Review Article

Black Mirror and Narratives of Technological Malaise: Analysing Contemporary Science Fiction Tropes in Terms of the Potential of Artificial Intelligence (AI) Technology to Harm or Benefit Human Culture

Terry Hyland[1]

1. Free University of Ireland, Dublin, Ireland

There can be little doubt that we are in the midst of a golden age for dystopian science fiction (SF), given added impetus in recent times by the public availability of artificial intelligence (AI) tools such as Chat-GPT, Oracle, Gemini and Bard, and the wealth of both academic and popular literature which such developments have generated. Charlie Brooker's dark and edgy SF series of Black Mirror films – now in its seventh iteration on Netflix – exploits heavily our fears about the dangers of AI technology threatening human culture in every sphere of activity from work, to education, to health care, and in social relationships. The question of whether this is pure, thrill-seeking entertainment – like stories about vampires and demons which have little or no impact on real life – or whether it is intended to warn us against the potential dangers of emerging technology and thus help us to establish safeguards to protect our cherished values and ways of life is a moot question. The persistent dystopian emphasis about AI dangers throughout the Black Mirror films suggest that there is indeed a serious intention behind the stories. After examining some of the core themes – and cognate connections with SF in general – this article goes on to consider whether this futuristic doom-mongering, especially about AI, is harmful or beneficial for present and future human culture in all its forms.

Corresponding author: Terry Hyland, hylandterry@ymail.com

1. Science Fiction Dystopian Visions

Very few science fiction (SF) stories about human contact with extra-terrestrial civilisations present such contact in a favourable light. It is true that the *Star Trek* series did include some fruitful collaborations between humans and Vulcans (in the form of Mr Spock) and indeed with non-human robots (Mr Data) but – from H.G Wells' *War or the Worlds* to *The Thing* and *Independence Day* – doom-mongering in the area of contact with other civilisations has been a standard trope. Moreover, such dystopianism is even more in the foreground when it comes to stories and films which feature robotic or artificial intelligence (AI) themes. As Adam Thierer[1] has expressed this emphasis:

> Dark depictions of AI and robotics are ever-present in popular modern sci-fi movies and television shows. A short list includes: ***2001: A Space Odyssey***, ***Avengers: Age of Ultron***, Battlestar Galactica (both the 1978 original and the 2004 reboot), Black Mirror, Blade Runner, ***Ex Machina***, ***Her***, ***The Matrix***, ***Robocop, The Stepford Wives, Terminator***, T***ranscendence***, ***Tron***, ***WALL-E***, ***Wargames*** **and** ***Westworld***, among countless others. The least nefarious plots among these films and television shows rest on the idea that AI and robotics are going to drive us to a life of distraction, addiction or sloth. In more extreme cases, we're warned about a future in which we are either going to be enslaved or destroyed by our new robotic or algorithmic overlords (p.1).

doi.org/10.32388/VC9ZBO

Clearly, much of the horror depicted in such films is motivated by the same intention to excite, thrill and frighten audiences as that which inspires stories of werewolves and vampires. However, there is increasing concern amongst scientists, philosophers, and technology writers that – in the case of current AI trends – some of this pessimism and dystopianism may be justified. A poll of scientists conducted by the Scientists for Global Responsibility (SGR) published in 2018 reported that:

> Over 80 percent believe there's a medium to high chance of things going badly wrong with Artificial Intelligence (AI)...while 96 percent say AI needs more regulation, and 82 percent thought that AI was most likely to create a dystopian rather than a utopian future (p.1).

Moreover, since that poll was conducted, AI tools – developed by Open AI, Google, Meta and other big tech companies – have been made widely available to the public..Although AI applications have been with us for decades – in phones, cars, banking, medicine, and the like – it was the appearance of the publicly accessible Chat GPT and similar applications and tools in November 2022 that has generated much intense and unrelenting academic and popular interest. In the last three years – as Open AI has developed Chat GPT- 4 and 5 with similar advanced tools emerging from Google, Meta, and Amazon – the breadth and range of such tools has impacted almost every public and private sphere including health, education, entertainment, working life, the economy and general culture. Hundreds of articles, both academic and popular[2] (57,000 publications by 2024) have appeared in the last few years, and the field has recently been scrutinised by Emily Bender and Alex Hanna[3] and their division of opinions and judgments about AI into 'boomers' (pointing to potential benefits) and 'doomers' (warning about potential harms) is instructive and helpful in making sense of this exponentially growing area of debate. Such controversies will be discussed later in order to provide a background to the analysis of *Black Mirror* and recent similar explorations of the impact of AI on all aspects of human life.

2. Charlie Brooker and Black Mirror

The line between doom-mongering satire and the portrayal of futuristic horrors as pure entertainment is necessarily fuzzy and porous, and this is well illustrated in Charlie Brooker's own account of his motives and inspirations in relation to the making of the *Black Mirror* (Channel 4, 2011-2013; Netflix, 2014-2025) series of films. In his book co-written with Annabel Jones, the executive producer of the first series, Brooker cites shows such as *The Twilight Zone* and *Tales of the Unexpected* as inspirations for the key ideas intended to be 'moral parable stories around the theme of social media, technology and AI advances'[4].

From the start, it appears that Brooker favoured the more pessimistic and negative aspects of new technological advancements. As he comments:

> Being a paranoid person, as soon as I see any advert where everyone's happy and smiley I immediately think it's a bit like a sinister advert in a dystopian movie...The fact that it looked so happy meant that it couldn't last, so I was immediately unsettled by that (ibid., p.199).

More recently, Brooker has spoken of the influence of his reading of the film *The Truman Show* and addressed the phenomenon of the uncanny capacity of *Black Mirror* to match the contemporary zeitgeist. As he observes:

> The soothsaying aspect comes up quite often. Unfortunately, I seem to have an early alert system for people sending me things like, "Have you seen this news story? This is quite Black Mirror." And all I can say is, "Well, I'm just exaggerating something I've seen." And it's just the way that, unfortunately, the world works. It seems that things get worse, so reality did catch up. But the timing of the episode ('Joan is Awful', Series 6) was crazy. Sometimes, with a story, you're like, I have to write this before someone else does this. At the time of writing, ChatGPT hadn't blown up yet. But, you know, you type something into it and get an immediate first reaction: fear. Because it's doing a convincing impersonation of coming up with thoughts, the more you play with it, the more you see how limited it is[5].

