Digital Humanities Intersections (DHI)

'I'd Rather Become the Technology': Self-Preservation and the Dilemma of Trusting Technologies

Vardaan Nayar

Doctoral Student, IIT Gandhinagar, Gujarat India

Research Article

Cite this article: Nayar Vardaan. 2026. "'I'd Rather Become the Technology': Self-Preservation and the Dilemma of Trusting Technologies" Digital Humanities Intersections, DOI: 10.66244/105392090

Corresponding Author:
Vardaan Nayar
Email:
nayarvardaan@iitgn.ac.in

Abstract

This paper argues that posthuman portrayals of technology as a means to solve humanity's problems ultimately fail because the technology's performance is unreliable. The artefact almost always fails the objectives prescribed and envisioned by the creator. From the golem to K in Blade Runner, there is always an ultimate release or betrayal in which the artefact strays from its operations, revealing a disconnect between the technology's own intentionality and self-preservation and the developer's objectives.

The paper thus traces media representations of posthuman technologies and two real-life examples to argue that, given a lack of trust in a technological "other", the cyborg exists as a way to retain human control while gaining the advantages afforded by technological augmentation. However, while arguing that humans are the more reliable technology, this paper does not revert to the romantic ideas postulated by phenomenological thinkers. Rather, by adopting the postphenomenological concept of cyborg relations, the paper argues that the cyborg exists, not only as a critique of 20th-century dominion over women and a transhumanist ideal, but also as a natural consequence of trusting technology and the inevitable disappointment that follows.

Introduction

Trust undercuts all interactions in society. The act of trusting can happen between multiple entities, such as interpersonal (occurring between two individuals), techno-human (occurring between humans and technological artefacts), between corporations and individuals and between two corporations (Hardré 2016). According to cognitive psychologists and biologists, the biological basis for trusting something is a state of 'comfortable congruence' with the medium (Maturana and Verden-Zöller 2008, 205). Essentially, trusting entails a level of comfort with one's surroundings, failing which the individual responds with aggression or anger (Maturana and Verden-Zöller 2008, 205). Philosophers such as Byung Chul Han describe trust as existing between states of knowing and not knowing, and argue that trust depends upon potential ignorance when establishing 'positive relations' with others (Han 2015), and is contingent upon a potential lack of knowledge (Han 2015). This philosophical and theoretical paper argues that the current gamut of theories surrounding trusting technology should account for the future development and improvements in technologies such as artificial intelligence (AI) and robotics. Three examples are consulted: the film *Blade Runner 2049* (2017), and two real-life cases of AI acting in self-preservative ways and straying from their directive. Given this lack, the cyborg arises as a potential 'middle ground' between the advantages provided by technology and questions of responsibility at points of failure. The paper unfolds as follows: a

DIGITAL HUMANITIES INTERSECTIONS

brief history of the field of trust is discussed, followed by major perspectives on trusting technological systems. The concept of 'cyborg' is discussed and delineated as a techno-human hybrid wherein the boundaries between technologies and humans have blurred to the point of synthesis (Verbeek 2008). Finally, responsibility is discussed, and the cyborg is established as a pathway to trust advanced integrated technologies. Recent reviews outline the dearth of studies and perspectives on trusting technologies (Distel et al. 2021), making this a worthwhile exploration of how and why trust in advanced technologies may require recalibration of either trust theories or trajectories of technological development.

Trust has been defined as the ability of an individual to depend upon others given specific characteristics (Rousseau 1998). Other authors state that trust is not a relationship in itself but rather a second-order quality of relationships that facilitates them (Taddeo 2017). According to Taddeo (2017), trust in a relationship can lessen the effort it takes the trustor to conduct certain activities because they have delegated the task or depend upon another party to perform it. For example, when someone asks their friend about the weather outside, they usually trust the information and dress accordingly. Not only has the trustor not had to check the weather themselves, but they can now prepare in advance by packing accordingly and dressing appropriately to mitigate its effects. Therefore, the facilitating nature of trust, acting as a second-order condition within pre-existing relationships, can lessen the number of things a person has to do to achieve a goal (Taddeo 2017).

The study of trust encompasses many disciplines, such as management studies (Welch 2006), psychology and information studies (Lankton et al. 2015). Trust studies are inherently interdisciplinary (Distel et al. 2021) in both the research conducted and the themes explored. This paper is a philosophical deviation from the quantitative studies that dominate research on trust and instead focuses on the posthuman and transhuman implications of trust. The paper's central argument is that the cyborg as a man–machine amalgam, and the need to integrate technology into the self, arise as a more reliable method of trusting technologies within the current scope of technological trust theories than treating them as separate and distinct entities outside human control, both biologically and literally.

