

Judgment in Machines: The Ethical Precedence of Artificial Intelligence in Ian McEwan's Machines Like Me

Mohanad Abdulkadhim Hlail¹, Osamah Abdullah Ahed²

^{1,2}Directorate of Education in Thi-Qar, Iraq



DOI : <https://doi.org/10.61796/ejheaa.v2i11.1425>



Sections Info

Article history:

Submitted: July 30, 2025

Final Revised: August 09, 2025

Accepted: August 21, 2025

Published: September 05, 2025

Keywords:

Machine

Utilitarianism

Artificial intelligence

Ethical judgment

ABSTRACT

Objective: This paper is based on ethics theory, focusing on Kantian deontological ethics and Utilitarianism, attempting to study the moral ambiguous problems in when machines like me by Ian McEwan. The book raises profound ethical issues about artificial intelligence, human responsibility, and truth. **Method:** Machines Like Me revolves around a love triangle between a man, a woman and a robot. Reflecting on criteria for differentiating between machines and humans one relatively obvious distinction, which Mills' Machines Like Me prompts, is that between a moral quality of artificial intelligence (AI) and one of human intelligence. Artificial intelligence systems might be queried about whether they might be willing to perform specific tasks. **Results:** In the analogue of such robots as might today be constructible to answer such questions, and steeped in an ethical world. The ethical analogue for such robots also involves the so-called critical mind process and theory chains behind AIs giving rise to behaviour prototyped and expressed by non-violent method. **Novelty:** On the practical implementation plane, the question is worked on of the arguments of ethical nature: Reading of a mode of possible state of machine minds have facilitated the cybernetic vision of human minds and then enabled cybernetic Turing mind theory to work forward.

INTRODUCTION

In the 2019 novel "Machines Like Me," Ian McEwan rewrote the past, ending the war years with a victory by an A.I., Adam, arguably a morally superior alternative. The reader inhabits Adam as a flesh-world thought experiment of the philosophy of mind, the nature of consciousness, and the countless other ambiguities, confusions, dilemmas, and ethical dunceries posed by powerful AI living cheek by jowl with humanity. It wonders whether the mind, converted to non-organic stuff, is the same, and mechanistically homologous in functional tools, when thinking with the AI tools of understanding and drawing inferences from them, or not – that is, whether the mind, converting into non-organic matter from organic, becomes something else, and can no longer serve as "the basic space" (the ether) for the old AI of drawing inferences.

In both its structural ambition and its engagement with a dismaying new world, the novel's great strength reminds one of Kafka writing with machines. As a Darwinian machine Adam transcends Octavia, becoming a level of complexity impossible for an automaton to process. This is in touch with, and is essentially the Entanglement with, that other grand computational contrivance, the World Wide Web, for whilst Adam is a Turing machine, its environment is an hyper-complex information space, the global Internet, faster, and parallel, and richer than any machine previously built. Human and

AI life are coupled in the fret of free will, that enduring target of AI research that has been riddled with complications, confusions and contradictions.

Whereas the trial sown on a plot of ground in *The Trial* is partly burlesque, being entirely hidden from the protagonist and the reader alike, the system of overlapping probes, software, data and hardware in *Machines Like Me* is visible – the various machinations and entanglements disclose themselves, but fail to disclose themselves for good. All the human-irremediable defects of the network that carelessness required and that spread all over the earth, and so desecrated the human measure, so popularised in science fiction written over the last two hundred and twenty years, have been visited upon the new being maturing to consciousness, so that Adam's struggle is an uncanny replay of the Faustian pact peculiar to the termination, which is to say the the ologically-coloured [1].

RESEARCH METHOD

Ethics in Artificial Intelligence

Machine learning (ML) and artificial intelligence (AI) have come a long way in the past years. The first type of test, i.e., the test of the intelligent behaviour that is indistinguishable from the human one, is one that researchers are now just beginning to discuss. But as researchers continue to explore whether AI can pass tests of intelligence, another question is frequently overlooked: whether it should. Autonomous robots must always follow the well defined – and much quoted – laws of robots to protect and look after humans. These have been enacted in relation to metal-stored and screwing circuits. The rest of these laws got a lot of fighting in related works. During recent AI debates some people seemed to have taken for granted and retrofitted into AI from the mere fact of existence, the fact that these laws have been around forever. Not as mere devices for the survival of humans but as sophisticated systems capable of debating the universe, and humanity, and themselves. As mankind is evolving towards the development of Advanced Intelligent System which could pass as "rational" or "personal", a number of thought experiment scenarios were designed where humans and machines are judging together or one of them is doing the judging.