The principal idea expressed here that it is not the new technology, social media or AI tools that leads to cruelty, harm and suffering but the uses of all this in the hands of perverse and malignant humans. This general view is confirmed in a recent Brooker interview for the BBC publicising the latest series which he intimates 'could just run and run'[6]. He comments that:

> there are also aspects of AI I probably don't understand, and that could be used as tools for good. It's about everyone being able to understand them and to use them correctly, and them being in the right hands...I can totally see the value of AI as a tool for creative people. The point at which it worries me is if you remove the people bit from that equation, or you're just hoovering up their work and regurgitating it, and they're not being paid (ibid., p.2).

As will be noted later, such fears that AI and the new technology might be misused if or when it falls into the wrong hands are neither unrealistic nor misplaced as recent concerning developments have demonstrated.

3. Philosophical Explorations

In addition to the popular cultural discourse about the *Black Mirror* series, there has been considerable philosophical interest and attention paid to the phenomenon. In the collection of reflections on *Black Mirror* by leading philosophers edited by Kyle Johnson[7] we find the assertion that 'ultimately *Black Mirror* does what science fiction does best: philosophy'. This is explained more fully in the observation that:

> It's not just that ***Black Mirror*** tells stories that philosophers might find useful for demonstrating philosophical ideas – though it certainly does that. According to contemporary philosophers Thomas Wartenberg, for example, fictional media (like ***Black Mirror***) can illustrate philosophical theories or serve as thought experiments – imaginary situations that philosophers often use to reveal our philosophical intuitions or refute philosophical theories using counter-examples. And that certainly counts as doing philosophy. Indeed, philosophers have been using counter examples and thought experiments, much more far-fetched than any *Black Mirror* episode, for centuries (Kindle edn., loc.940).

The thought experiments referred to here would include Plato's Cave, Descartes' Evil Demon and, in more contemporary writings, philosophical zombies, the brain-in-a-vat, and the idea that the world is a simulation[8].

The very title chosen for the series reflects both the principal intentions and the key inspirational themes of the series. Elaborating upon Brooker's statement that the "'black mirror of the title is the one you'll find on very wall, on every desk, in the palm of every hand: the cold, shiny screen of a TV, a monitor, a smartphone", Johnson observes that:

> When you watch Black Mirror, you're watching a dark reflection of society – one that is just slightly cracked – that depicts our flaws, our fears, and our possible future[7].

Similar perspectives on the series are offered by Alexander Schnell[9] who insists that Black Mirror is 'a philosophical work' (p.7) and suggests that the episodes

> are of considerable philosophical importance, particularly in fathoming out the meaning and status of reality. In addressing this philosophical problematic, Black Mirror looks at the the basic features of the present age (p.14).

Schnell goes on to explore some of the principal ideas and themes of the series by reference to the philosophical canon. Drawing inspiration from Kant, Fichte and, in particular the 'techical-cyberneric' element in Heidegger's work, it is argued that:

> In the world of Black Mirror, 'virtuality' refers to the combination (based on 'transcendentality' in Kant, Fichte and Husserl) of the 'real', the figurative 'imaginary', and the 'technical'...***virtuality* is *reality***! The mental acrobatics needed here – namely that reality is *at the same time* a constituent part (i.e. in a sense part of the whole) of virtuality and 'the same' as it (i.e. also the whole itself) – is what ***Black Mirror*** demands of us in order to attain a proper understanding of reality[9].

In later chapters, Schnell goes on to consider – with reference to a wealth of philosophical sources drawn from phenomenology, psychoanalysis, psychology and metaphysics – such themes as time, love, death and nature as these concepts are dealt with in the series. Some of this material will be utilised below in the consideration of specific *Black Mirror* episodes and key themes.

3.1. Black Mirror: Appearance and Reality

As mentioned earlier, there is a philosophical consensus about the idea that a key *Black Mirror* theme is concerned to challenge our perspectives about what reality actually is, on the links between the imagination (especially as this is expressed digitally through film) and the so-called 'real' world. As Schnhell[9] puts it, the 'series sets out to depict the *point of transition* between visible reality and underlying virtuality – insofar as it is reality-forming'[9].

The philosophical discourse about what is real goes back at least 2000 years as far as Plato who constructed his 'allegory of the cave' to provoke reflection about whether what we take is the real world might actually be mere appearance. The Allegory (or Simile) of the Cave in Plato's *Republic*[10] – whereby prisoners are imprisoned in a cave and are only able to see the world by looking at shadows of people

passing thrown upon the cave wall by a fire situated behind their fixed position – is one of the earliest example of the philosophical idea of the 'two worlds' thesis. Basically, the core argument – marking a distinction between appearance and reality (an especially moot one in the context of contemporary simulation notions and quantum metaphysics) – is that the world we see around us is an imperfect reflection of an infinitely more perfect, more accurate and truthful world, which may be accessed through initiation into certain forms of knowledge and reasoning. In a similar sort of sceptical thought experiment, in his search for absolute truth, Descartes[11] vowed to 'reject as being absolutely false everything in which I could suppose the slightest reason for doubt, in order to see if there did not remain after that anything in my belief which was entirely indubitable' (p.53). Using this method which questioned all hitherto accepted and established beliefs in science and philosophy, Descartes arrives at his famous *Cogito ergo sum* – I think, therefore I am – which was thought to be 'so certain and so self-evident that all the most extravagant suppositions of the sceptics were not capable of shaking it, I judged that I could accept it without scruple as the first principle of the philosophy I was seeking'[11].