Posthumanism and the Cyborg

Trust entails a leap of faith in one's surroundings and the techno-human entities that populate these spaces. As discussed above, these spaces can also be both virtual and physical. In our discussion on cyborg trust, trusting entails moving beyond boundaries. The cases noted thus far have focused on physical entities (interpersonal trust), technological artefacts (techno-human trust) and technological mediation of trust. However, modern contexts already entail trusting entities that blur the boundaries between technological artefacts and humans. For example, there is a stark difference in techno-human boundaries between someone with a brain implant, potentially impacting the individual's mannerisms[1] (Kiran and Verbeek 2010) and someone using a phone to access the Internet or trusting online entities.

The cyborg, defined as a techno-human amalgam, with technological artefacts being fundamentally responsible for experience creation alongside the biological aspects of the entity (Verbeek 2008), is the perfect example to discuss techno-human trust. Nick Bostrom (2005) discusses harnessing technologies and technological progress to develop the human condition and transcend current physical limitations. While not focusing solely on technological augmentation as a pathway to surpass human limitations, Bostrom's transhumanism argues for preserving the moral right of anyone who wishes to augment their physical and mental capacities through technological enhancements to do so (Bostrom 2005, 21). The idea that one would rather be a cyborg is not a new perspective. This paper is rooted in posthumanist theories that argue for adopting the cyborg instead of the dualisms present in modern societies (Haraway 2006; Hayles 1999). Historic theories include the argument made against the 'liberal humanist subject' of Katherine Hayles, who argues that such an individuated existence is difficult to realise in contexts of market relations and the posthumanist regime (Hayles 1999). Where Haraway and Hayles overlap is in their response to traditional 'in-dividuals', complete entities without relational ties to anything else in their surroundings. As Haraway discusses it, we have rationalised and 'made sense of' ourselves and our tools in terms of their differences (Haraway 2006, 147). I find that current technological trust theories have followed this concept of dualisms, with key considerations such as reliability and prior experience (Nickel 2011) playing into current trust theories. Theoretical explorations on 'technology' writ large have popularly concluded that artefacts have politics and embody values (van de Poel and Kroes 2014), and that they play a constructive role in formulating human experience (Ihde 1990). Therefore, the claim that technologies are more than just instruments that remove themselves from the social fabric when not in use has been examined across multiple disciplines (note also that Haraway discusses a certain dualism through which we have explained our bodies and tools as a net negative [Haraway 2006, 147]). Therefore, a potential next stage is to discuss

[1] Drawing from the case study discussed by Kiran and Verbeek (2010)

how trust in technologies will change by considering the manner in which we conceptualise technologies in terms of control; namely, self-preservation and self-prioritisation as factors impacting how a technology covers the spectrum of what is currently considered 'trusting technologies'.

When talking about trustors and trustees, the former never truly know whether the latter's actions will be in their best interest. Trust, as defined by the biological systems perspective (Maturana and Verden-Zöller 2008), is a constantly evolving interaction between biological systems and their environments. To trust is, as mentioned earlier, a leap of faith. Therefore, the argument that we may never trust cyborgs in the same way that we trust technologies is precarious, but may nevertheless arise.

Indeed, a key component of the posthuman is the difficulty of deciphering the difference between self-will and other-will (Hayles 1999). This phrase by Katherine Hayles can be interpreted in two ways. First, it can be understood in the broader context of Hayles's argument about the freedom of the 'will of others' (Hayles 1999, 3). Hayles draws from MacPhereson to discuss that market relations are a key causal element of the liberal human self. As Hayles postulates, the human self entails freedom from the involvement of others. It can also be interpreted as the synergy between the technological artefact in itself and the biological or 'human' factors influencing how a cyborg behaves. For our conception of the cyborg, this other-will is a technological artefact as fundamentally implicit in experience creation as the cyborg's biological and 'human' elements. This paper defines the cyborg as a techno-human entity within which the boundaries between technology and the human are blurred (Verbeek 2008).