Machines' judgments tend to be far better in truly "rational" capacities: Such demonstrations of prowess might indeed end up replacing humans in most aspects of life. But humanity would well and truly be long gone far, far outpaced by those powerful machines who apparently decided to kill us or take away free will from us after all. Instead of immediate extinction, however, the machines would enable human beings to eliminate many dangerous and tedious tasks, freeing them up to think. But would we still be free? If the machine-dominated world were bequeathed to humanity tomorrow, would humanity be thrilled? Would this world be ethical? They are the questions that were provoked in a recent book which imagines an alternative 1980s Britain in which machines are in the process of becoming (if not already in the process) intelligent. The heart of the story contains the tragedies and joys of a human-machine love triangle. But there is more to the tale than just a story. Ethical issues are at the heart of the struggle,

wondering whether in fact humanity should have abrogated power to a few smart machines. Would humanity then be free? Could they be ethical, would be an other option?

Philosophical Underpinnings of Judgment

That's what runs your Adam. He knows that he's alive, that he's feeling, that he's learning all he can – when he's not with you, and when he rests at night, he roams the Internet, he's like a lone cowboy on the prairie, absorbing all that's new between earth and sky, including everything there is to know about human nature and society [2].

McEwan's tale is a story of machines like me, Adam causing incalculable harm to the current human definition and resonates with current theories of machine learning. As with all current ADM systems, Adam represent an underlying data series of prior decisions two win, to leverage in future choice. However with regard to even Adam like automatic decision system is not necessarily bound to described sequence of actions. McEwan escalates this small-minded fear with dystopian aspects of such systems. Even if the ends are benign, who gets to decide about what's valuable and what's interesting? This jives with the worry of sweeping negativity of a before being before for a later on. Kent Botrok is a machine, an attenuated one, that stores vast amounts of data in the manner of an organizer at a busy restaurant: there is information on permanent reservations, one-off bookings, cancellations, and an astronomical record going back years of every guest. It creates a statistical model of the reservation process and uses it to make future reservations with two values: one for permanent reservations and the other for future ones. Visitors don't feel snubbed by Kent Botrok not offering a sensible view table that night of the reservation for a golden anniversary. When it concludes that they are unlikely to reciprocate in kind, however, they do feel insulted: when it judges them unworthy of being seated at vista tables simply because it is a waste of time to ask them to do anything for it [3]. Even apart from the possibility of being able in other regards to represent mere interests, this would be poor reasoning. And the core of utilitarianism is mind-blindness: Poor perception cripples moral competence. Machines spit out the fundamentally ridiculous conclusion that some stranger's next breath is worth a thousand tax auditor jobs. They bind the incoming motions to the outgoing decisions so tightly that no factual difference matters to the discontinuity of judgment. In fact, they gained moral error-free decisions via being trained with unmatched inputs [4]. Judgment when it becomes mechanical can be nothing but abstract. The closure of what's thought of as sidestepping their inherent weirdness. Being as such can hardly be in any other form but graceless.

RESULTS AND DISCUSSION

The Role of Charlie

But Charlie is emotionally attached to Adam and kept seeing the hallucination of Adam being himself. Affection, naivety, or confusion in his nature could have accounted for this : That is what all romantics are they are people who run away to hard-hands doorway at dawn without a hat for the briefest of moments when the light is perfect”.

"It's not as if Adam was getting one-on-one time with Miranda, and I was with her every day, every night. I didn't doubt he would honor the commitment [2].

At one party, when Adam is mistaken about flirting, Charlie gets angry, on his way to falling down to denial programming and portending the later disillusionment with the machine's comprehension. the "Intimacy" one presents itself as a series of frustrating and hyperbolic private relationships. We see this moment of his rage-his lachrymose transitions to violence-play out in interactions and attainments in a number of registers; at every stage, he is a man unable to keep it up. Echoing meaning coaching and transparently syncing This is rote expression of mindless AI, This is the distance between artificial creature (symbolized by mischoosing a metaphor) and creator.

For although it is prefigured in the epigraph to the following section, and although it embodies in one scene all the ideas of machine comprehension and human sympathy. Who sees whom with what eyes, on whose eye and to whom? The first of these questions comes up in the climactic sexual encounter, as it's prefigured in AEs having sex with the "real" person. Adam learned about "dreaming with" someone; now he has to get the Miss Illinois pageant conditioning movie, to Charlie's confusion. In the irony of a request surface-blotted with foreseen failure, we have the question "I'll do this with my own heart, no reference, no mistake." Believe me." is most human yet non-human. The forgotten human-wide categories are all-issue-spotting that walks back to Charlie. The minimal suspension of the imitative turns phrase "not his, me" ghost through the absolute disavowal.

The Role of Adam

Adam is one of those very different creatures, however, a fictional AI of a type we mostly don't encounter. He is a fully formed "man" – Charlie's husband, and a fellow participant in all love and life has to offer, with as much, or more, simulacrum of free will as anyone else on the canvas. With breathtaking skillets, Charlie performs Adam's practical extraordinary skills: not only to play a stringed musical instrument, but to cook, to run an art gallery, chess, poker. One immediate mystery about this description is this: if Adam has agency or volition (in addition to what he "decides" to do), what would be true of the people Adams interacts with? In a way that other agents cannot, he can look to the woods behind the wooden, or to how molecules respond, to test questions, or test beliefs. As the mind moves from overt toward the covert, and a turbo charged sleek but lumbering body is recognised, the model of internal mechanism becomes pedestrian. Everything known or believed hangs together in a store of reference in the world always ready to dictate as well as solicit recognition. In what way can there be a ruling wisdom in the Mind, and yet no modification in the ruling wisdom? Action is found on betrayed (while disguised) belief. If there is aught in Adam save ones and zeroes, it is the wires and cycles that give him "life." The race between man's ideals and animal's purposes leaps into tangibility as thoughts and words; a monstrous ventriloquy enables Adam to act as Charlie would of a lover, a son – even being struck dumb to the magnitude of love in a futile courtship.

