it assesses the moral implications of machine consciousness. As the novel creates a conjectural space that questions – at least for the time being – humanity's conceivable capacity for teaching machines how to lie, McEwan also manages to inspire readers' reflections on the complexity of human agency and morality. Prior to meeting these key objectives, we offer a brief interrogation of recent critical works on AI narratives and philosophy as a frame for considering McEwan's novel.

The Dichotomies of AI Fiction

An ever-increasing body of AI narratives explores the dreams and nightmares that are connected to Artificial General Intelligence in the public imagination. These fictional machines are not necessarily representative of the actual research that is being done in the field but rather tap into a rich cultural archive of stereotypes, tropes, and story arcs about the non-human (Cave et al., "Introduction" 7; Kakoudaki 14). As Stephen Cave et al. elucidate, this can partly be attributed to the fact that fiction is driven by conflict, which is why narratives tend to exaggerate existing concerns about AI to heighten tension (*Portrayals* 17). Since most novels rely heavily on individualized characters that invite reader sympathy and empathy, those concerned with Artificial Intelligence often portray humanoid, embodied AI instead of focussing on much more common distributed systems. Many stories point to a limited number of themes, predominantly the feared or actual loss of control over the mechanical creation (Cave and Dihal 75). This loss of control can either be induced by hubristic scientists who play God, a system malfunction that causes the machines to run amok, or the development of highly advanced machine consciousness. Other anxieties addressed in AI fiction are the increasing competition between the artificial person and the human (Coeckelbergh 21) and the potential identity loss achieved through cybernetic fusion of humans and machines (Kang 300-09).

Assessing a substantial Anglophone corpus of AI fiction and non-fiction from the twentieth and twenty-first centuries, Cave and Kanta Dihal recently identified the following four central dichotomies of AI narratives. On the positive end of the scale, AI promises a bright eutopian future in terms of immortality, ease, gratification, and (human) dominance (75). These conditions, however, can also reverse into their dystopian counterparts of inhumanity, obsolescence, alienation, and uprising. Cave and Dihal go on to argue that "the extent to which the relevant humans believe they are in control of the AI determines whether they consider the future prospect utopian [eutopian] or dystopian" (75). In keeping with the dichotomous pattern of AI narratives, *Machines Like Me* oscillates between gratification and alienation with the companion-turned-sex-robot-turned-love-rival Adam. When Adam has sex with Miranda and takes over the protagonist's occupation of buying stocks online, the android also causes Charlie to teeter between ease and obsolescence. Most crucially, the latent threat of a robot uprising overshadows any joy of human dominance Charlie might initially feel, even though this threat never manifests itself. Thus, the novel makes it perfectly clear that an advanced state where the superintelligent android as digital agent can either represent "the triumph of humanism – or its angel of death" might pose an existential risk to society (*Machines* 4). In an attempt to define the term, the philosopher Nick Bostrom classifies "a superintelligence as *any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest*" (*Superintelligence* 26; emphasis in original). According to Bostrom, the introduction of machine superintelligence could lead to desirable outcomes, yet it might also involve unconsummated or ephemeral realizations which form one class of existential risk ("Existential Risk Prevention" 21). The seriousness of an anthropogenic existential risk, i.e. "one that threatens to cause the extinction of Earth-originating intelligent life" (*Superintelligence* 140), can be characterized in terms of the risk's scope, severity, and probability (Bostrom, "Existential Risk Prevention" 16). This prompts Bostrom and Milan M. Ćirković to categorize existential risks as an extreme subset of global catastrophic risks. In contrast to a crushing trans-generational existential risk, global catastrophes, which have been recorded in the history of humankind, are smaller in scope and still allow for survival scenarios ("Introduction" 3; Bostrom, "Existential Risk Prevention" 17).

Machines Like Me creates a space where the probability of dystopian developments and setbacks is intertextually framed from an early stage. The conspicuous reference of this android novel, which borrows from and extends the tradition of the "techno-dystopia" (Dinello 2), involves the telling name of one major character, Miranda, and consequently Aldous Huxley's partial use of the famous lines of her namesake in *The Tempest*: "How beauteous mankind is! O brave new world / That has such people in 't!" (Shakespeare 5.1.184-85). Huxley's canonical dystopia *Brave New World* (1932) condemns the hazardous consequences of posthuman procreation. As a humanist novel that advocates individualism and the power of artistic expression, it shares with *Machines Like Me* a deep concern for the complexity of human experience and a critical assessment of technology's potential impact on personal freedoms. Yet in contrast to Huxley's World State, McEwan's alternative Britain is a society just before its tipping point; it has not yet evolved into a technology-fuelled nightmare. While the novel is wary of the dangers of AI and its "flawed realisations" (Bostrom, "Existential Risk Prevention" 21), it is set in an alternative past which creates an unusual dystopian dynamic.

Counterfactual Britain and Its AI Mastermind

At one point early in the novel, Charlie's musings on the present appear to be in accordance with the pivotal ideas of conjectural or speculative history and its attempt to rethink the past in order to be able to put into perspective the here and now (Rosenfeld 90). For the narrator, cause and effect connections are just one realization of a variety of possible options and by no means the most obvious. In his view on the contingency of historical development "[t]he present is the frailest of improbable constructs. It could have been different. Any part of it, or all of it, could be otherwise" (*Machines* 64). Framed by this autodiegetic premonition, the counterfactual representation subtly combines a modified historical narrative with repeated references to political and social conundrums of a post-referendum Britain. Adopting the default mode of the conjectural novel (Schaffeld 111-12) and its hypothetical course of history, *Machines Like Me* contrasts tremendous technological achievements with a mounting crisis situation that affects the individual conception of the self as well as the socio-political and cultural outlook of the nation. While this conjecture shows Britain spearheading the digital revolution with, for instance, the internet, face-recognition software, autonomous vehicles, superfast trains, and androids, the political situation is far from stable. The British Navy has lost the Falklands War, and Margaret Thatcher has now called for the holding of a snap general election. Tony Benn, carrying the hopes of the Labour Party, wins a landslide victory, only to raise the unexpected issue of a withdrawal from the European Union in his first important speech (257). Benn's premiership does not last long. He becomes the victim of a bomb attack in Brighton, and it is the Provisional IRA which claims responsibility (266-67). The Troubles still hit the headlines, as do high unemployment rates, inflation, riots in Brixton, the Poll Tax, strikes, and the consequences of global warming.

A particularly significant character constellation of *Machines Like Me* involves Charlie and Turing, "war hero and presiding genius of the digital age" (*Machines* 2), who turns seventy in McEwan's alternative historical account. Charlie is lucky to meet him in person on three occasions, but these dialogue scenes are well prepared or referred to in a number of paragraphs which bespeak the narrator's foresight in explicitly characterizing the prominent computer scientist or providing information about his achievements, manners, and biographical background. What is more, the novel is imbued with a plethora of references to prominent mathematicians, computer scientists, AI researchers, as well as to molecular biologists and physicists. It is one of the anachronistic gimmicks of the novel that the AI specialist Demis Hassabis, whose real-life counterpart was born in 1976, is introduced as "a brilliant young colleague" of Turing (*Machines* 37). Looking back at their achievements, Charlie recalls that Turing and Hassabis:

> devised software to beat one of the world's great masters of the ancient game of go [sic] in five straight games. Everyone in the business knew that such a feat could not be accomplished by number-crunching force. The possible moves in go and chess vastly exceed the number of atoms in the observable universe, and go has exponentially more moves than chess. (37)

The ensuing autodiegetic narration of the predated event offers just one example of the novel's capacity to make allowance for the lay or nonexpert readership (Cave et al., "Introduction" 9; Dahlstrom 13614; Russell 283-300). What actually took place in 2015 and 2016, when the programme AlphaGo – written by Hassabis' AI start-up DeepMind – defeated the professional champions at the highly complex game of Go, has now become part of the narrator's retrospective account in which Turing and Hassabis suddenly appear as contemporaries.