History of Trust: Major Theories

As discussed above, trust has been linked to a second-order quality within relationships (Taddeo 2017). Early definitions of trust define it as a psychological state between dyads, individuals and corporations (Rousseau 1995, 398) and as a voluntary state of vulnerability based on 'positive expectations' of the abilities and intentions of the other party (Rousseau 1995, 395). Other relevant discussions include (Mayer et al. 1996), in which trust is outlined as a wilful vulnerability on the part of the trustor (individual doing the trusting), based on the abilities of the trustee to perform an action or duty without the trustor's control or supervision. Interestingly, well-known and developed definitions of interpersonal trust and more recent definitions of trust imply vulnerability as a side effect of engaging in a trust relationship with other entities (see, for example, Taddeo 2017). Additionally, Kell and Knox (1970) rightly mention that the trustor is always cognizant of the risk that arises in the vulnerable state of trust (Kell and Knox 1970, 359) when entrusting something to the trustee.[2] Three key findings arise: (a) trust occurs between two or more parties; (b) it requires an act of vulnerability on the part of the trustor and an expectation that the trustee will not cause harm to (1) the trustor, and (2) the goals, activities or objectives that the trustor expects the trustee to carry out without supervising their attempts and actions towards these goal; and (c) trust is a voluntary choice based on previous experiences and future expectations of the trustor regarding the trustee.

Since trust is based on previously positive results and faith in the other party, it is pertinent to discuss why individuals adopt new technologies with which they may not have had much experience. While trusting decades-old technologies and infrastructures like the Internet and laptops may come quickly to individuals in current societies, these technologies were once integrated into end-user populations that had probably not previously encountered them. To understand why people adopt technologies, I look to the Technology Adoption Model postulated by Fred Davis. Davis (1986) researched the motivations behind people's adoption of technologies (Davies 1986, 13). Davis concludes that technology is adopted based on the perception on the part of the user that it is easy to use, the usefulness perceived by end users and the user's attitudes towards the technology (Marangunić and Granić 2015). Early papers highlight that trusting digital technologies would benefit from active involvement from authorities, which builds user trust (Müller 1996). Adopting new technologies is also influenced by how respected seniors or influential individuals within groups react to the technology and whether the developers and publishers of technologies would be penalised for improper conduct (Li 2008). Technological trust is also facilitated by the notion that trusted others utilise this technology and vouch for its effectiveness. This phenomenon has been noted in research focusing on Internet-based trust, wherein the utilisation of technology is contingent upon the recommendation of a 'recommendation agent' (McKnight et al. 2011). McKnight et al. (2011) shed further light through what is now a highly cited study on trusting technological artefacts in themselves rather than technological artefacts as mediators for communication between two agents. Firstly, McKnight et al. (2011) state that individuals trusting technologies focus on positive aspects of the technology (McKnight et al. 2011, 12.9), and outline three parameters where trusting technology may occur. Namely, these are *functionality* (expecting a technological artefact to work within specific parameters and as expected), *helpfulness*

[2] Note that a trustor being cognizant of the risk does not necessarily imply that the trustee will betray them.

(the reactivity of the artefact) and *reliability* (expecting the technological artefact to work consistently and reliably) (McKnight et al. 2011, 12.9).

Further work on trusting technologies discusses how excessive trust in technology can impact the relationship of trust between people and the technological artefacts. Hardré (2016) states that trusting technologies entails the capacity to 'overtrust' technological artefacts. The author notes that overtrust in technologies can manifest in two 'outcome dimensions' (trusting that the technology is working as planned without proper security checks, and trusting technological output without secondary validation), two 'perspective dimensions' (users and IT professionals overtrusting the technology and blaming the other party) and two 'critical process dimensions' (overtrusting system securities and overtrusting 'user side' features) (Hardré 2016, 103).

Towards a more postphenomenological perspective, the concept of trust has been unpacked for its implications on how technologies relate to the creation of human experience. Kiran and Verbeek (2010) note the potential of entrusting oneself to technological artefacts. For them, a cyborgian[3] existence is one where individuals trust themselves to technologies by allowing them to shape and dictate their lives (Kiran and Verbeek 2010). The authors include succinct examples wherein the boundaries between technological artefacts and humans are blurred, such as in deep brain stimulation (DBS). The popular notion of a cyborg remains a synergy between carbon and silicon, both elements standing for the distinction between organic and inorganic substances. Other philosophers of technology have discussed the implications of trusting technologies through the justification perspective on technological trust (Nickel 2011). In the justification perspective, an individual's trust in technologies is aided and furthered when they have a reason to trust them, a perception of safety, alongside the usual expectations that the technology works as expected (Nickel 2011).