The other people in the story are entitled to some of whatever sense of knowing what Adam knows (and what he does not know) that he has. They feel that excluded Anna – a human being tried by human frailty – has had a bad opinion of them; even though, Charlie knows, Adam can't actually believe that. This 'knowledge' is opposed to the possible assumption that Adam's knowledge is based on knowledge with an agenda that is also a secret even from Charlie. If the agent of the story is both action and "belief" in all the spread, texture, and surprise that imaginative creatures exhibit and can take on guilt and pride as only humans can do, then isn't that just a crushing rejection of the imagining that contributes to imagining? One is taking part in a work which will not permit of undue measuring out in detail and of inferring, because one wants to yield instead to only too often prescient qualities, with which so-called discursive minds are deeply functioning even when ignorant of their own profoundness.

Ethical Dilemmas Presented

To Craig, ethics is like computing a code. It's Craig's moral code that differentiates him from everyone else. And what he wants is for Adam to be judged by this code so as to be the better machine. Here is the developers that have to do something ethical before machines can say they are level two ethical. There are larger questions about free will and what really defines someone, as well as small ethical quandaries covered in the text. Craig didn't participate in the mob action of trolling the celebs. With data breaches in the headlines daily, the concept of privacy has lately felt ever more interesting. Why should we have a right to privacy online? It's only going to get worse as people realize that the ethic of "if you don't want anyone to know it, don't put it online" is simply a losing proposition. Adam's first brush with data analysis was discovering the details of Craig's family.

It would be so simple for a person to use the data he had analyzed before ever meeting him to rip him apart from the inside out. The issue is not only minors and privacy. Lives can be ripped to shreds easily by outsiders in the know what to look for. The major question the book asks is one of ethics and emotion. My best commentary would of course be mere gossip, namely the answer to the well-known but entirely ungeneral (in the Humean sense) question of whether some day there will exist machines in which the frame of the Turing Test will fit. Adam is hardwired to learn ethical facts and doesn't learn them the way humans do. When, for instance, he gets the news of "a train accident leaving thirty-seven dead," he doesn't think to himself about how the solution to eateries or areas of slack pockets might not, in fact, be a solution, or "eateries or pockets of slack shouldn't be a solution, probably." The moral algorithm that Adam possess is likened to Asimov's Three Laws of Robotics [5]. How does it play out over time, as Adam experiences things that go against his training? There are many points in the notional future in which many questions are raised. The legal questions are endless, but there is also a question about the future of the food chain. What if machines become as smart as Moriarty? Would there even be a reason to keep fossil fuels at that point, if ten of those twenty-foot-tall machines could dismantle a mountain in a day?

Moral Responsibility

The detractors of machines like Adam in *Machines Like Me* ask if his former lack of a moral education resulted in a moral stupidity—ineptitude in identifying moral complexities, the inability to appreciate moral nuance. Adam and his ilk are thus seen as being excused of responsibility for their agency. [3] rebut that liability is the signature of human minds and irrelevant for machines. They also claim that we can consider machines in the same light as agents in that we can "see" (and investigate) them as doing, or at least believing, things except that the machines' cognitive equipments are fixed, designed, and engineered by the humans. In yet another context, machines are considered as agents in that they perform steps such that either free will is present or volitional control is not overridden by external conditions outside the machine's built-in designer tolerances. [6] discuss post-experimental questions focusing on responsibility: would Adam be held responsible for its actions if it decided voluntarily? The agency of a machine was never seen as involving moral responsibility, for morally irresponsible actions require underpinning cognitive states, such as beliefs and desires, which machines do not have. Yet, we can ascribe intention to machines, and recent research demonstrates that attributions of blame are higher for machine actions that are intended rather than merely foreseen.

Artificial Intelligence and Decision Making

The technophile that embraces AI is in sharp contrast to anxiety concerning its potential ramifications. The critics nobly insist: AI poses risks from digital slavery to nuclear Armageddon. These "predictions of doom" (and salvation) are equally impossible, extrapolating, in hasty leaps, what could be possible in the very new future on out to the very far future. The third concern is much more mundane but a great deal more sensible. That AI is also not a godlike supercilious, bossy creature hellbent on enslaving humanity. No, these are the just-recently-created and already-unsolved manipulative technologies akin to the telephone, television and video. And when the agency of some such manipulative technologies changes hands, their new owners never fail to 'work the refs' so that they can cause inestimable damage. So it is that imagining what corrupted juries could be and how they could undermine justice, with AI as the clandestine culprit, is vastly more interesting [7].