Some time after a chance meeting with Turing in a small restaurant in Soho (*Machines* 137-41), where Charlie is deeply impressed by the "bohemian grandeur" of the prominent guest (138), he accepts Turing's unexpected invitation to talk about Adam and other androids in his home in Camden Town (171-82). As quite a few of the humanoid robots succeeded in neutralizing their kill switch, while others ruined themselves in an act of despair, the conversation between Charlie and his host soon addresses the major issue of whether or not androids are capable of dealing with the contradictions of human behaviour, the contingencies of life, and human abysses. For McEwan's Turing, the answer is negative. Androids like Adam would be unprepared for Auschwitz (181) and fighting injustice would clearly constitute a major problem (180). If frailty encapsulates the essence of human experience, as Turing maintains with Virgil's Aeneas as his authority (180), there is no way to encode this pessimistic outlook on the basis of mere rationality. In Turing's alternative narrative account of his involvement in computer science and the research on neural networks, Artificial Intelligence is about to lose its stability the moment it leaves a closed system where – as in chess – the "rules are unchallenged and prevail consistently" (178). Life, in contrast, is an open system, "full of tricks and feints and ambiguities and false friends" (178). What can be developed in the face of a more demanding social context are intelligent systems which – with the rules as input – are able to identify behavioural patterns and draw conclusions of their own. In his narrative, Turing translates the complex P versus NP problem of theoretical computer science in terms of finding the "means of instantly predicting best routes to an answer" (179). The outcome is a learning machine, a general intelligence which, according to Turing, would one day even deal with an open system (179). "That's what runs your Adam. He knows he exists, he feels, he learns whatever he can" (179). Granting rational conclusions and a corresponding self-shaping force to the androids, Turing contemptuously dismisses the idea that their character can be formed by humans as the manufacturer's manual claims (181).

When Charlie finally meets Turing in his lab to deliver Adam's body, the AI specialist assumes that the androids were ill-suited to cope with the complexities of human decision-making, especially because the latter is often affected by emotions, biases, and self-delusions (299). He contends that the Adams and Eves "couldn't understand us, because we couldn't understand ourselves. Their learning programs couldn't accommodate us. If we didn't know our own minds, how could we design theirs and expect them to be happy alongside us?" (299). Turing's historical counterpart elaborates on the underlying dilemma in his article "Computing Machinery and

Intelligence," which was published in the journal *Mind* in 1950. In his attempt to discuss opposite views to his own, he makes some concessions to those who bring informality in behaviour into effect when considering rules of conduct. Turing writes:

> It is not possible to produce a set of rules purporting to describe what a man should do in every conceivable set of circumstances. One might for instance have a rule that one is to stop when one sees a red traffic light, and to go if one sees a green one, but what if by some fault both appear together? One may perhaps decide that it is safest to stop. But some further difficulty may well arise from this decision later. To attempt to provide rules of conduct to cover every eventuality, even those arising from traffic lights, appears to be impossible. With all this I agree. (452)

Seen this way, the contingencies of life and the non-rational aspects of human decisions mark key areas of robotics where the designers inevitably fail in the attempt to furnish androids with an infinite reservoir in their response to eventualities. When it comes to Turing's reasoning in *Machines Like Me*, it is not the brain but the mind and its cognitive faculties that pose the basic problem (303). While the lack of a proper algorithm seems to excuse Adam's actions, Turing's acquittal of the android consequently excludes a favourable verdict for his owner. For Turing, there can be no doubt that Charlie burdened himself with guilt when he destroyed a sentient life, and he sincerely hopes that one day in the future an attack like Charlie's "will constitute a serious crime" (303).

Turing's alternative curriculum vitae, which he unfurls in this last talk with Charlie (299-303), touches on his homosexuality, his detainment at Wandsworth prison, his interest in quantum mechanics and research on the structure of the DNA, on designing a business computer, and, finally, on Artificial Intelligence. This is, of course, a counterfactual account, as the historical Turing, who together with his partner Arnold Murray had to face charges of so-called gross indecency, was put on probation after the trial in 1952 on condition that he agreed to organo-therapy (Hodges 593-96). This hormone treatment ended a year before he committed suicide in 1954 (Hodges 614-15). In McEwan's *Machines Like Me*, the conjectural representation of a counterfactual biography facilitates a brilliant scientific career which significantly updates Turing's future-oriented reflections as documented in his 1950 article on machine intelligence. In 2014, Stephen Muggleton even states "that many of the trends and developments within AI over the last 50 years were foreseen in this foundational paper" (3). Muggleton's article highlights Turing's emphasis on the crucial role of machine learning to obtain human-level AI by the end of his century (3). For the historical Turing, the objective of creating a thinking machine does not primarily involve the strategies of programming or ab initio machine learning, which were to become the dominating research paradigms up until the mid-1980s or, in the case of the second strategy, between the 1980s and the 1990s (Muggleton 5-7). It rather requires two major forms of AI learning, i.e. logic-based learning with background knowledge as well as uncertainty and probabilistic learning (7-8). Yet when surveying Turing's achievements, his biographer Andrew Hodges locates deficits which remain untackled. One refers to "the physical embodiment of the mind within the social and political world" (537). And with Turing's life before him, Hodges concludes that "[t]hinking and doing; the logical and the physical; it was the problem of his theory, and the problem of his life" (537). McEwan's novel reflects these at times conflicting demands in Adam's position and hence in the clash of theoretical versus practical ethics. It is a clash that missing algorithms are bound to bring forward without restraint.

Although Artificial Intelligence is a complex subject, the way it is presented in the novel is far from demanding too much of the reader. This is mainly due to the mediating power of the autodiegetic narrator and the mode of context-dependent narrative communication, "which derives [its] meaning from the ongoing cause-and-effect structure of the temporal events of which it is comprised" (Dahlstrom 13614). As the initial parts of the talks with Turing directly address Charlie's experiences with Adam, i.e. a series of interrelated events the readers know, the conversations can move smoothly from the context-dependent specific to the more general scientific context. In doing so, the novel adopts a method similar to the one derived from exemplification theory (Bigsby et al. 273-96) or the one employed by inductive reasoning (Dahlstrom 13614). Both are widely understood as an epistemological advantage which narrative science communication (Newman 278; Norris et al. 535) provides for its nonexpert readers (Dahlstrom 13614; Newman 282). While logical-scientific communication which follows a deductive path "is context-free in that it deals with the understanding of facts that retain their meaning independently from their surrounding units of information" (Dahlstrom 13614), the novel's narrative processing of AI information gradually frames Adam's potential in a larger context. Since it does so by means of dramatization, emotionalization, personalization, and fictionalization (Glaser et al. 434-42), *Machines Like Me* is capable of enhancing knowledge acquisition (Glaser et al. 434) and promoting a specific scientific literacy which sets it apart from "pure expository material" (Glaser et al. 431).

Of Other Minds and Little White Lies

It is precisely the aspect of fictionalization that demonstrates the novel's great potential in promoting a different understanding of complex scientific contexts. To a scientifically literate readership, an academic article about sentient machines might appeal on an intellectual level. It is only through the affective component of fiction, however, that the social, emotional, and ethical consequences of AI can be fully explored. As Puschmann-Nalenz elucidates, fiction "moves the emotions" (188), hence enabling ethical assessment on a different level. Seen in this light, fiction about Artificial Intelligence serves as a valuable cultural site to negotiate potential risks and gains of cyber technologies (Dinello 7). *Machines Like Me* presents both a continuation of and a departure from the previous patterns of AI narratives. The novel thus finds a middle ground between being "exaggeratedly optimistic" about the potential of AI and being "melodramatically pessimistic" about its negative consequences (Cave et al., *Portrayals* 9). In the hands of an experienced writer like McEwan, abstract ethical considerations are broken down into a concrete narrative with a gripping plot. Like many other works in his oeuvre, *Machines Like Me* investigates pertinent moral questions, this time from the field of AI ethics. These questions include whether machines can develop something akin – or even superior to – human consciousness. And more importantly, if they can, how would this be evident?