Having established what he takes to be an incorrigible fact about his existence as a thinking being, Descartes then extends his sceptical doubt to the possibility that he might be dreaming the fact (the 'life is a dream' idea is a popular fictional trope in science fiction) that he is thinking but dismisses this objection since he would need to be thinking and existing to entertain this possibility. However, suppose that he was being deceived by an all- powerful evil genius (think 'deepfake' in modern digital terms) about everything he perceives and believes. As he expresses it:

> I shall suppose, therefore, that there is, not a true God, who is the sovereign source of truth, but some evil demon, no less cunning and deceiving than powerful, who has used all his artifice to deceive me. I will suppose that the heavens, the air, the earth, colours, shapes, sounds and all external things we see, are only illusions and deceptions which he uses to take me in[11].

Descartes eventually resolves the issue through the (much criticized and controversial) claim that he has constructed an argument for a benevolent God which validates the foundation of knowledge.

In contemporary philosophy, the simulation argument advanced by Nick Bostrom can be regarded as the modern descendant of the challenges issues by Plato and Descartes. Bostrom's original paper[12] has an abstract which usefully sums up the main argument.

> This paper argues that at least one of the following propositions is true: (1) the human species is very likely to go extinct before reaching a "posthuman" stage; (2) any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof); (3) we are almost certainly living in a computer simulation. It follows that the belief that there is a significant chance that we will one day become posthumans who run ancestor-simulations is false, unless we are currently living in a simulation[12].

Bostrom [24] adds the following gloss to his original statement of intentions:

> Many works of science fiction as well as some forecasts by serious technologists and futurologists predict that enormous amounts of computing power will be available in the future. Let us suppose for a moment that these predictions are correct. One thing that later generations might do with their super-powerful computers is run detailed simulations of their forebears or of people like their forebears. Because their computers would be so powerful, they could run a great many such simulations. Suppose that these simulated people are conscious (as they would be if the simulations were sufficiently fine-grained and if a certain quite widely accepted position in the philosophy of mind is correct). Then it could be the case that the vast majority of minds like ours do not belong to the original race but rather to people simulated by the advanced descendants of an original race. It is then possible to argue that, if this were the case, we would be rational to think that we are likely among the simulated minds rather than among the original biological ones. Therefore, if we don't think that we are currently living in a computer simulation, we are not entitled to believe that we will have descendants who will run lots of such simulations of their forebears. That is the basic idea. The rest of this paper will spell it out more carefully[12]

Since Bostrom advanced his original arguments in 2003, computing power – especially the development of quantum computing – has come on by leaps and bounds to the extent that the simulation argument now seems reasonable, plausible and, perhaps, possible.

The astrophysicist and Nobel prize winner, George Smoot, has stated boldly that 'you are a simulation and physics can prove it'[13]. Smoot's key arguments range from the fact that our brains did not evolve to discriminate clearly between reality and illusion, that it will soon be possible to upload a complete human brain to a computer, and that inconsistencies in physics (in particular, quantum mechanics and general

relativity) are exactly what we would expect in a simulated world. In the same vein, Jenkins[14] – utilising ideas proposed by the futurist Ray Kurzwiel – is confident that the simulated hypothesis is almost certainly correct. Jenkins suggests that:

> by 2050 it would be feasible to have a completely realistic historical simulation running on every desktop, and that these simulated worlds would outnumber the real one by a factor of millions or even billions to one. This makes it almost certain that we live in one of the simulations if a future society has the motivation to create them.[14]

All such speculation – along with intense philosophical debates about the nature of reality inspired by quantum physics and attempts to solve the hard problem of consciousness[15] – has provided fertile material for science fiction films such as the *Matrix*[16] and there are clear and direct parallels here with some central themes of *Black Mirror* discussed in the next section.

4. Some Key Black Mirror Themes

The key theme of challenging reality discussed above needs to be interpreted within the framework of the ways in which new technological developments, AI, social media, and the proliferation of digital applications have impacted human culture and behaviour in all spheres. The Black Mirror series repeatedly poses questions about who controls the technology, who profits and who pays the costs. AI and social media apps are used to demonstrate how the amplification of human impulses – such as jealousy, desire, validation, revenge, control, security, love – can distort thought and behaviour in ways which produce new forms of oppression, loss of identity, fear, isolation, and a whole host of ethical dilemmas. Social media is portrayed not just as a new tool but as a catalyst that magnifies human weaknesses and structural socio-economic and political issues such as surveillance, attention extraction for profit, reputation economies, group hatred and cruelty, and algorithmic control[17]. Depictions of such issues often show how ordinary, natural emotions and desires – for approval, fame, ease, love, and so on – may become catastrophically harmful when mediated by the new technological tools. It would be useful to illustrate some of these features by reference to specific episodes.

4.1. Black Mirror Themes

Performative Identity and Reputation

In *Nosedive* (series 3, 2016) we are shown how a social-rating app makes every interaction a transaction which receives a score. The upshot is a society in which people perform to achieve higher ratings thereby intensifying widespread social anxiety, simulated behaviour, social stratification based on rating scores. The central character, Lacie, becomes obsessed with improving her ratings, a malaise which ruins her life demonstrating how the suppression of natural, authentic relationships can distort human behaviour and foster social harms and cruelty. Schnell[9] characterizes the world of Lacie Pound in this episode as one in which 'only pseudo-popularity counts' (p.33), a world which eventually almost destroys her as she comes to rebel fiercely against the illusory reality of her life. This can be seen as an overt criticism of the crude algorithmic demands of social media apps for summary like/dislike judgements which – though shallow and transient – can have a devastating impact on the mental health of online subscribers.