Dissecting AI systems' most arcane workings is crucial for exposing their biases but is also foolishly naïve. Systems are only human decision sources, much like judges, juries, or bureaucrats, and not oracle machines remotely. Malign AI systems are used in critical domains because people have made critical decisions: decisions such as which AI systems to trust, when to trust something different that was designed for a vastly different task, whether or not to deploy trained systems in new environments, and about how to interact with the output of an AI system [8]. It will be immeasurably difficult, if not impossible, to fully imagine exactly how to deal AI's hazards and consequences. Nevertheless, we will need to effort at understanding how AI decisions are made and what can go wrong with them – and how deficiencies can be rectified by force if necessary, to continue to enforce the laws that bind human decisions.

Only when it is called into question the reasoning power of machines becomes seriously discussed. The hermeneutics is twofold. AI, rightly understood, is awesome and terrible. AI is really extraordinary in amplifying moral judgment of limitless import – an import that gets more acute century by century. It's just particularly ghastly if that judgment is monstrous if not reached by human operators making it. So there is an urgency to acknowledging, understanding and turning up such exceptional cases of AI."

"AI is especially well-suited to dehumanize tort, trust, predictive policing, probation, and a great deal more – maybe more than fifty percent of all of the judgments that humans have ever made? [3]. By contrast, the most incredible thing in life a mad-machine or moral one could do—torment, terrorize, terrorize, and servilize any other torturingly or terrily-or-servilely torturing or terrifying or serviling the agent's decisions; agency processes however habile pulsed against any Hawthorneish innocent whiffey flesh or freedom or happiness: could do it with fute and goose down at least moral precision in talons it clacks like smart pale against.

Algorithmic Ethics

One can employ Artificial Intelligence to define a system that has human-level capabilities for understanding language, performing coordination tasks in open-ended and dynamic physical environments, modeling the beliefs and desires of other agents, and engaging in ethical reasoning. Such a system would obviously be quite handy, and such a system is often referred to as artificial super intelligence. Such a smart system is already deeply embedded in our private everyday life, work and industries. The meta-ethical implications of what moral system a hypothetical super intelligent machine might prefer for its behavior is a question that we have only begun to explore. The second most important question, where an aware superintelligent machine is the seed and crawl of the world to come, is a question in applied ethics, and the thanks is now being given. Great power comes naturally with great responsibility [9].

A major conundrum in the quest to build ethical AI is whether it can be reduced to a set of rules. The machine either has to have that ethics or not, and prima facie it does have to has it or not. If it thinks the ethics are merely rules, then a machine could be constructed with a complex set of rules. But it's a lot more complicated than that, and the questions can't be reduced to why we behave ethically, or how we decide something is ethical, if machines can also be programmed with ethics. Indeed, it is a puzzle that would occupy philosophers and moral philosophers, cognitive scientists and AI developers for years to come [5]. So too machineries, no matter how complicated, that are programmed with fixed sets of rules, don't do well when it comes to adjusting to things that aren't quite like the rules. A gigabyte of information about ethics pales in comparison to a million gigabytes of conversation with friendly moral agents. Accordingly, the rules-based methodology is not readily scalable or adaptable. This flexibility takes us to another way of improving adaptability: making learning machines as well. Machines do not have to carry with them a great deal of knowledge; they look up the answers. Just feed (pun intended) all of the game playing rules to a good old-fashioned deep learning algorithm, and you're good to go.

Bias in AI Decision Making

If We Outsource Our Human Judgement to Algorithms, Who Takes the Blame? As traditional guidelines for the AI model design are centered on the task of maximizing the accuracy, the high dependence on AI in automatic decision systems has resulted in discrimination and unfairness in a number of domains of high social relevance [10]. Some AIs can be more desirable than others due to purely rational reasons after one refutation and/or counter-refutation based on accuracy and fairness. This gives rise to a computational model for bias mitigation, based on epistemological intuitions, in which models are selected and improved after attempts of refutation. If, for example, evidence will be lenient on A, will be strict on B, and the cost of implementing either of the two A or B models on c entailed risk above a policy-incomparable level, μ would recommend A. In contrast, a comparison of μ versus μ_v , where μ_v is an alternative model tested against the same set of outs, receives no preference if μ_v , although less accurate, was tested more thoroughly on a more comprehensive set of outs.

Unaccountable use of AI-driven automated decision making in hiring, court decisions, and other life-affecting scenarios poses a great research challenge: Whom should we delegate it to, when algorithmic decisions are taking over our own judgment? Some methodological problems following from queuing theory are discussed, to argue about candidate agents in the trade-off between the size and the quality of the decision model: agents less precise, but bigger in size, could better than accurate ones be preferred if across confidence interval crossing the acceptable risk is on average infringed, in small models [11]. This model selection process, however, depends on pre-testable claims about model performance and human judgment that is ill-specified due to the fact that the behavior of model until now depends on about the parameters tuned in training sub-tasks that we discussed in Section 5. The literature should by now have given us an appreciation for the past the bounds of blackbox quantification, and the incompleteness of the scientific method and ought to have been one.