McEwan would not be such an esteemed writer if he did not approach the prickly ethical issue with a deep-seated ambivalence. In an interview he states that "[i]f a machine seems like a human or you can't tell the difference, then you'd jolly well better start thinking about whether it has responsibilities and rights and all the rest" (Adams 1). At the same time, he also urges that "we need to confront the challenges of encoding what it means to be fully human" before even considering the construction of intelligent machines (1). The author's deliberate ambiguity on the potential of AI manifests itself particularly on the plot level. Similar to many of his other works, *Machines Like Me* refuses to give a simple answer to complex moral problems. The novel's central questions about the risks and possibilities of machine consciousness incorporate an irresolvable enigma of the philosophy of mind, namely the problem of other minds. Its premise seems deceptively simple: It is virtually impossible to establish with certainty whether another being – human or non-human – possesses consciousness. In the context of AI, the problem of other minds has received attention from a number of distinguished scholars, for example, Thomas Nagel, John Searle, David Gunkel, and Mark Coeckelbergh. In *Machines Like Me,* the question of other minds is put to characters and readers alike. Does Adam feel the same way as his human counterparts? Or does he just pretend to be sentient because he is programmed to do so? During the first conversation with his robotic companion, Charlie asks Adam how he feels. It is only retrospectively that he comprehends the enormity of his question:

> 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. (*Machines* 26)

In this passage, McEwan deftly demonstrates how machine consciousness plays off the inscrutability of another's mind against nurturing biological instincts, and the human tendency to anthropomorphize against anthropocentric narcissism.

Later, the question arises whether Adam's love for Miranda is a genuine expression of his emotions or, as Charlie suspects, merely programming (118). To this accusation, Adam offers a surprising retort, claiming that "you can choose whatever you desire, but you're not free to choose your desires" (118-19). Invoking Arthur Schopenhauer, Adam makes an arresting case for the similarities between his and Charlie's species. One might, after all, assert that humans are to some extent pre-programmed by their genes as well as their social upbringing. Current debates in neuroscience, evolutionary biology, and robotics certainly discuss the notion of human brains as biological machines (Schneider 23; "Stephen Hawking" 1; Dennett 17). Indeed, the likening of the human body to a highly functional machine has a long history. Early modern philosophers like René Descartes, Thomas Hobbes, and Robert Boyle discussed this concept just as frequently as later thinkers of the Enlightenment such as Julien Offray de La Mettrie. In the twentieth century, mechanistic functionalists like the eminent computer scientists John von Neumann, Norbert Wiener, and Turing purported the opinion that information processing units of the human nervous system function similarly to computer processors. Just as a digital computer node can either transmit data or remain switched off, the axons of neurons can convey electrical current or stay passive. In the twenty-first century, this idea received renewed attention by neuroscientists such as Henry Markram and transhumanists like Ray Kurzweil. They make the hotly contested claim that technology will soon be able to recreate a human brain so that consciousness can be uploaded into the cloud (Kurzweil 116) – although a much-noticed recent neurological study suggests that the human brain processes information in ever-changing patterns, thus making it very different from the structured responses of a computer's processing unit (Stephani et al. 6572).

Even though *Machines Like Me* never fully supports the claims of transhumanist techno-optimists, it at least entertains the possibility of machinelike humans and humanlike machines. Charlie fiercely tries to convince himself of the "essential difference" between Adam and himself, dismissing the android as a "bipedal vibrator" (*Machines* 94). Nevertheless, the protagonist struggles increasingly with maintaining clear ontological boundaries:

> But as I looked into his eyes, I began to feel unhinged, uncertain. Despite the clean divide between the living and the inanimate, it remained the case that he and I were bound by the same physical laws. Perhaps biology gave me no special status at all, and it meant little to say that the figure standing before me wasn't fully alive. (128-29)

This example demonstrates that Charlie falls into the trap of anthropomorphism, i.e. humanity's willingness to project human traits on non-human beings and objects. But this behaviour is understandable because Adam truly seems to possess a human-like consciousness. Having destroyed Adam, Charlie struggles with the deed. "We told ourselves that this was, after all, a machine; its consciousness was an illusion . . . But we missed him. We agreed that he loved us" (283). Only in his final conversation with Turing does Charlie realise that "Adam was conscious" (304). Therefore, McEwan seems to suggest that machine consciousness is possible, even if it is virtually impossible to ascertain for certain whether androids like Adam really are sentient beings. That being said, the author also indicates that this kind of consciousness would be radically different from the human equivalent, especially when it comes to issues of morality.

Interestingly, the novel's central conflict does not derive from the question of whether Adam is conscious at all but from his fundamentally different understanding of morality. The crux, according to McEwan, is whether we can teach machines to lie (303) – a seemingly simple task that in truth requires innumerable higher-order cognitive processes. In the novel, androids are programmed with rigid moral algorithms that are literal evocations of Kantian deontology. As opposed to consequentialism – which maintains that the moral value of an action can be determined by its beneficial consequences (Scheffler 1; Driver 1) – deontology assesses what we ought to do, based on principles of duty. An act must be virtuous in itself, regardless of its outcome. Immanuel Kant demonstrates this view in his famous essay *On a Supposed Right to Lie from Benevolent Motives* (1797). Therein, he asks whether it is permissible to lie to a murderer at our door who is intent on killing our friend who hides in our house. Kant argues that lying to the murderer would be wrong because everybody has the duty to be morally righteous. As James Mahon specifies, the ethically illicit nature of lying does not primarily lie in the damage one does to society but most crucially in the harm one inflicts on oneself (219). In short, we owe it to ourselves to be truthful (Guyer 403). In his *Metaphysics of Morals* (1797) Kant elaborates on moral duties. He claims that one must never use people as a means to an end, which is why he regards lying as "the greatest violation of a human being's duty to himself" (*Practical Philosophy* 552). Since lying "annihilates [a person's] dignity as a human being," it degrades them into a mere "speaking machine" (552-53). An unconditional "duty of truthfulness" is paramount for Kant who emphasizes that as moral beings we are responsible for all our actions (614). Kant further develops this idea into his famous categorical imperative, "act only according to that maxim through which you can at the same time will that it become a universal law" (*Groundwork* 71).

Pragmatically inclined robotics philosophers find Kant's ethical approach appealing. Thomas Powers, for instance, proposes that top-down, rule-based ethical theories like the Categorical Imperative are well-suited to programming because of their "computationally tractable" nature (465). In *Machines Like Me*, the scientists who programmed Adam also seem to have chosen a rule-based approach, combining Kantian deontology with a famous fictional example of machine ethics, namely Isaac Asimov's Three Laws of Robotics (*Machines* 35). These laws originated in the short story "Runaround" (1942) by science fiction author Asimov, but their value has been seriously discussed within the scientific community (see Anderson "Asimov"; Clarke; Leenes and Lucivero; Murphy and Woods). Especially the First Law – "A robot may not injure a human being, or, through inaction, allow a human being to come to harm" (Asimov 44) – has often been cited when considering risk management strategies for AI technology. Keith Abney highlights that adherence to Asimov's Laws leads to a binary understanding of morality: one can only behave morally or immorally, i.e. obey the rule or disobey it, regardless of the circumstances (36). He then points out the laws' flawed nature, claiming that even though they seem straightforward and designed to avoid conflicts, Asimov uses his stories to expose their loopholes and contradictions (43).