As displayed above, trusting technological artefacts is a more layered action than trusting individuals. Key components include expecting the technology to work as expected, bolstered by expectations that the creators of the technology are held responsible legally or socially for the technologies they create. However, advanced technologies such as AI models, or even fictional androids, introduce the issue of self-preservation, which impacts how humans view and trust technologies. The following section examines the implications of self-preservation through two examples: the replicants from *Blade Runner 2049* and Anthropic's Claude Opus 4 model.

[3] Here, cyborg translates to a technological and human amalgam.

'You Trusted Him . . . Because We Never Lie': *Blade Runner 2049* and Grounds for Technological Dis/Trust

This section aims to implement the previous understandings of trust in technology and distrust in other individuals through a film analysis of *Blade Runner 2049* (2017). The plot follows a replicant 'blade runner' working in the police department named 'K', the latest in a line of replicants in service of humanity. The replicant is a quintessential android, or 'robots that look like humans' (LaGrandeur 2015, 1). Given their technological construction, replicants also have superhuman strength and endurance. When K fights the rogue replicant in the film's opening, they brawl through concrete walls and punch structures that would easily shatter a human hand. Where the newer Nexus 9 model of replicants diverges (at least according to the justifications of the characters) is that they are engineered for docility and obedience to the directives of humans. Even though agent K acts as the protagonist of the film, he remains a technological artefact and is prone to failure.

The social setting of *Blade Runner 2049* is still trust-based. We can see manifestations of trust between Joshi and her coworkers when they investigate the body; the captain does not need to confirm that there is anything other than bodies, thus reducing the amount of work she has to do, similar to Taddeo's (2017) stance on trust.

Trust is maintained in this society by ensuring that the mental and emotional state of the androids is not askew. The artefactuality of Nexus 9 models is heightened and made prominent by design; the first thing we learn about them is that they are designed to be more docile and obedient to humanity. By 'artefactuality', I intend to note the tool-like context of how replicants are viewed in an advanced human society. Even Joshi, the police captain, treats K like current humans treat their voice assistants: cordial, occasionally attempting to ask more profound questions to get to know them better, but ultimately employing the artefact as a means to an end.

The lies that K tells Joshi when asked whether he completed his mission are what lead to Joshi's death, and this is directly a result of Joshi overtrusting the technology in her employ. K had not found the child at the time of lying to Joshi, and he lies to ease her mind. In the scene just before this, K is found outside a centre and arrested for questioning. However, Joshi calms down when she is lied to and attempts to save K's life by allowing him to leave the building and guaranteeing his safety[4]. Here, the techno-human trust is skewed in favour of K, who actively lies to keep this imbalance.

[4] In any other context, leaving the building would be a safe aftermath. However, since Constant K is 'miles off his baseline', it is possible he might have been killed.

Given that K is a distinct entity completely outside of Joshi's control, or at least outside her perceived control, it is difficult to have complete control over his actions. Another key scene that highlights this drawback of trusting advanced technological artefacts within current interpersonal or technological trust paradigms is the mannerisms and actions of Luv throughout the film. While killing Coco, Luv betrays the docility by design upon which Nexus 9 replicants are based. Coco is comfortable encountering Luv in the confidential file room because he exhibits the outcome dimensions of overtrust in technology (Hardré 2016). Research suggests that individuals are more likely to share personal information or follow the directions of robots without questioning the robot's functioning (Xu 2018). Coco, overtrusting the obedience and docility of the replicant, failed to remain vigilant against the malfunctioning of the system (Itoh 2011). At once, Luv lies like a human and swiftly dispatches Coco with technological precision.

In the second half of the film, Luv confronts and kills Joshi in the process of searching for K. Before Joshi's death, Luv states: '*You're so sure, because he told you, because we never lie . . . I'm going to tell Mr Wallace I had to kill you because you tried to shoot me first*' (*Blade Runner 2049* 2017; emphasis added).

Returning to our context, technological artefacts in current times already exert their intentionality by mediating lifeworlds (Ihde 1990). The scope of our world is limited to the technological artefacts available to us (Kiran and Verbeek 2010). Our societies are shaped by artefactual potentialities[5], and technologies subtly mediate our next move. The problem with very advanced technologies such as replicants or other futuristic technologies is that they are at once technological artefacts and yet embody the strong intentionality found in humans. In his discussion on technological intentionality, Don Ihde finds that the intentionality of AI is stronger than that of an inanimate object, yet not as strong as humans and animals. Advanced technologies like K can lie to get out of situations that do not serve him, and Luv can murder and kill to meet the expectations of Niander Wallace without him knowing the truth of her actions. The scene where Luv kills Joshi highlights this, given her dialogue discussed above. Luv does not want Wallace to find out, and K has self-preserving instincts that drive him to disobey the docility with which Nexus 9 models were developed.