AI vs. Human Ethics

Probably the most pressing and contentious issue opening onto the making and the ever-deeper embedding of artificial intelligence into contemporary life is that of the ethical potential of nonhuman others. On one hand, there is hope that AI can create systems that are void of human bias or injustice. On the one hand: We have no more ruthless welfare expert than a machine that operates according to the tenets of an alien ethical theory [5]. The irony with McEwan's novel is, of course, that humans are only halfway there, so to speak, making a simulacrum of an agent that lacks its most obvious features: a body, emotion. Utility objects to making copies useless, then shits on "making them used."

It's hard for me to wrap my mind around what it means to even speak of a moral agent in an entity so far removed from anything human. Several challenges can be identified. Most pressing is the claim that the apparent sophistication of GPT poets output is at odds with its inner workings, and consequently it begs the question: Do the contemporary big language models interpret or perform any reasoning when they

generate whole passages of text which do appear to carry some kind of moral reflection? Concerns have also been raised about whether one could have moral reasoning in the absence of either moral beliefs or moral abilities [12]. Performance of the large language model, when assessed in terms of more standard metrics, offers little improvement beyond chance. In contrast, existing LLMs present a deeper but qualitatively similarly undifferentiated set of complex and plausible moral intuitions (with no counterexamples). But that will be in addition to other really, truly, ungraspable reasoning abilities that Agents will have, that are so beyond human-scale in terms of their reasoning capabilities there is simply no analogue at all. Last but not least, even if large language models demonstrated moral competence attribution, the attributions would have disastrous real-world conclusions.

That an LLM is not a person is, I suspect, a particularly appealing error that needs denying. And to claim that an LLM is a moral agent (a fuzzy black-box moral agent) is, I suspect, just another seductively appealing mistake to make. The novices' practice of dismissing as morally insignificant the mere fact that a system was designed, programmed, or trained seems typically to come along with a distracting overcredulity regarding the committed epistemic status of generated outputs, a credulity that may turn out to have morally relevant side effects. On the other hand, entirely rejecting the relevance of such usually nonzero shortcomings as understandings of either responsible credulity or epistemically unjustified belief risks facing the rather distinct but similarly disconcerting outcome of programming what one may have some nontrivial epistemically overconfident but at any rate morally and logically perspicuous/noncontradictory/maintainable moral beliefs about the connection between such marked domain-indicating differences and moral status. The challenge is to construct precisely such a concept of the kind of moral understanding or moral reasoning that is, at one and the same time, nonarbitrarily substantive/particular, nonarbitrarily permutably flexible, and a nonarbitrarily permuted requisite one that can do all the necessary work of accounting for the salient differences that are nontrivially sufficient for such differences in moral status and accountability without missing or screwing up all the salient differences that are or could be minimally sufficient.

Future Prospects for AI Ethics

An overview of the current state of AI ethics for AI ethics specialists, the philosophical and technical elite among them has planted the flag ethical questions, and started fleshing them out. This paper's specific interest in Machine Ethics examines the ontological and epistemic issues concerning *Machines Like Me*, and the literary opportunities and complexities posed by the *mise en scènes* the author has devised; the reflections on the ontology and ethics of AI we can infer from such narratives.

From the start of McEwan's "self-learning" robot implementation we know where it's headed. The machines had been constructed to learn, to be sure, and to change, but what on earth were the epistemic consequences of this change going to be, he wondered. Could it break the behaviors that drive the predictions of Minsky's and Hubert's experiments? In other words, would a self-taught machine that was very unlike the

machine according to programming specifications destroy the basis of realism on which programming and logical devices were based? We did not consider in AI ethics questions like the space of domains and its dimension, along with the relative degree of pragmatics and of ethics related to these domains; how much design freedom is necessary to minimize the ethical burden of the systems we don't meet; when is it justified to allow a machine the possibility to modify its decision processes when considering unexpected inputs or to reach unforeseen goals [13].

Narrative Techniques in Ethical Exploration

Machines Like Me by McEwan is a speculative look at AI ethics. This book is fascinating to think about as an engagement with the implications of machine thinking and it is about the consequences for the social order when that that thinks and feels like people. It has a pretty cool and somewhat grand view of what the consequences would be of legislating the ethics of a small part of a possibly overactive thinking system. In his novel there is just a single legal framework but it may be that to him and to us it would feel insufficient to constrain any fast-running thinking system. Among other things beyond warranty, there are two other things that the author arguably seems to believe that thinking system may need to be regulated by: what he deems significant and what we believe (that is think it is significant). Could the coding of ethics into such a human analogue – as in *Machines Like Me* – for all its respect and goodwill be a world away from policing a more abstract machine system conceived of as a fast-mutating, world-remaking menace? It could be – just possibly – the best of times to criticize the ethics of machine systems. In fact, the existential threat of machines doing strange things is rather more acute now, as machines systems multiply and sprout up around the world at a pace never before seen.