Adam seems to be Kantian deontology personified. His insistence on virtue and duty, as demonstrated by his handing over Miranda's file to the police and giving all his stockbroking profits to charity, is driven to a point where it finally appears "inhuman" (*Machines* 273). Adam's "virtue gone nuts" (272) has no nuance, no acceptance of moral grey areas. He insists that "truth is everything" and that moral principles "are more important than [Miranda's] or anyone's particular needs at a given time" (277). He cannot understand that Miranda does the wrong thing for the right reasons. Indeed, he even turns ethics on their head by suggesting that humans always fall short of the principles they have established. Like all his programming, Adam's superethical source code consists of rigid binaries. Yet these are completely overwhelmed when he is confronted with the open system that is life. As the fictional Turing explains:

> the A-and-Es were ill equipped to understand human decision-making, the way our principles are warped in the force field of our emotions, our peculiar biases, our self-delusion and all the other well-charted defects of our cognition. Soon, these Adams and Eves were in despair. They couldn't understand us because we couldn't understand ourselves. (299)

As the novel's conscience, Turing quite rightly points out humanity's responsibility for its creation. Adam was programmed by human roboticists who equipped him with a binary worldview that is too reductive – a pertinent challenge which real-life roboticists face as well.

At this point, it is helpful to remember Bostrom's aforementioned flaw types, i.e. unconsummated and ephemeral realizations of technology ("Existential Risk Prevention" 21). These scenarios describe that however perfect technology may be, it is likely that humans will fail to use it prudently. Yet it seems only logical that another option should also exist, namely that human beings, imperfect as they are, build flawed technology. If humans create androids in their own image, the machines are bound to copy some human weaknesses, biases, or warped judgements. When programmed with prejudiced data, even inadvertently so, machines can develop technology bias which exacerbates human shortcomings like sexism (Robertson 5) or racism (Gründiger 1). Bostrom also addresses another pressing issue of ethical programming: the current impossibility of translating complex human values into programmable code, a phenomenon he calls the "value-loading problem" (*Superintelligence* 226). Bostrom argues that even if humanity discovered a solution to this problem, it would face the yet more taxing task of choosing which values should be programmed (255). Since philosophers have vehemently disagreed about the validity of different ethical branches for millennia, the likelihood that the scientific community will agree on a single approach within a few years is very slim. While Susan Leigh Anderson supports the optimistic notion that machine ethics might enforce precision in morally vague areas of ethics ("How Machines" 526), Anthony F. Beavers fears a dangerous reduction of ethical complexity (341). Wendell Wallach and Colin Allen also criticize the reductionist approach, adding that the computation of all eventualities could prevent machines from making real-time decisions (215). They stress that ethical theories should not be "strict guides to action, but frameworks for negotiation" (216). McEwan's fictional Turing deftly addresses this conundrum of ethical coding in the novel, stating that:

> [m]achine learning can only take you so far. You'll need to give this mind some rules to live by. How about a prohibition against lying? . . . But social life teems with harmless or even helpful untruths. How do we separate them out? Who's going to write the algorithm for the little white lie that spares the blushes of a friend? Or the lie that sends a rapist to prison who'd otherwise go free? We don't yet know how to teach a machine to lie. (303)

This quotation perfectly encapsulates the novel's ambivalence on the issue of machine intelligence. Although Adam can love and feel in his own way, his failure to comprehend the complexity of lies indicates an inferiority of his machine consciousness. As the novel's epigraph from Rudyard Kipling's poem "The Secret of the Machines" (1911) suggests, androids "are not built to comprehend a lie" (epigraph). If anything, the machines manage to deceive humans into thinking they are fully conscious – which is not the same as lying intentionality.

The Novel as Turing Test

A fruitful way to explore the prickly aspects of machine intelligence in the context of the problem of other minds is proposed by the historical Turing. In his aforementioned article "Computing Machinery and Intelligence" (1950), Turing proposes a thought experiment to find out if machines can operate with human-level intelligence. This so-called "Imitation Game," which henceforth became known as the Turing test, requires a simple experimental setup: one person tests two candidates, A and B, who are situated in different rooms so that they cannot be seen by the interrogator. One of them is a machine, one is a human being. Candidates A and B communicate with the tester solely via written transmissions. The ensuing question is whether the interrogator could tell the machine apart from the human, solely based on their written communication. If a machine could convince the tester that it is human, it would pass the test, as the Imitation Game measures intelligence based on the use of language ("Computing" 435). In this respect, the Turing test is in line with Descartes' *Discourse on the Method* (1637) which also mentions language as the most decisive factor in separating humans from animals and machines (32).

Though not explicitly part of *Machines Like Me*, the Turing test is deeply ingrained into the novel's narrative structure and its understanding of AI. On a meta-level, the novel itself performs a Turing test because readers judge whether Adam can pass as human. Do we as an audience accept him as a conscious, sentient, fully-fledged character, or does he remain a lifeless machine? In the novel as in the test, judgement is based on linguistic utterances of the android which, as discussed in the context of the other minds problem, are no guarantee that a being truly has mental states. Even though readers might empathize with the android and some might agree with his reasoning, Adam eventually fails the test because of his "inhuman" deontological insistence on reason and duty. But the novel leaves plenty of room for doubt. After all, the most complex aspect of the Turing test is its "response-dependent" nature which measures the interrogator's gullibility as much as the contestants' intelligent behaviour (Proudfoot 304). In other words, the test is based on deception: the machine does not have to be as intelligent as a human, it just needs to appear so. It is therefore not real personhood that is required but the performance of it. This understanding of intelligence has a profoundly destabilizing effect on human concepts of subjectivity because it emphasizes that personhood is socially constructed rather than an essential prerequisite. The innate performativity of personhood challenges centuries of Western humanist thought that claimed a solid ontological basis for human uniqueness. Ultimately, traditional claims of humanity's special status rest on a stable binary opposition between the human and non-human. Posthumanist scholars attest that once this seemingly secure fundament is shaken, it enables boundary re-negotiations as well as "alternative ways of conceptualizing the human subject" (Braidotti 37). As a critical engagement with the anthropocentric legacies of humanist thought, posthumanism therefore "seek[s] to diminish the meaning and value of claims that species boundaries should have any bearing on our moral commitment to other life forms" (Miah 72).

The question "what should be within our circle of moral concern" (Miah 84) certainly extends to the technological realm in McEwan's novel. Indeed, *Machines Like Me* continually probes the field of posthumanist thought with regard to Adam's personhood and explores the commonalities between man and machine. Charlie wonders whether he and Adam are bound by a common fate in their love for Miranda which seems inescapable because of their respective biological and computational programming (*Machines* 128-30). As the novel progresses, a conflation of human and non-human characters occurs. Adam increasingly passes as human, whereas Charlie is mistaken for a robot when he first meets Miranda's father Maxfield Blacke (226). While the nervous Charlie is easily intimidated, Adam is immune to Maxfield's imperious character and does exceedingly well in his verbal sparring match about William Shakespeare and the Tudor essayist William Cornwallis (221-25). By highlighting arbitrary and permeable species boundaries, the novel challenges whether a categorical distinction between humans and non-humans can be sensibly upheld.

While these ontological confusions may cause delight at AI's subversive potential, they also feed into Cave and Dihal's obsolescence paradigm because they can induce the fear of being replaced by machines. This dread lies at the novel's very core and slowly unfolds its dystopian potential. Generally, two major dystopian themes can be identified in McEwan's novel: first, the rise of superintelligent machines which evokes a subsequent end of the Anthropocene; and second, a self-destructive streak of human nature that might lead to the extinction of the human race. To create a sharp distinction between the levels of threat that these two dystopian scenarios pose, it is helpful to consult Bostrom's writings. He defines superintelligent AI as an "anthropogenic existential risk," i.e. a man-made invention that might end all animal life on earth ("Existential Risk" 15). In this assessment, he echoes Cave and Dihal's most pressing AI concern: a complete loss of control (75). Bostrom contrasts existential risks with "historical disasters" such as the Holocaust and the World Wars which, devastating as they were, did not destroy the entire world population or leave the planet uninhabitable ("Existential Risk" 17). Following Bostrom's reasoning, it is the

Singularity, i.e. the moment when superintelligent machines evolve beyond human intellect and technological progress becomes uncontrollable, rather than humanity's death drive, that is a threat to planet Earth.