The attention economy and attention addiction

In *Smithereen*s (Series 5, 2019) an app company's business model and the obsession with fame and being noticed generated by the new culture is shown to drive reckless behaviour ending in personal tragedy. Similar themes are explored in *The Waldo Moment* (Series 2, 2013) and *Fifteen Million Merits* (Series1,2011) which satirize celebrity-for-entertainment tropes that mimic influencer dynamics and talent-show shallowness and pseudo-vitality. Such themes – obviously inspired by films such as The Truman Show, and real-life celebrity zand talent shows such *as Big Brother* and the *X Factor* – seek to demonstrate an impoverished culture, stripped of ordinary human values such as compassion and respect for persons, which values transient fame and celebrity for its own sake. Schnell[9] describes these worlds as places where 'reality and singularity are suppressed and replaced by a superficial world of appearance and meaninglessness', a scenario which 'leaves the viewer with disgust and revulsion' (p.30). In the harm inflicted on social media addicts – especially young people – is now well documented in real-world studies. Use of generative AI can lead to a loss of critical thinking skills and the capacity for open-ended, flexible thought and speech[18], and the links between high social media exposure and feelings of loneliness, isolation and depression have been established in a number of studies[19].

Surveillance, Control and Manipulation

Men Against Fire and *Hated in the Nation* (Series 3, 2016) addresses the ways in which humans may be psychologically manipulated by technological apps and implants. The episodes demonstrate how relatively easy it is for malign and autocratic governments to incite public hatred and violence to achieve grotesque self-serving ends. This is not far from real-life wartime propaganda, and we have current examples of this in the totalitarian control of information and messaging in Russia and Israel in support of unilateral war aims. Schnell[9] links these general ideas with other episodes such as *Metalhead* (Series 4, 2017) and *Mazey Day* (Series 6, 2023) as examples of the mass Orwellian control of all forms of communication, a process which is greatly assisted by developments in the current digital age. It is a classic Black Mirror theme – the loss of individual identity and agency in the face of an all-encompassing and relentless mass media presence – and it is without question a dystopian trend which the film series is asking us to understand and challenge[20].

Consumerism and Capitalism

Critiques of mindless consumerism in a neo-liberal capitalist economy can be glimpsed in many Black Mirror episodes but are foregrounded in a number of episodes which portray the lethal cocktail of the merger of unfettered consumer capitalism with unregulated social media and surveillance algorithms. *White Christmas* (Series 2, 2014),for instance, demonstrates how personal data – which we all give willingly, unwillingly but mostly unthinkingly to the big tech companies – is commodified and manipulated by algorithms designed to exploit human vulnerabilities. Consumerist greed thus comes to dominate and mutate all aspects of everyday existence, reaching its apex of awfulness in *Fifteen Milliion Merits* (Series 1, 2011) in which individuals compete for merits earned by performing tedious and monotonous work. The further irony is that such merits are then spent in ways strictly determined by the consumerist exploitation which motivated the activity in the first place. Yanis Varoufakis[21] has described these developments as a new post-capitalist economic arrangement which he labels ‘technofeudalism’ which makes serfs of us all as – through our captivity by the new technology – as we labour without pay providing valuable data under the fiefdoms of the tech barons. This theme will be discussed further in the conclusion.

4.2. Dystopian Spin-Offs

The key themes of *Black Mirror* – especially the distortion of everyday life by AI technology, social media, and algorithmic surveillance and manipulation of basic human needs and emotions – have parallels in a number of other films, books, podcasts and video platforms. The *Altered Carbon* series (Netflix, 2018,2020), is set in a future in which in consciousness and memories can be downloaded onto cortical stacks and transferred from dead to living bodies. The unsettling mixture of the real and the virtual here is reminiscent of the *Black Mirror* episode 'USS Callister' (Series 4,2017) in which a game manufacturer is able to create digital sentient clones of real people. There are clear parallels here with the virtual reality (VR) developments planned by Meta involving VR headsets and glasses[22], and direct links with the AI deepfakes currently flooding cyberspace. Similar themes are explored in the classic Philip K Dick's *Electric Dreams*[23] which inspired a popular Channel 4 series (2017-18).

Perhaps the most popular and influential SF tech series in recent years has been *Years and Years* (BBC, 2019) which follows a single family through momentous changes characterized by the climate catastrophe, growing political authoritarianism, and elements of social and moral decay amidst the rapid advance of new AI technology. Unlike *Black Mirror*, the narrative in *Years and Years* is based on real-life relationships within a recognizable political and socio-economic milieu. For this reason, the serious harms to individuals and society generated by the technological trends – the enormous power of tech companies and their influence on politics, the over-reliance on smart phones and screens for information, the blurring of the real and the virtual, the spread of deepfakes and conspiracy theories, and the manipulation of the lives of ordinary people through corporate algorithms – seem all the more heinous and unforgivable. If only some of these dystopian warnings are plausible, challenges need to be mounted and safeguards established as soon as possible.

5. Perspectives on Recent AI Developments

As mentioned earlier, hundreds of articles, both academic and popular have been published in recent years and – since there is no space to rehearse all the key findings here (I have written elsewhere on specific AI matters, [24][25][15]) – it would be useful to examine issues under the headings outlined by Emily Bender and Alex Hanna of 'boomers' and 'doomers'[3].

5.1. Boomers

On the positive side benefits of using AI tools have been outlined a vast range of areas such as business, education, health, science, transport, defence, work and just about every aspect of social life[26]. Many of such alleged AI gains tend to be located within business and work environments within which there are claims about increases in productivity through the automation of repetitive tasks and processes, more efficient management and team working, and the acceleration of data collection and analysis. As will be argued below, all of these putative benefits have serious negative downsides linked to job losses, discrimination, misinformation and general social harm. However, on the boomer side it is worth mentioning some clear AI gains and achievements.

In medicine and health care, research undertaken by the *World Economic Forum*[27] has reported a wide range of AI benefits including the following (pp.2-4):

- new AI software is "twice as accurate" as professionals at examining the brain scans of stroke patients
- AI can spot more bone fractures than humans can
- A new AI machine learning model can detect the presence of certain diseases before the patient is even aware of any symptoms
- A new brief from the World Health Organization (WHO) and partners, 'Mapping the application of artificial intelligence in traditional medicine', shows how AI can enhance traditional, complementary and integrative medicine (TCIM) while protecting cultural heritage.