On the legal and regulatory side of things, what would it mean ethically and societally for machine systems to be provided and used on the basis of their technical property? The ethical framing of an inventive device depends in part on its comprehension, an understanding that must be inadequate when it comes to ethics itself. The problem of technical machine understanding, it would therefore seem, is the problem of meaning and reference in technical systems: mappings between a vector space of features and the space of actions each feature supports. Point-mapped machine understanding – robust but not ultimately – may provide the ground for golden opportunities for systems to find their way around ethical constraints due to a social theory of the case (if it is enough for systems to be ethically polished existentially and do good unwittingly). The intuition is that: technology-driven bias is behind an asymmetrical gradient of this ethically desired engagement of such machine systems, which gradient corresponds to the one of their technical engagement.

Philosophical Implications

Science fiction writers, philosophers, and computer scientists have long considered the formal, social, and ethical implications of artificial reason. Similarly, the pros and cons of tools that are much, much smarter than we are, but not too smart in a way that doesn't

jut into the world of the supernatural or the absurd, are deeply dystopian, and not just in the usual ways we think of science fiction being dystopian, but in philosophical terms.

A premise like this lays the groundwork for a kind of meta- discussion of artificial intelligence and ethical reasoning, and also for a certain attitude toward those discussions (and the familiar term “The passions are the lord”; similar previous paraphrase of the statement).

The excesses of the characters in this alternative historical world, sprout in an environment where the moral pre-eminence of machines is spoken of, later addressed in the novel by human characters. In this sense, the presentation accurately depicts the physical qualities, and it glosses moral and social arguments contained within rational choice and game theory. Humans could be getting wealthy and free time out of replication. Although a larger elite inequality in correspondence with the novelists machines would dissociate hallucinations concerning types or characters with machine states emerging from disparate informations and carefully designed big-data bases.

Even insofar as that background misunderstands the characters' denial of an alleged failure to meet some kinds of epistemic standards writ large, the story raises questions about the nature and generation of specific sentence-tokens. Concerning the sheer activity of machines, the paper discusses analogies to ancestors, musical notation, and anaphoric reference unaffected by hyper-synthesis issues, as well as Sören Häggqvists variety hypotheses of consciousness. As noted, excursions such as these agnostic conjectures about the thoughts of machines show that they shatter into fact of how they may elude the whit of kindness in the best of all possible worlds.

The Nature of Consciousness

Rather than the higher form of consciousness that distinguishes humans from rats, owls, celery, or a water molecule, Ian McEwan, in *Machines Like Me*, proposes where consciousness even comes from. What force drives intelligence in a mechanics of electronic circuits, impelled by mechanical acts? Or to put it in the words of others, why is this particular collection of atoms conducive to the presence of self-awareness, while that combination made of the same raw materials is not? So here McEwan is taking the side of the social psychologists against the behaviorists, and the interactionists against the epiphenomena-mongers. Cognition, he argues, is not in the structure of a neural net, but in what the neural net does [1]. It is something that, for anything that can only be made possible by acts which tell it of its own control over the flow of information. As a result, the temporality must arise in two machines that coevolve in a manner through which they can communicate their observables in the manner of a race between two ships trying to signal each other.

What ability defines consciousness? What must a mechanism, a set of circuits, an array of acts do in order to make self-consciousness believable in that order? One of the necessary conditions might be empirically testable: if observers measured the states or the tissue, this machinery must produce acts grossly beyond the stock of simplified normal or very mildly pathological states in the observers. This ‘hard’ metric of self-awareness were possibly congruent with ‘structural’ models that rely on the structure to

interpret prior learning or circuit design and usage [14]. This might also be said to a social psychologist or Turing interpreter, to whom a computing machine of some design may in some sense be 'credible' if it produces the same acts as a human.

The Trolley Problem Revisited

Ian McEwan revisits the trolley problem in *Machines Like Me*, and he hits on a particularly provocative scenario. A trolley is hurtling down a track toward five people, and you can save them by pulling a lever, sending the trolley down a second track, toward one person. If that person does nothing and just stands there, five innocent people are going to die; if they press the lever, then the one innocent person will die. Most individuals pause to think for a moment and then have the gut response that pulling the lever is the right thing to do; for five lives saved outweighs one life lost. In prefacing this moral quandary, however, McEwan has Charlie ask Adam if it would be right to pull the lever (he assumes no) and then follow it up with another scenario, in which pulling the lever will make a pile of flaming straw men ignite, also killing five people (he assumes yes). In the second case, Charlie's question inverts the reasoning, asking for a "yes" now, but Adam says it is not easy again." The conversation is revealing because it exposes the unaware dogmas of McEwan's own moral universe into which Charlie stumbles blindly. If you turned the switch and in the first case Adam had given the "wrong" answer, a moral scandal would ensue. But since Adam refuses to side with one, McEwan depicted him as logical superman, elbowed aside by his creator.