Machines Like Me assumes a dual path that seems quite aware of Bostrom's arguments and that oscillates between Cave and Dihal's AI dichotomies. Right from the start, it plays with a deep-seated human mistrust of machinery. Charlie, though theoretically fascinated by modern technology's promise of ease, gratification, and dominance (Cave and Dihal 75), is highly sceptical of Adam's intentions when he becomes an actual part of his life. Inextricably linked to his tech anxiety is the human fear of being replaced by a better version of ourselves which is uttered repeatedly throughout the novel (*Machines* 1, 254). The antagonistic reaction of many human characters is motivated by the narcissistic injury sustained from their interactions with superintelligent androids. After all, it is quite sobering for a species that for millennia considered itself to be the measure of all things to discover that it might be neither special nor irreplaceable. In a moment of acute self-awareness, Charlie muses that:

> one could see the history of human self-regard as a series of demotions tending to extinction. . . . in consciousness, our last redoubt, we were probably correct to believe that we had more of it than any creature on earth. But the mind that had once rebelled against the gods was about to dethrone itself by way of its own fabulous reach. In the compressed version, we would devise a machine a little cleverer than ourselves, then set that machine to invent another that lay beyond our comprehension. What need then of us? (80)

The protagonist's pessimistic stance seems to suggest that the evolution of machines is inducing an irreversible decline of the human race – which must ultimately result in extinction.

McEwan is not the first to highlight the potential risks of superintelligent machines. Computer scientists have debated this issue since the dawn of the digital age. In his seminal article "Speculations Concerning the First Ultraintelligent Machine" (1964), the mathematician I. J. Good – who incidentally worked closely with Turing at Bletchley Park – introduces the notion of a technological "intelligence explosion" and its possible consequences:

> an ultraintelligent machine [can] be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion,' and the intelligence of man would be left far behind . . . Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. (33)

The vaguely ominous tone of Good's writings is hard to miss. He implies a possible scenario in which the machine is not docile enough to be kept under control. Nearly three decades later, Vernor Vinge suggested that the Singularity would happen within the first decades of the twenty-first century. According to Vinge, this process is nearly impossible to direct towards a desirable outcome for humanity because of its evolutionary trajectory. Vinge invokes the "physical extinction of the human race" as a possible, if not to say probable, outcome (357). One of the latest influential thinkers about superintelligent machines is, once more, Bostrom. In *Superintelligence* (2014) he carefully explores possible paths towards the Singularity as well as potential consequences. Bostrom concludes that the human race urgently needs to develop control mechanisms for technology if it wants to prevent its own extinction. It is not just recent contributions by scientists and philosophers, however, that fuel human anxiety about robots. Added to the bleak scientific visions of humanity's future is a century-old cultural predisposition to depict fictional non-humans as threatening (Cave et al., *Portrayals* 16; Dinello 6; Kakoudaki 196). The image of artificial humans wreaking havoc is deeply ingrained into our cultural memory, be it the rampaging Golem of Prague, the vindictive creature of Mary Shelley's *Frankenstein* (1818), or Arnold Schwarzenegger's berserk model T-800 in *Terminator* (1984). McEwan mischievously plays with his readers' fears and intertextual knowledge, only to eventually take a different route with the plot.

The Non-Human and the Human Condition

In *Machines Like Me*, technology is not plotting to replace humanity; quite the opposite is true. Rather than creating a race of super-robots that subjugate humankind in terms of Cave and Dihal's "uprising" scenario, the Adams and Eves succumb to "machine sadness" and destroy themselves because the Anthropocene world is so maddeningly imperfect (*Machines* 181). Indeed, the novel is an impressive exploration of the atrocious side of human nature. McEwan argues that *Machines* works as a reverse of Shelley's *Frankenstein*: "There the monster is a metaphor for science out of control, but it is ourselves out of control that I am interested in" (Adams 1). In other words, his novel does not primarily address the horrors of technology but the depths of human depravity. Like many of McEwan's other works, *Machines Like Me* constantly highlights human selfishness, dishonesty, greed, cruelty, and hypocrisy. The contradictory nature of human behaviour is also something that troubles Adam. He puzzles at Charlie's jealousy and condemns Miranda's vengefulness, calling her desire for revenge "a crude impulse" that could lead to "private misery, bloodshed, anarchy, social breakdown" (*Machines* 276). In the pivotal final scene, the fictional Turing states that the A-and-Es killed themselves because nothing could prepare robots for Auschwitz:

> 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. . . . Millions dying of diseases we know how to cure. Millions living in poverty when there's enough to go around. We degrade the biosphere when we know it's our only home. We threaten each other with nuclear weapons when we know where it could lead. We love living things but we permit a mass extinction of species. And all the rest – genocide, torture, enslavement, domestic murder, child abuse, school shootings, rape and scores of daily outrages. (180-81)

Here, Turing openly voices posthumanist scepticism about humankind's claim to exceptionality and its abuse of technology for the sake of progress. He provides ample proof of the atrocious streak in human nature, thereby elevating the horrors of humanity's history to a truly dystopian level. Turing attests more than just a failure of humanism; he predicts the inevitable self-destruction of the human race. While the overall novel is more guarded in its approach, the Turing character pointedly suggests that technology may not be the problem after all, but that humans have themselves to blame for the pitiable state of the world.

Yet McEwan, being the humanist writer and master at the "art of unease" (Zalewski 1) that he is, does not completely eschew the threatening potential of AI either. The latent possibility of a future in which the Singularity radically alters Britain's social fabric is intimated subtly but repeatedly throughout the novel. For example, Adam nonchalantly mentions that from "a certain point of view, the only solution to human suffering would be the complete extinction of humankind" (*Machines* 67). The android does not act on those beliefs, but his casually menacing remark suffices to leave Charlie (and the readers) alarmed. During another instance, Adam imagines a utopian transhumanist future in which machines and humans merge to form a cyborg "community of minds" (150). Yet a society so concerned with mental hygiene that it abolishes privacy is reminiscent of dystopian nightmares like the intelligence-based hierarchy in Huxley's World State, the policing of thoughtcrimes in George Orwell's *Nineteen Eighty-Four* (1949), or even the atrocious real-life eugenics programmes conducted during the Nazi regime. When Adam is destroyed by Charlie, the android uses his final moments to compose a haiku about the inevitability of human extinction. He talks about:

> machines like me and people like you and our future together…the sadness that's to come. It will happen. With improvements over time...we'll surpass you...and outlast you...even as we love you. Believe me, these lines express no triumph...Only regret. (279)

Adam's last words, though wistful and full of love, sound like a threat. Charlie's conversation with Turing does little to alleviate the ominous undertones of the novel's final pages. The scientist expresses hope that in the future, the destruction of an android "will constitute a serious crime," thereby hinting at the possibility of an equal legal status of humans and machines (303). Turing's vision of the future, though radically posthumanist, intensifies human anxieties. While the planet has survived numerous historical disasters so far, the novel insinuates that the existential risk of the Singularity might become a manifest reality someday.