The impact of Ai on education and academia in general has been far-reaching[25]. On the optimistic interpretations, there are claims that machine learning better facilitates personalized learning, and generally supports 'the promotion of learning and wellbeing of students, teachers, and other educational stakeholders'[28]. In a similar vein, in relation to the central topic of sustainable development, there are optimistic claims about the potential of AI in alleviating global warming, pollution and the degradation of the planet[29].

5.2. Doomers

In their recent critique of the wide range of discourse about AI, Bender & Hanna[3] point out the intriguing state of affairs in which – in what they describe as the topsy-turvy world of 'AI hype and con' – boomers and doomers are often in the same camp. As they put it:

> Scratch a doomer and find a boomer...Doomerism and boomerism are supposedly diametrically opposed camps, but both see AI as inevitable and desirable (p.148).

The explanation for this surface irrationalism is to be found in the need for the big tech companies to maintain the hyperbolic publicity about the tremendous power of their AI tools. Thus, Sam Altman of *Open AI*[30] may freely admit that AI will cause substantial job losses, pose threats to national security and wreak general havoc on human communications, but he does so in a context in which AI development is said to be happily advancing from strength to strength.

Moreover, many of the more grandiose concerns expressed by doomsters about the potential for superintelligent AI robotics to destroy human civilisation are, for Bender & Hanna[3], pure science fantasy (no doubt boosted by some of the SF dystopianism being examined here!) designed to distract us from the real harms that AI is inflicting on humanity right now not in some sci-fi future. As they comment:

> The danger is not from some hypothetical extinction-level event. The danger emerges from rampant financial speculation, the degradation of informational trust and environments, the normalization of data theft and exploitation, and the data harmonization systems that punish people who have the least power in our society by tracking them through pervasive policing systems. But the Doomer/Boosters would have us looking the other way from all these real harms bedazzled by their dystopian/utopian visions (pp.151-2).

Deciding not to look the other way requires individuals and organisations to scrutinize very carefully the current Open AI, Google and Amazon and other AI developments, to monitor the claims made, and to challenge any potentially harmful, anti-social or retrogressive outcomes. An important first point to make about the new AI large language models (LLMs) such as Open AI's Chat GPT series, Google's Gemini and Microsoft's Llama models is that they are neither rationally intelligent nor capable of human-like thinking and reasoning. The IBM site[31] defines them as follows:

> Large language models (LLMs) are a category of deep learning models trained on immense amounts of data, making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks. LLMs are built on a type of neural network architecture called a transformer which excels at handling sequences of words and capturing patterns in text (p.1).

Bender & Hanna[3] insist that they should be called by their true names as 'text-extruding machines' (p.17) which fool the unwary into thinking they have some human-like understanding because they can answer complex questions in human language.

The philosopher, John Searle's famous 'Chinese Room' experiment serves to warn us about making such potentially disastrous errors. Searle (2004) argues that the Chinese Room argument – dismantling the claim that mere computation or information-processing, no matter how complex or stunningly fast, equates to intelligent reasoning and understanding – 'strikes at the heart of the strong AI project' (p.43) and he has been able to defend his position against some key philosophical objections. Arguing along similar lines, Noam Chomsky[32] has commented on the 'false promise of ChatGPT' arguing that – although such applications are 'marvels of machine learning' – the science of linguistics and epistemology indicate that 'they differ profoundly from how humans reason and use language' (p.3). The philosopher Philip Goff[33] argues forcefully that 'ChatGPT can't think – consciousness is something entirely different to today's AI' (p.2). In addition, the Nobel Prize winning physicist, Sir Roger Penrose, has stated unequivocally in recent lectures that AI has been misunderstood and will never achieve human-like intelligence or consciousness[34]. In a similar vein, the physicist, Carlo Rovelli, has debunked the AI myth about super-intelligence and called for more human as opposed to artificial intelligence[35].

Notwithstanding such philosophical and scientific critiques of AI, the doomsayers (some of them such as Geoffrey Hinton, see Heaven[36], with inside knowledge of AI companies) still rightly wish to warn us about the threats and dangers of uncontrolled and unregulated developments in this rapidly developing domain, and it is just here that *Black Mirror* and its spin-offs may provide an extremely valuable public service. Some of the principal issues are discussed in the next, concluding section in which the key threats – and ways of countering them – are examined against the background of the dystopian scenarios.

6. Conclusion: Black Mirror and AI Doomers, Boomers, and Pragmatic Ethics

Although Charlie Brooker and *the Black Mirror* team appear, on the surface, to have achieved a reasonably happy balance between offering chilling, thought-provoking entertainment and prophetic doom-laden warnings about our current technological malaise, this balance breaks down on closer inspection to reveal pessimistic forebodings. In direct reversal of Bender and Hanna's claim about AI hype in general –

scratch a doomer and you will glimpse a boomer[3] - it can be argued that *Black Mirror's* superficial boomerism (almost mirroring the pro-tech hype) about the joys and benefits of technology in terms of love, pleasure, social esteem, scientific progress, and so on, is really a mask which serves to make the dire and urgent warnings about grave potential harms to humanity more palatable and effective.

Almost every negative and harmful outcome of AI technology, social media, and the manipulation of reality featured in the series can be, to a greater or lesser degree, discerned in contemporary culture, economics, politics and general social life. As mentioned earlier, Bender & Hanna's[3] critique of AI serves to bring out the widespread harms of the new applications in real time, in the here and now, rather than in vague SF speculations about the dangers to humankind of super-intelligent robots taking over our lives. We can begin by highlighting the negative downsides of the boomeristic claims about AI in a number of salient fields.

There are claims, for example, that AI can be an invaluable tool in managing the transition from fossil fuel to net zero socio-economic conditions. As Stern, et al[29] argue:

> The net-zero transition is not simply a mitigation strategy and should not be thought of as a "cost", but rather a great opportunity for innovation and sustainable, resilient and inclusive economic growth...Artificial intelligence (AI) is well-positioned to accelerate this transition and, as general-purpose technologies, AI can be applied to speed up this process of profound systems' transformation by increasing the speed, efficiency, and effectiveness with which innovation processes are scaled and capital is deployed (pp.12-13).