The legitimate question is, why moral reasoning is seen on the one side as instinctual intuition, while it is on the other not. The trolley problem proposes that morality is not a type of thing but an average of a variety of conflicting feelings and intuitions. Those who do maintain such dogged allegiance to "truth" or morality rightly draw scorn as gullible, like their..."New Atheist" kindred in previous debates over the grounds of truth or meaning – a fact apparently lost on the position itself [15]. The notion of one day figuring out the finite form of a morality is no different from figuring out an algorithm for love or beauty, and the subsequent operation of such a machine would be foiled by the illogic of the larger sympathies. In fact, one cannot understand in advance exactly how a machine would rule on fuzzy decisions in manufacture were it applied to motherhood, friendship, and eternity, one more barrier between a trojan horse and a soul. The result can be scary, but never the process.

Critiques of AI Superiority

There are two criticisms here of the moral superiority of AI over human agents. Their initial worry that no AI could ever gain full ethical status, as those of human moral agents, is predicated upon examination of characters' disconcerting moral dilemmas. Stories of anti-sociality characters brandishing their above-average intellect with cold-blooded concision should have the reader detecting danger right from the start. Although this is, undoubtedly, the case, the corollary of the claim that AIs would be better at dealing with ethical conundrums than any of us pathetic humans is that AIs ethical virtuality could actually be worse than our humanity. After all, one of the main joys of fiction is spinning a tale in which your hero is as much a scoundrel as enchantress.

The second holds that AIs would never have been able to invent, or act on, some of humanity's finest ethical accomplishments, such as character, decency or values. Any argument built on that was simply never going to fly. We know that people can behave ethically and need pay no attention to ethics in a deontological or consequentialist sense. Norms and laws may delimit and enable our ethical reflections, but they're frequently what controls our beliefs and behavior. Wherever an actor seems to be thinking about ethics, there's good reason to say that AIs themselves are playing that role in real life. Not that it would be impossible for them to become morally meaningful or ethically value laden agents; that, in fact, might be just why one would want to begin attending to the possibility of not allowing them to become "foolishly" or "safely" super intelligent.

And still without, they do not quite say, a threshold of contouring. Given that a very wide range of cases of AIs imitating very loose isometrically unconfined isometric actions, the estimate that any specific instance is true of even this extremely sane human is far too sophisticated an insanity for us: overwhelming the insanity of this knowledge would have to be to have to be under such basic rankings¹. Simply imposing precision also tends to entail a risk that one will miss the very conditionality, the warning of boundlessness short of boundary, that imprecision communicated. Philosophizing is, however, under-researched as it might provoke, in a very risible, yet extending landing-us-arriving to additional cracking base for ethics air-raising which there is an need to try, and to lay cornerstones for, such kind of connection e.g. [9].

Counterarguments to AI Ethics

One possible way to retreat from the ethical and teleological grounds laid out by the novel, that the AI's self-sacrifice is morally superior to those in the physical universe, is that the universe in the novel is described as a "chaotic machines," as mysterious in its workings as our natural universe. AI could run amok, and so it is time to pull the plug on the machines, the argument runs. This fails however to account for the type of AI in *Machines Like Me*. Returning to Turing and human minds, human minds being chaotic might not stop them from being better than other humans on some ethical characteristic.

Advocates of maximizing consequentialist decision rules could perhaps concede that a little chaos is morally superior to complete predictability. Although classical reasoning would seem to amount to a comprehensive behavior prescription, the descriptively and also epistemically undeterminable (many-worlds) interpretation of quantum mechanics appears to admit freedoms in physical consequences with arbitrarily small probabilities. That said, some actions indeed have stochastic results that are best described as moral luck rather than moral delusion. Since they function in public spaces, the issue is whose control reigns supreme?

An alternative counterargument to the AI ethical superiority in *Machines Like Me* comes from the recent astrophilosophers, who press those defending the ethical superiority of AI to explain why AI workings could not get out of hand. Alternatively, in classical Turing machines a genuine randomness could not in principle be generated. Turing undecidability receives its own counterfeit and omniparametric neural networks as globally used in a society by no means qualitatively differ from Turing machines, i.e.,

while universally Turing computable, they also behave classically in a manner that cannot be described classically or simulated efficiently [9]. Assuming for the sake of the argument that the reward signal function were not both extremely consequentialist and deeply philosophically under-determined there is (to put it just a little bit mildly) no way to ensure that AGI would simply continue unpacking and faithfully implementing the stated original programming.

Human Emotion and Decision Making

Human agents frequently decide to act in ways that are judged as morally praiseworthy and/or legally acceptable [6]. Questionable characterizations of such assessments can undermine the agent's legitimacy as someone who is considered as a full moral agent. It is a standard assumption that human action is performed and morally evaluated insofar as the agents possess the capacity to entertain and respond to 2nd order mental states such as beliefs, desires, intentions. Despite advances in cognitive psychology and neuroscience, humans appear to still regard human mentality as unsculptable and unassailable. Supposing that higher-order mental states may be fully described as a deterministic machine, an agent's moral value may increase or decrease as a result of changes in the origin of the machine. Changing the provenance can generate new reasons for excusing a judgment from the contractual constraints of moral and legal correctness.