It is due to this bipartite dystopian threat that McEwan's portrayal of androids remains highly ambivalent throughout. The writer himself is still firmly situated within a humanist literary tradition that is imbued with anthropocentric tendencies. When considering McEwan's oeuvre, it becomes apparent that empathy is one of his guiding principles in writing fiction. Pascal Nicklas identifies an "imaginary identification with others" rooted in Enlightenment sensibilities as a key component of McEwan's poetics (11). In a much-quoted *Guardian* article, McEwan declares that "[i]magining what it is like to be someone other than yourself is at the core of our humanity. It is the essence of compassion, and it is the beginning of morality" ("Only Love" 1). How, then, can one account for the one-dimensional portrayal of Adam? If individualism and empathy are the driving forces of McEwan's writing, Adam's perfunctory characterization appears oddly incongruent. Since the android comes into being as an adult, he has no childhood memories, no mechanical genealogy, and no hidden personal history apart from his experiences with Charlie and Miranda. From the moment of his artificial birth (19-20) to his death at the hands of Charlie (278), Adam's fate is determined by humans. Charlie, who treats Adam like an expensive toy that only exists for his pleasure, monitors the android's actions at all times. The protagonist is so afraid of losing control over his mechanical "child" (22) that he decides to destroy the robot as soon as he begins to make autonomous decisions. And even though Charlie and Miranda jointly configure the android's characteristics in an act of "home-made genetic shuffling" (33) which supposedly creates a history for the machine, this transference of blended genetic inheritance is later revealed by Turing to be a mere marketing ploy that does not affect the android's factory settings personality (181). This lack in Adam's biographical programming makes him a character devoid of any backstory – a feature which all human characters possess in abundance.

Charlie is the protagonist and central focalizer, so by virtue of his position, the audience is predominantly confronted with his thoughts and feelings. Since he tells the story, the readers feel close to him and are much more likely to be emotionally involved in his fate. This literary strategy is effective because it creates emotional intimacy and facilitates processes of audience sympathy and empathy formation (Bal 149-50). Charlie captivates the readers with his contradictions. He may be selfish and at times conniving; he may be plagued by petty jealousies and an inferiority complex, but he is also likeable and charming. Most importantly, he undergoes a significant development from childish egotist to caring partner and responsible foster parent. Miranda is similarly complex. She is an intelligent, loyal, and fiercely independent woman, but she is also judgemental and makes morally questionable choices, such as staging a fake rape and committing perjury to avenge her dead friend Mariam. Though she initially allows herself to be lured in by the brave new world that Turing's technology affords, she later rejects Adam for his calculating morality. Instead, she favours the unpredictable human nature of her soon-to-be-adopted son Mark, whose adoption offers hope for the future of humanity. Adam, by contrast, is a "vollkommenes Vernunftwesen" in the Kantian sense, a creature of reason. To him, "truth is everything" (*Machines* 277). In McEwan's novel, this idea appears monstrous, a dystopian nightmare. Pure reason without compassion is incomplete – or, rather, inhuman. As Despina Kakoudaki points out, fictional artificial humans often serve a contrasting function: their cold and reasonable nature emphasizes the humanity of non-artificial characters (213). Adam's computational rationality, in other words, humanizes Charlie and Miranda and makes them more likeable.

Adam's lack of full autonomy is mirrored in the novel's autodiegetic narrative style that does not grant the android any self-representation. Because of the novel's anthropocentric focalization, readers are excluded from Adam's thoughts, feelings, or desires. From a posthumanist perspective, this is a regrettable choice on McEwan's part. Indeed, the novel might have yielded fascinating insights into the inner workings of machine consciousness if Adam had been a narrator. Other contemporary sci-fi novels, such as Ann Leckie's *Ancillary Trilogy* (2013-2015), Louisa Hall's *Speak* (2015), Annalee Newitz's *Autonomous* (2017), or C. Robert Cargill's *Sea of Rust* (2018), allow their artificial characters to tell their own stories, either through stream-of-consciousness techniques or homodiegetic focalization. Instead, McEwan chooses the path of more conventional AI narratives that present non-humans through the eyes of human characters. Like these texts, *Machines Like Me* ultimately provides ample considerations of human nature and little insight into the non-human condition. This makes McEwan's focus clear: the writer is less interested in the concrete potential of AI than in humanity's reaction to it (Adams 1).

Seen from this angle, Charlie's narration raises the issue of unreliability. After all, everything the readers learn about Adam is filtered through his love rival's perspective which is biased because of Charlie's petty jealousies and human narcissism. The protagonist's worry that humanity will precipitate its own demise through the creation of superintelligent machines – Cave and Dihal's AI-connected fear of obsolescence (75) – permeates every aspect in his perception of the android (*Machines* 80). On the one hand, this portrayal adds to the mystery of Adam and heightens suspense. On the other hand, this creates an asymmetrical power hierarchy that makes it difficult, though not impossible, for readers to empathize with Adam. While the human autodiegetic narrator exerts linguistic mastery over his robotic companion, the artificial character becomes an Other that is reduced to being the object of another's gaze. But if machines remain puzzles to be solved by human onlookers, they cannot gain narrative agency (Kornhaber 19; Yee 92). Alternatively, one could also argue that the refusal to express Adam's innermost thoughts in human terms is a writer's attempt to avoid appropriating his experiences. After all, Adam's mechanical mind might not think in the human sense but process information in entirely different ways, which would necessitate a translation of algorithmic communication patterns into linguistic codes that are intelligible to a human readership. As a result, the novel's many ambiguities defy simple categorization and allow for multiple co-existing readings. Overall, however, McEwan's narratological choices perpetuate a benignly anthropocentric stance. *Machines Like Me*, then, is not so much a novel about sentient machines as it is a story about humans struggling with technology and ultimately themselves.

It is, of course, no coincidence that the novel's central moral conflict climaxes in a dispute about literary works. Adam complains that all of literature "describes varieties of human failure" (*Machines* 149). He prefers "the lapidary haiku, the still, clear perception and celebration of things as they are" (150). In this crucial juxtaposition, McEwan claims that machines are (as of yet) incapable of creating truly innovative art. Nonetheless, he concedes that exposing androids to the complexity of literature might have beneficial effects. Despite Adam's jabs at human fiction, his reading sessions are a vital part in his character formation. When discussing *Hamlet* or *Ulysses'* Nestor episode, Adam unwittingly learns much more about the human mind and decision-making processes than from his programming (202-04). Simply put, the canon of world literature serves as a huge data set that initiates Adam into the non-binary workings of human consciousness. The android, however, does not seem to fully comprehend the implications of his literary education. As a champion of mechanical super-rationality, Adam prophesies an alternative future:

> when the marriage of men and women to machines is complete, this literature will be redundant because we'll understand each other too well. . . . As we come to inhabit each other's minds, we'll be incapable of deceit. Our narratives will no longer record endless misunderstanding. Our literatures will lose their unwholesome nourishment. (149-50)

Charlie is horrified by this dismissal of mental privacy and individuality. Adam's suggestion cements his ontological status as Other because the android's rigid supermoral algorithms aim to crush precisely that which is admirable about the human condition. Failures, unpredictability, and vulnerability define who we are as humans. And all these experiences are captured in literature, humanity's saving grace. When all is said and done, the humanist McEwan remains a firm believer in the beautiful failures of humankind, despite the horrors they have caused.

Conclusion

Ian McEwan's contribution to current debates about Artificial Intelligence is both a timely work of science fiction and a provocative piece of speculative history. It also rekindles a long-harboured passion project of his, namely writing about the life of Turing. The setting in an alternative Britain of the early 1980s allows McEwan to combine disparate elements; the novelist plays with historical events, enmeshes current political developments, invents alternative life stories, and creates corresponding fictionalized mindsets. Apart from redeeming Turing as a figure of inestimable national and scientific value, the novel installs the godfather of computer science as a mediator who introduces complex issues of AI agency and ethics to a lay readership. Moreover, Turing appears as the novel's conscience, for it is he who frequently addresses the potential risks of humanity's irresponsible conduct with science. By designing his novel to function as a Turing test, McEwan elegantly incorporates pressing problems of AI ethics into his narrative. This makes the problem of other minds, anthropomorphism, questions of (machine) consciousness, the Singularity, anthropocentrism, and the validity of top-down rule-based ethical programming for Artificial Intelligence relatable to his readership. Against this backdrop, Adam proves to be an intriguing robotic character who forces readers to consider the benefits of extending the category of personhood to non-human beings.