The ongoing impact of the rapidly growing climate catastrophe means that it is a moral and pragmatic imperative to challenge such claims of AI beneficence with scientific evidence. There is now a wealth of evidence that the energy-hungry expansion of massive data centres required by AI development will cause an exponential increase in greenhouse gas emissions. As *EuronewsGreen* noted recently, 'Data centre electricity demand is predicted to rise to around 945 terawatt hours by 2030, more than the entire electricity consumption of Japan'[37]. Bender & Hanna[3] fully endorse this position on AI energy demands, stating unequivocally that 'AI is hastening the climate catastrophe' (pp.156ff.). Moreover, adding insult to injury, the many errors and fantastically mistaken citations and judgments – quaintly labelled 'hallucinations' by the tech industry – will require even more energy in the attempt to remove them from LLMs. Wei Xing[38] explains that this is because the reduction of AI hallucinations requires

substantially more 'training' – that is massive trawling of data sets, often infringing copyrights of all kinds – and this is extremely energy intensive' (p.1-2).

It will be important to foreground such messages in order to explore ways of challenging the apparently unrelenting doom spiral towards the destruction of the planet. As Arshin Adib-Moghaddam[39] urges, 'the public needs to challenge the good AI myth pushed by tech companies'. He elaborates in commenting:

> the public is at the receiving end of a distinctly hierarchical top-down system, from the big tech companies and their governmental enablers to users. In this way, we are made to consume, with little to no influence over how the technology is used. This positive AI ideology is therefore primarily about money and power. As it stands, there is no global movement with a unifying manifesto that would bring together societies to leverage AI for the benefit of communities of people, or to safeguard our right to privacy (p.2).

There are many more reasons for challenging the 'good AI myth', not least – as highlighted throughout the Black Mirror series – because of the deleterious impact the new technology and its attention-hungry algorithms is having on the mental health of individuals, communities and the general culture.

Against the claims that AI can enhance educational development, there is growing evidence that over-reliance on AI learning tools may erode crucial cognitive abilities and critical thinking skills[18]. Moreover, the ubiquitous and all-embracing social media platforms facilitated by the new technology show every sign of having negative consequences for the mental health of users. The most harmful impacts are felt by the younger age groups who use the platforms more than older groups. Investigating this demographic, research by Khalaf et al[40] reported 'privacy concerns, cyberbullying, and bad effects on schooling and mental health are all risks associated with this population's usage of social media' (p.1). In the same vein, the meta-analysis of research in the field by Kenta Minamitani at Stanford[41] concluded that:

> The link between social media and mental health issues has been well documented in numerous studies and research papers. A systematic review found that the use of social networking sites is associated with an increased risk of depression, anxiety, and psychological distress. The associations, though not by itself proof of causation, at least some reason for concern. Additionally, this association is particularly strong in adolescents compared to younger children. Moreover, in the United States, the 12-month prevalence of

> major depressive episodes among adolescents increased from 8.7% in 2005 to 11.3% in 2014 The new media screen activities have been suggested as one of the causes of the increase in adolescent depression and suicide (pp.2-3).

Related research by Lucy Osler[42] has reported that 'AI-induced psychosis' generated by over-reliance on chatbots by mentally vulnerable people has led in some cases to violence and murder. Given all such findings it is little wonder that there are strenuous efforts in Britain to ban the use of smartphones in schools and – following tragic cases of adolescent suicides – to severely restrict the content available to youngsters on social media platforms[43].

All the new technological developments discussed above are located within a neo-liberal capitalist framework in which ever-expanding consumer spending is the key driver of algorithmic design and implementation. Just as the internet itself was initially designed by Tim Berners-Lee as an open access, free discussion platform, it was inevitable that in a capitalist economy it would be captured – much to the deep dismay of its originator[44] – and grossly mutated into a voracious consumerist platform so the original non-profit projects of Open AI and other companies were quickly taken over in the squalid money-making scramble to cash in on AI mania to grab as much of the profits as possible[45]. Open AI is currently launching an erotic adults-only platform[46], and its rivals search for similar new terrain in the race for a piece of the lucrative market which now – thanks to the inflated bubble created by venture capitalist interest in AI – has a combined value exceeding £2 trillion. However, the prevailing market analysis is that this bubble – like the similar dot.com bubble in 2000 – is about the burst with huge costs to taxpayers as well as businesses[47]. This pessimist message has been endorsed more recently by the Google CEO, Sundar Pichai[48] who has described the 'irrationality' of the AI boom and has warned that if/when the bubble bursts 'no company is going to be immune, including us' (p.1).

As seriously concerning as this dire prognosis is Pichai's open admission that AI is 'prone to errors' which is potentially more worrying as is his advice that people should always use AI tools alongside other standard research apps since LLMs cannot be trusted (p.2). This warning and its dangerous implications for all aspects of culture and society are, as outlined earlier, echoed by many observers – both from within the tech industry – and from the fields of science, philosophy and general humanities[15]. In addition to all this, the economic and social dominance of the main technological corporations has far-reaching and potentially harmful implications for the whole political, cultural and moral fabric of society. In this area, Yanis Varoufakis[21] presents convincing evidence that the rise of tech company dominance

over all aspects of social, political, economic and cultural life – and the overwhelming power of tech barons such as Elon Musk, Jeff Bezos and Mark Zuckerberg – has resulted in a post-capitalist state of affairs that he calls 'technofeudalism'. As he explains:

> Markets, the medium of capitalism, have been replaced by digital trading platforms which look like, but are not, markets, and are better understood as fiefdoms. And profit, the engine of capitalism, has been replaced with its feudal predecessor: rent. Specifically, it is a form of rent that must be paid for access to those platforms and to the cloud more broadly. I call it ***cloud rent*** (original italics; p.34).

Moreover, the cloud capitalists are able to extract more profit than their more orthodox predecessors because they have absolute control over all their platforms which traders have to use to sell products. In addition, all users, all of us who click on Amazon, Facebook or any other social media platform are working for these 'cloudaliists' for free by providing them with our personal data as well as our business.