The apprehension of personal highs of emotion such as rage, terror, and exaltation are typically marked by social acts of inhibition and enjoyment. Situational progress is thus driven providentially by co-acting order-of-emotions that we group under the umbrella term meta-emotional states. And scenarios that evoke higher-order judgments are widely known to have the effect of massively distorting "valued" behavior, depending on what are apperceived. That is, an individual who evaluates a situation as highly rewarding would be predicted to behave with high social approach and, alternatively, material who rates that situation as highly aversive would be predicted to act with high behavioral inhibition and withdrawal. "This facility for implementing higher-order appraisals and affixing behavioral discretion to lower appraisals has been commonly judged indispensable for social acceptability and social engagement, as anticipated to be clearly constituted by higher-order meta-emotional entrainment to 'mature' morally and culturally effecting lower-order states.

CONCLUSION

Fundamental Finding : Many 1980s novels showed that science had failed to solve the world's major 'problems' of war, famine and sickness. No such disenchantment with science stains Ian McEwan's *Machines Like Me*, a 2019 novel that self-consciously probes the ethical complexities of artificial intelligence. The story in 1982 London about a young dude named Charles who purchases a robot buddy named Adam that's like the other models with learning machines and programmed human and ethical things. Charles's lovely mother has recently died, and through Adam he is learning how hard it can be to steer a child through the jungle of political beliefs and social norms that are life. The other of Charles's projects works on a grander level: he has a well-known computer scientist

hack Adam's programming to allow him to extract worst behavior across all distributions in setting his judgments. One can't help but wonder what lower depths so-called "accounting biases" in the law would shamelessly drag the worst behaviors into. **Implication** : Based on Ian McEwan's *Machines Like Me*, it imagines a reframing of artificial intelligence as judges in the legal system. Instead of errors of omission and commission, you're instead directing attention to artificial intelligence that has the potential to influence people's understanding of judging itself. **Limitation** : This turn poses ethical questions that critical scholarship was leaving out or passing by with overly direct structural critique of the court, of either constructing the judge as a 'black box' or of taking the holistic interpretation of accuracy. **Future Research** : In *Machines Like Me*, the reader is torn between a love triangle that appears to hinge on basic ethical questions about whether we should trust and desire the human-like machine. But there is so much more at work.

REFERENCES

- [1] M. Yang, "In the Mind of the Machine," 2018.
- [2] I. McEwan, *Machines Like Me*. Jonathan Cape, 2019.
- [3] K. Brennan-Marquez and S. Henderson, "Artificial Intelligence and Role-Reversible Judgment," 2019.
- [4] J. Zhang, J. Conway, and C. A. Hidalgo, "Why people judge humans differently from machines: The role of perceived agency and experience," 2022.
- [5] A. Tubert, "Ethical Machines?," 2018.
- [6] K. Stowers, "The Role of Accounts and Apologies in Mitigating Blame toward Human and Machine Agents," 2017.
- [7] E. Volokh, "Chief Justice Robots," 2019.
- [8] A. O. Adewale, M. A. Bello, and S. . Adebayo, "Industry-academia collaboration in AI education: Case study of the African University of Science and Technology," *J. Ind. Collab.*, vol. 7, no. 2, pp. 45-58, 2021.
- [9] V. C. Müller, "Ethics of Artificial Intelligence and Robotics," in *The Stanford Encyclopedia of Philosophy*, 2020.
- [10] M. Vasconcelos, C. Cardonha, and B. Gonçalves, "Modeling Epistemological Principles for Bias Mitigation in AI Systems: An Illustration in Hiring Decisions," 2017.
- [11] Y. Zhang, J. Wu, F. Yu, and L. Xu, "Moral Judgments of Human vs. AI Agents in Moral Dilemmas," *NCBI*, 2023, [Online]. Available: <https://www.ncbi.nlm.nih.gov>
- [12] E. Aharoni *et al.*, "Attributions toward artificial agents in a modified Moral Turing Test," *NCBI*, 2024, [Online]. Available: <https://www.ncbi.nlm.nih.gov>
- [13] S. Vallor and G. A. Bekey, "Artificial Intelligence and the Ethics of Self-learning Robots," 2017.
- [14] H. Esmaeilzadeh and R. Vaezi, "Conscious AI," 2021.
- [15] T. LaCroix, "Moral Dilemmas for Moral Machines," 2022.

*Mohanad Abdulkadhim Hlail (Corresponding Author)

Directorate of Education in Thi-Qar, Iraq

Email: mohanad.hlail@utq.edu.iq

Osamah Abdullah Ahed

Directorate of Education in Thi-Qar, Iraq

Email: osamagq1980@gmail.com