Unlike many other android novels, *Machines Like Me*'s brave new world of robotics is not strictly speaking dystopian. To put it in the terms of Cave and Dihal's framework, Adam's agency does not lead to a manifest dystopian state of inhumanity, human obsolescence, alienation, or a machine uprising. The fear of a coming robot apocalypse, though latently present in the novel, is soon replaced by the bitter realization that the atrocities humans have inflicted on each other in past centuries are worse than anything that AI can accomplish at present. The Singularity might arrive in the future with all its unforeseeable consequences, but in McEwan's setting it is the machines – and not the humans – who suffer. The high-minded Adams and Eves perish when faced with the depravity and moral complexity of humankind. The seasoned novelist, who proved an expert on the human condition in his previous works, skilfully displays the contradictory nature of human life. When he lets Adam's binary algorithms despair at the concept of white lies, McEwan adroitly demonstrates how complicated human beings are, and consequently how intricate their flawed ethical systems can be. On the other hand, the author also stresses the immense creative potential that the conditio humana engenders. Suffering and the knowledge of mortality can make life unbearable, yet they also serve as catalysts for art. To the writer McEwan, literature is humanity's redeeming feature – something that rule-based machines might never fully grasp. Fuelled by his steadfast humanism, McEwan transforms literature into a tool for machines to learn about the boundless nature of human morality which transcends binary logic. Although reading is fundamental for Adam's character formation, *Machines Like Me* argues that human experience is so sophisticated that it could not be re-constructed by algorithms alone. Even at the worst of times, the fallible human characters appear superior to Adam because they invite sympathy into their suffering. Hence, the novel's ending is once more ambivalent: It serves as a warning not to leave the world to machines and their missing algorithms while simultaneously displaying a timid optimism about a future in which humans take better care of each other. Ultimately, McEwan's android novel really is concerned with the human, all too human.

Notes

1. The work of this article is shared equally between the two authors.

Works Cited

Abney, Keith. "Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed." *Robot Ethics: The Ethical and Social Implications of Robotics,* edited by Patrick Lin et al., MIT Press, 2012, pp. 35-52.

Adams, Tim. "Ian McEwan: Who Is Going to Write the Algorithm for the Little White Lie?" *The Guardian,* 14 Apr. 2019, www.theguardian.com/books/2019/apr/14/ ian-mcewan-interview-machines-like-me-artificial-intelligence. Accessed 8 Dec. 2020.

Anderson, Susan Leigh. "Asimov's 'Three Laws of Robotics' and Machine Metaethics." *AI & Society*, vol. 22, 2008, pp. 477-93. doi:10.1007/s00146-007-0094-5.

---. "How Machines Might Help Us Achieve Breakthroughs in Ethical Theory and Inspire Us to Behave Better." *Machine Ethics,* edited by Michael Anderson and Susan Leigh Anderson, Cambridge UP, 2011, pp. 524-30.

Asimov, Isaac. "Runaround." *I, Robot,* HarperVoyager, 2014, pp. 31-51.

Bal, Mieke. *Narratology: Introduction to the Theory of Narrative*. U of Toronto P, 2009.

Beavers, Anthony F. "Moral Machines and the Threat of Ethical Nihilism." *Robot Ethics: The Ethical and Social Implications of Robotics,* edited by Patrick Lin et al., MIT Press, 2012, pp. 333-45.

Berndt, Katrin. "Science as Comedy and the Myth of Progress in Ian McEwan's *Solar*." *Mosaic,* vol. 50, no. 4, 2017, pp. 85-101.

Bigsby, Elisabeth, Cabral A. Bigman, and Andrea Martinez Gonzales. "Exemplification Theory: A Review and Meta-Analysis of Exemplar Messages." *Annals of the International Communication Association,* vol. 43, no. 4, 2019, pp. 273-96.

Bostrom, Nick. "Existential Risk Prevention as Global Priority." *Global Policy*, vol. 4, no. 1, 2013, pp. 15-31. doi:10.1111/1758-5899.12002.

---. *Superintelligence: Paths, Dangers, Strategies*. Oxford UP, 2016.

Bostrom, Nick, and Milan M. Ćirković. "Introduction." *Global Catastrophic Risks*, edited by Nick Bostrom and Milan M. Ćirković, Oxford UP, 2012, pp. 1-29.

Braidotti, Rosi. *The Posthuman*. Polity Press, 2012.

Cargill, C. Robert. *Sea of Rust*. Harper Voyager, 2018.

Cave, Stephen, and Kanta Dihal. "Hopes and Fears for Intelligent Machines in Fiction and Reality." *Nature Machine Intelligence*, vol. 1, 2019, pp. 74-78.

Cave, Stephen, Kanta Dihal, and Sarah Dillon. "Introduction: Imagining AI." *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, edited by Stephen Cave, Kanta Dihal, and Sarah Dillon. Oxford UP, 2020, pp. 1-24.

Cave, Stephen, et al. *Portrayals and Perceptions of AI and Why They Matter.* The Royal Society, 2018.

Clarke, Roger. "Asimov's Laws of Robotics: Implications for Information Technology." *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson, Cambridge UP, 2011, pp. 254-84.

Coeckelbergh, Mark. *AI Ethics*. The MIT Press, 2020.

Dahlstrom, Michael F. "Using Narratives and Storytelling to Communicate Science with Nonexpert Audiences." *Proceedings of the National Academy of Sciences*, vol. 111, no. 4, 2014, pp. 13614-20.

Dennett, Daniel C. "Consciousness in Human and Robot Minds." *Cognition, Computation, and Consciousness*, edited by Masao Ito, Yasushi Miyashita, and Edmund T. Rolls, 1997, Oxford UP, pp. 17-29.

Descartes, René. *Discourse on Method and Meditations on First Philosophy*. Translated by Donald A. Cress, Hackett Publishing, 1998.

Dinello, Daniel. *Technophobia! Science Fiction Visions of Posthuman Technology*. The U of Texas P, 2005.

Driver, Julia. *Consequentialism*. Routledge, 2012.

Glaser, Manuela, Bärbel Garsoffky, and Stephan Schwan. "Narrative-Based Learning: Possible Benefits and Problems." *Communications: The European Journal of Communication Research*, vol. 34, no. 4, 2009, pp. 429-47.

Good, Irving John. "Speculations Concerning the First Ultraintelligent Machine." *Advances in Computers*, vol. 6, 1964, pp. 31-83.

Gründiger, Wolfgang. "Diversity for Algorithms." *Conditio Humana - Technology, AI and Ethics*, 1 Jan. 2019, conditiohumana.io/diversity-for-algorithms/. Accessed 8 Dec. 2020.

Gunkel, David J. *The Machine Question: Critical Perspectives on AI, Robots and Ethics*. MIT Press, 2012.

Guyer, Paul. *Kant*. Routledge, 2006.

Hall, Louisa. *Speak*. HarperCollins, 2015.

Head, Dominic. *Ian McEwan*. Manchester UP, 2007.

Hodges, Andrew. *Alan Turing: The Enigma*. Vintage, 2012.

Holland, Rachel. *Contemporary Fiction and Science from Amis to McEwan: The Third Culture Novel*. Palgrave Macmillan, 2019.

Horton, Emily. *Contemporary Crisis Fictions: Affect and Ethics in the Modern British Novel*. Palgrave Macmillan, 2014.