Discussing ways of escaping from this serfdom, Varoufakis recommends democratising the cloud through strategic 'rent strikes' and state regulation to control content and scale. Unlike the USA and Britain, Europe now has a reasonably robust AI regulation system[49] – its effectiveness can be discerned by how much the big tech platforms object to it – and similar systems are worth implementing globally, along with the increasing number of legal challenges designed to bring the tech companies down to earth and under democratic control[50][51]. To be sure, large and vastly wealthy corporations are – like oil tankers – not easy to turn around or change direction. However, a combination of people power and political will can be most effective as is illustrated in the recent Australian ban on the use of social media platforms such as Tik Tok, Snapchat, Facebook and Instagram by youngsters under the age of 16[52].

Charlie Brooker's *Black Mirror* team has provided us with a treasure trove of imaginative and endlessly provocative ideas about what the future of AI and related technology might have in store for us in the not too distant future. The intriguing dystopian narratives provide graphic illustrations of what might befall us if we allow ourselves to be 'cloud serfs' controlled by the big tech corporations. The lessons are there for all to see: we need to be masters of the new technology rather than its mindless servants. *Black Mirror* shows us the horrors of what might happen if we fail to heed such lessons.

Statements and Declarations

Funding

No funding was required for the research and writing of this article.

Conflicts of Interest

There are no conflicts of interest.

References

1. ^*Thirer A (2022). "How Science Fiction Dystopianism Shapes the Debate Over AI & Robotics." Discourse. https://www.discoursemagazine.com/p/how-science-fiction-dystopianism-shapes-the-debate-over-ai-robotics.*
2. ^*Science (2025). "China Tops the World in Artificial Intelligence Publications, Database Analysis Reveals." Science. https://www.science.org/content/article/china-tops-world-artificial-intelligence-publications-database-analysis.*
3. a, b, c, d, e, f, g, h*Bender EM, Hanna A (2025). The AI Con: How to Fight Big Tech's Hype and Create the Future We Want. London: Vintage.*
4. ^*Brooker C, Jones A (2018). Inside Black Mirror. London: Ebury Press.*
5. ^*Jackson D (2024). "Black Mirror' Creator Charlie Brooker Tries to Steer Clear of Painting Technology as a Bad Thing: "It's How You Use It"." Deadline. https://deadline.com/2024/08/black-mirror-charlie-brooker-interview-1236034654/.*
6. ^*Saunders E (2025). 'Black Mirror Could Just Run and Run', Says Charlie Brooker. London: BBC. https://www.bbc.com/news/articles/cj45vn5k78go.*
7. a, b*Johnson DK (ed) (2020). Black Mirror and Philosophy: Dark Reflections. London: Wiley Bla.*
8. ^*Hyland (2019). Adventures in the Matrix. Mauritius: Scholars' Press.*
9. a, b, c, d, e, f, g*Schnell A (2024). The Philosophy of Black Mirror. Bucharest: Zeta Books.*
10. ^*Plato (1982). The Republic. Harmondsworth: Penguin. Trans. H.D.P. Lee.*
11. a, b, c*Descartes (1968). Discourse on Method. Harmondsworth: Penguin. Trans. F.E. Sutcliffe.*
12. a, b, c*Bostrom N (2003). "Are You Living in a Computer Simulation?" Philos Q. 53(211):243–255.*

13. ^Smoot G (2014). "You Are a Simulation & Physics Can Prove It: TedX Talk – You Tube." YouTube. https://youtu.be/Chfoo9NBEow?si=-BStIXUhyBtNzf6X.
14. a, bJenkins PS (2006). "Historical Simulations – Motivational, Ethical and Legal Issues." J Futures Stud. 11(1): 23–42.
15. a, b, cHyland T (2024). Morality, Neo-Idealism and Conscious Realism. London: Dodo Books Ltd.
16. ^Open Culture (2017). The Philosophy of The Matrix: From Plato and Descartes, to Eastern Philosophy. Open Culture. https://www.openculture.com/2017/03/the-philosophy-of-the-matrix.html.
17. ^Raj RI, Rani S (2024). "Ramifications of Digital Technology - A Critical Study of Black Mirror: Bandersnatch." Literary Voice. 2(2):53–61.
18. a, bCerella A (2025). "How Generative AI Could Change How We Think and Speak." The Conversation. https://theconversation.com/how-generative-ai-could-change-how-we-think-and-speak-267118.
19. ^Riehm KE, Feder KA, Tormohlen KN, et al. (2019). "Associations Between Time Spent Using Social Media and Internalizing and Externalizing Problems Among US Youth." JAMA Psychiatry. 76(12):1266–1273.
20. ^Duarte GA (2021). "Black Mirror: Mapping the Possible in a Post-Media Condition." In Duarte GA, Battin JM (Eds.), Reading Black Mirror: Insights Into Technology and the Post-Media Condition. 1st ed. pp. 25–50. http://www.jstor.org/stable/j.ctv371bnwf.4.
21. a, bVaroufakis Y (2025). Techno Feudalism: What Killed Capitalism. London: Vintage.
22. ^XRToday (2025). "Meta Invests $100 Billion Into Augmented Reality Smart Glasses, Virtual Reality Headsets, Mixed Reality Quest." MetaversePortfolio. https://www.xrtoday.com/augmented-reality/meta-invests-100-billion-into-augmented-reality-smart-glasses-virtual-reality-headsets-mixed-reality-quest-metaverse-portfolio/.
23. ^Dick PK (2017). Electric Dreams. London: Awesome Books.
24. ^Hyland T (2023). "AI Applications and Non-Human Persons." Philosophy of Education Society of Great Britain. https://www.philosophy-of-education.org/ai-applications-and-non-human-persons/.
25. a, bHyland T (2023). "Artificial Intelligence (AI) Developments and Their Implications for Humankind: A Critical Analysis of Contemporary Trends and Perspectives." Curr J Appl Sci Technol. 42(21):12–26.
26. ^Oneadvanced (2025). The Benefits of Artificial Intelligence (AI): A Strategic Overview. Oneadvanced. https://www.oneadvanced.com/resources/the-benefits-of-artificial-intelligence-ai-a-strategic-overview/.
27. ^World Economic Forum (2025). 7 Ways AI Is Transforming Healthcare. World Economic Forum. https://www.weforum.org/stories/2025/08/ai-transforming-global-health/.