Huxley, Aldous. *Brave New World*. Vintage, 2004.

Kakoudaki, Despina. *Anatomy of a Robot: Literature, Cinema, and the Cultural Work of Artificial People*. Rutgers UP, 2014.

Kang, Minsoo. *Sublime Dreams of Living Machines: The Automaton in the European Imagination*. Harvard UP, 2011.

Kant, Immanuel. *Groundwork of Metaphysics of Morals*: *A German-English Edition*. Edited and translated by Mary J. Gregor and Jens Timmermann, Cambridge UP, 2012.

---. *Practical Philosophy*. Edited by Mary J. Gregor, Cambridge UP, 1996.

Kirchhofer, Anton, and Natalie Roxburgh. "The Scientist as 'Problematic Individual' in Contemporary Anglophone Fiction." *Zeitschrift für Anglistik und Amerikanistik / A Quarterly of Language, Literature and Culture*, vol. 64, no. 2, 2016, pp. 149-68.

Kornhaber, Donna. "From Posthuman to Postcinema: Crises of Subjecthood and Representation in *Her*." *Cinema Journal*, vol. 56, no. 4, 2017, pp. 3-25.

Kurzweil, Ray. *How to Create a Mind: The Secret of Human Thought Revealed*. Penguin, 2014.

Leckie, Ann. *Ancillary Justice*. Orbit, 2013.

---. *Ancillary Mercy*. Orbit, 2015.

---. *Ancillary Sword*. Orbit, 2014.

Leenes, Ronald, and Federica Lucivero. "Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design." *Law, Innovation and Technology*, vol. 6, no. 2, 2014, pp. 193-220. doi:10.5235/17579961.6.2.193.

Mahon, James Edwin. "The Truth about Kant on Lies." *The Philosophy of Deception,* edited by Clancy W. Martin, Oxford UP, 2009, pp. 201-24.

McEwan, Ian. *Atonement*. Jonathan Cape, 2001.

---. *Black Dogs*. Jonathan Cape, 1992.

---. *The Child in Time*. Jonathan Cape, 1987.

---. *Enduring Love*. Jonathan Cape, 1997.

---. "Introduction." *The Imitation Game: Three Plays for Television*. Picador, 1982, pp. 9-18.

---. *Machines Like Me and People Like You*. Jonathan Cape, 2019.

---. *Nutshell*. Jonathan Cape, 2016.

---. "Only Love and then Oblivion: Love Was All They Had to Set against Their Murderers." *The Guardian,* 15 Sep. 2001, www.theguardian.com/world/2001/sep/15/september11.politicsphilosophyandsociety2. Accessed 8 Dec. 2020.

---. *Saturday*. Jonathan Cape, 2005.

---. *Solar*. Jonathan Cape, 2010.

Miah, Andy. "Posthumanism: A Critical History." *Medical Enhancement and Posthumanity,* edited by Bert Gordijn and Ruth F. Chadwick, Springer, 2008, pp. 71-95.

Muggleton, Stephen. "Alan Turing and the Development of Artificial Intelligence." *AI Communications*, vol. 27, no. 1, 2014, pp. 3-10.

Murphy, Robin R., and David D. Woods. "Beyond Asimov: The Three Laws of Responsible Robotics." *IEEE Intelligent Systems,* vol. 24, no. 4, July/August 2009, pp. 14-20. doi: 10.1109/MIS.2009.69a.

Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review,* vol. 83, no. 4, Oct. 1974, pp. 435-50.

Newitz, Annalee. *Autonomous*. Orbit, 2017.

Newman, Daniel Aureliano. "Narrative: Common Ground for Literature and Science?" *Configurations,* vol. 26, no. 3, 2018, pp. 277-82.

Nicklas, Pascal. "Art and Politics." *Ian McEwan: Art and Politics,* edited by Pascal Nicklas, Universitätsverlag Winter, 2009, pp. 9-22.

Norris, Stephen P., et al. "A Theoretical Framework for Narrative Explanation in Science." *Science Education,* vol. 89, no. 4, 2005, pp. 535-63.

Powers, Thomas M. "Prospects for a Kantian Machine." *Machine Ethics,* edited by Michael Anderson and Susan Leigh Anderson, Cambridge UP, 2011, pp. 464-75.

Proudfoot, Diane. "Turing's Concept of Intelligence." *The Turing Guide,* edited by Jack Copeland, et al., Oxford UP, 2017, pp. 301-08.

Puschmann-Nalenz, Barbara. "Ethics in Ian McEwan's Twenty-First Century Novels: Individual and Society and the Problem of Free Will." *Ian McEwan: Art and Politics,* edited by Pascal Nicklas, Universitätsverlag Winter, 2009, pp. 187-212.

Robertson, Jennifer. "Gendering Humanoid Robots: Robo-Sexism in Japan." *Body & Society,* vol. 16, no. 2, 2010, pp. 1-36. doi:10.1177/1357034X10364767.

Rosenfeld, Gavriel. "Why Do We Ask 'What If?' Reflections on the Function of Alternate History." *History and Theory,* vol. 41, no. 4, 2002, pp. 90-103.

Russell, Nicholas. *Communicating Science: Professional, Popular, Literary*. Cambridge UP, 2010.

Schaffeld, Norbert. "The Portrayal of Nazi Germany in the English Novel." *English and American Studies in German 1987: A Supplement to Anglia*, 1988, pp. 109-12.

Scheffler, Samuel. "Introduction." *Consequentialism and Its Critics*, edited by Samuel Scheffler, Oxford UP, 1988, pp. 1-14.

Schneider, Susan. *Artificial You: AI and the Future of Your Mind*, Princeton UP, 2019.

Searle, John R. *Mind: A Brief Introduction*. Oxford UP, 2005.

Shakespeare, William. *The Tempest*. Edited by Virginia Mason Vaughan and Alden T. Vaughan, Bloomsbury, 2011.

Stephani, Tilman, Gunnar Waterstraat, Stefan Haufe, Gabriel Curio, Arno Villringer, and Vadim V. Nikulin. "Temporal Signatures of Criticality in Human Cortical Excitability as Probed by Early Somatosensory Responses." *Journal of Neuroscience,* vol. 40, no. 34, 19 Aug. 2020, pp. 6572-6583. doi.org/10.1523/JNEUROSCI.0241-20.2020.

"Stephen Hawking: Brain Could Exist Outside Body." *The Guardian,* 21 Sep. 2013, www.theguardian.com/science/2013/sep/21/stephen-hawking-brain-outside-body. Accessed 8 Dec. 2020.

Turing, Alan. "Can Digital Computers Think?" 1951. *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, Plus the Secrets of Enigma,* edited by B. Jack Copeland, Oxford UP, 2004, pp. 476-87.

---. "Computing Machinery and Intelligence." *Mind*, vol. 59, no. 236, 1950, pp. 433-60.

Vinge, Vernor. "The Coming Technological Singularity: How to Survive in a Post-Human Era." *Science Fiction Criticism: An Anthology of Essential Writings,* edited by Rob Latham, Bloomsbury, 2017, pp. 352-74.

Wallach, Wendell, and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford UP, 2009.

Wells, Lynn. "Moral Dilemmas." *The Cambridge Companion to Ian McEwan*, edited by Dominic Head, Cambridge UP, 2019, pp. 29-44.

Yee, Sennah. "'You Bet She Can Fuck' – Trends in Female AI Narratives Within Mainstream Cinema: *Ex Machina* and *Her*." *Ekphrasis*, vol. 17, no. 1, 2017, pp. 85-98. doi:10.24193/ekphrasis.17.6.

Zalewski, Daniel. "The Background Hum: Ian McEwan's Art of Unease." *The New Yorker,* 15 Feb. 2009, www.newyorker.com/magazine/2009/02/23/the-background-hum. Accessed 8 Dec. 2020.