28. ^Bittencourt II, Chalco G, Santos J, Fernandes S, Silva J, Batista N, Hutz C, Isotani S (2024). "Positive Artificial Intelligence in Education (P-AIED): A Roadmap." Int J Artif Intell Educ. 34:732–792. doi:10.1007/s40593-023-00357-y.

29. a, bStern N, Romani M, Pierfederici R, et al. (2025). "Green and Intelligent: The Role of AI in the Climate Transition." Climate Action. 4(56):3–13.

30. ^Daws R (2025). "Sam Altman: AI Will Cause Job Losses and National Security Threats." AI News. https://www.artificialintelligence-news.com/news/sam-altman-ai-cause-job-losses-national-security-threats/.

31. ^IBM (2025). What Are Large Language Models (LLMs)? https://www.ibm.com/think/topics/large-language-models.

32. ^Chomsky N (2023). "The False Promise of Chat GPT." New York Times.

33. ^Goff P (2023). "ChatGPT Can't Think — Consciousness Is Something Entirely Different to Today's AI." The Conversation. https://theconversation.com/chatgpt-cant-think-consciousness-is-something-entirely-different-to-todays-ai-204823.

34. ^This Is World (2025). Gödel's Theorem Debunks the Most Important AI Myth. AI Will Not Be Conscious | Roger Penrose. YouTube. https://youtu.be/biUfMZ2dts8?si=oQF8UCpGH5Xl39pD.

35. ^Rajan A (2025). "The Weaponisation of Science: How to Avoid a Global Catastrophe (Carlo Rovelli)." Radical Podcast. https://www.bbc.com/audio/play/m002jtw6.

36. ^Heaven D (2023). "Geoffrey Hinton Tells Us Why He's Now Scared of the Tech He Helped Build." MIT Technology Review. https://www.technologyreview.com/2023/05/02/1072528/.

37. ^Bailey J (2025). "Climate Champion or Catastrophe: Can AI 'Offset' Its Own, Growing Emissions?" EuronewsGreen. https://www.euronews.com/green/2025/04/11/ai-climate-champion-catastrophe-carbon-emissions.

38. ^Wei Xing (2025). "Why OpenAI's Solution to AI Hallucinations Would Kill ChatGPT Tomorrow." The Conversation. https://theconversation.com/why-openais-solution-to-ai-hallucinations-would-kill-chatgpt-tomorrow-265107.

39. ^Adib-Moghaddam A (2025). "Here's Why the Public Needs to Challenge the 'Good AI' Myth Pushed by Tech Companies." The Conversation. https://theconversation.com/heres-why-the-public-needs-to-challenge-the-good-ai-myth-pushed-by-tech-companies-259200.

40. ^Khalaf AM (2023). "The Impact of Social Media on the Mental Health of Adolescents and Young Adults: A Systematic Review." Cureus. 15(8):e42990. doi:10.7759/cureus.42990.

41. ^Minamatini K (2024). "Social Media Addiction and Mental Health: The Growing Concern for Youth Well-Being." Stanford Law School. https://law.stanford.edu/2024/05/20/social-media-addiction-and-mental-health-the-growing-concern-for-youth-well-being/.
42. ^Osler L (2025). "AI-Induced Psychosis: The Danger of Humans and Machines Hallucinating Together." The Conversation. https://theconversation.com/ai-induced-psychosis-the-danger-of-humans-and-machines-hallucinating-together-269850.
43. ^UK Parliament (2025). Smartphones in Schools: Practice, Policy and International Perspectives. London: House of Lords Library. https://lordslibrary.parliament.uk/smartphones-in-schools-practice-policy-and-international-perspectives/.
44. ^Berners-Lee T (2025). "Why I Gave the World Wide Web Away for Free." The Guardian. https://www.theguardian.com/technology/2025/sep/28/why-i-gave-the-world-wide-web-away-for-free.
45. ^Hassett J (2024). The Truth About AI. https://www.youtube.com/watch?v=eFQdWAeZxlE.
46. ^Thorne S (2025). "ChatGPT Is About to Get Erotic, But Can OpenAI Really Keep It Adults-Only?" The Conversation. https://theconversation.com/chatgpt-is-about-to-get-erotic-but-can-openai-really-keep-it-adults-only-267660.
47. ^Bhardwaj A (2025). "If the AI Bubble Does Burst, Taxpayers Could End Up With the Bill." The Conversation. https://theconversation.com/if-the-ai-bubble-does-burst-taxpayers-could-end-up-with-the-bill-269115.
48. ^BBC (2025). "Don't Blindly Trust What AI Tells You, Says Google's Sundar Pichai." BBC News. https://www.bbc.com/news/articles/c8drzv37z4jo.
49. ^European Commission (2023). "Legal Framework Governing Medicinal Products for Human Use in the EU." European Commission. https://health.ec.europa.eu/medicinal-products/legal-framework-governing-medicinal-products-human-use-eu_en.
50. ^Singh S, Dahlmann F (2025). "Regulating AI Use Could Stop Its Runaway Energy Expansion." The Conversation. https://theconversation.com/regulating-ai-use-could-stop-its-runaway-energy-expansion-258425.
51. ^Kokkoris I (2025). "Despite Google's Recent Victory, a Flurry of Competition Cases Could Still Change How the Tech Giants Do Business." The Conversation. https://theconversation.com/despite-googles-recent-victory-a-flurry-of-competition-cases-could-still-change-how-the-tech-giants-do-business-264890.
52. ^Ritchie H (2024). "Australia Approves Social Media Ban on Under-16s." BBC News. https://www.bbc.com/news/articles/c89vjj0lxx9o.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.