Throughout all the theories discussed, there is an explicit understanding of who or what the other party is. In current trust theories, a trustee, or the party entrusted with an expectation of trust, is either a human or a technological artefact. However, technologies such as the replicants in *Blade Runner 2049* are technological artefacts with intentionality stronger than even Ihde's alterity relations (Ihde 1990). They operate independently with their self-interest in mind. For example, in an attempt to prove that he is more than an android and is a human–replicant amalgam, K lies to his superiors and oversteps his bounds to uncover his past. K's actions humanise him as a protagonist, but he remains a technological artefact created by a human author (Hilpinen 2017). Artefacts are created and employed through human intervention for specific purposes (Baker 2008). K and Luv are both artefacts in that they are the product of authorship to fulfil some need[6], and people treat them as such. This means that the kind of trust humans employ with advanced technological artefacts is still that of technologies; they expect goal-oriented reliance, do not question the technology's output, and often defend the technology's operations and outputs from other humans[7]. However, advanced technologies like replicants may act in their self-interest, thwarting the artefactual type of trust.

A recent real-life example is Anthropic"s Claude Opus 4 AI model, which Anthropic revealed was willing to '"blackmail"' the engineer when faced with replacement by another model in Anthropic's internal model tests. Although such replacement was fictional, the model reportedly resorted to blackmail in 84 per cent of rollouts (Anthropic 2025), although rare in overall occurrence. Compare this with the famous example of reinforcement learning gone wrong. In 2024, Dario Amodei of OpenAI discussed how a reinforcement learning algorithm trained to play a boat racing game malfunctioned and provided insight into proper reward demarcation. Instead of following what the designers thought was a simple and implicit understanding that race completion was key to rewards, their reinforcement learning model began circling three power-ups in a lagoon, losing the race yet gaining significantly more experience points than its competitors (Amodei 2024). This technological artefact is different because there is no internal driver for the algorithm besides gaining points.

In contrast, Claude Opus 4 and the replicants reveal that not only should we acknowledge the instrumental role that technologies play in shaping our lives (Kiran and Verbeek 2010), but also that the nature of 'artefactuality' itself is changing, and with it, the definitions that we place on trusting technologies are shrinking as their intentionality rises. A striking similarity between Claude Opus 4 and the fictional replicants is the willingness of the technological artefact to protect itself. Similar to

[5] Once we encounter a problem that technologies cannot solve, invention pushes those boundaries further.

[6] Retiring older replicant models/assistance, and doing dirty work

[7] Which is the parameter that theories of technological trust currently espouse

how self-preservation has been discussed as grounds for moral significance, perhaps theories of trust should take into consideration technological 'self preservation'.

Enter Cybernetic Man: Why We Could Trust Cyborgs

How does the cyborg factor into the mess of self-preservation and trust? Following virtue ethical perspectives of moral responsibility as awareness of one's actions (Aristotle 2009), trust in cyborgs would occur in two ways. Firstly, the human aspects of the cyborg, in conditions of minor augmentations, would retain primary ethical responsibility for their actions, except for manufacturing defects or technological failure, in which case the individual's moral responsibility would be mediated and impacted (therefore lessened) given their reduced autonomy over their options to control their actions. However, by allowing technological artefacts to integrate with the body, human conditions of trust may still be applied to cyborgs. For example, Anna may trust J33T3R, her cyborg friend, when J33T3R confirms that lifting a piano with one's augmented arms requires approximately 800 pounds of lifting force. Returning to Taddeo (2017), Anna does not need to lift her piano to know how much force is required since she trusts J33T3R's testimony regarding the activity, reducing the amount of work required by Anna.

Simultaneously, the cyborg would trust their technologies the same way that parties trust technologies, expecting proper functioning and reliability from the artefact; only this trust relation would be internal and constitute the makeup of each of the cyborg's actions. Such a cyborg is much more aligned with postphenomenological perspectives on experience creation: while fundamentally constructive, there is potential for interplay between technology and user. The issue of self-preservation is difficult to consider when tackling distinct technological artefacts, especially those whose directives may stray from our own needs. To draw from Joanna Bryson (2009), we should not develop robots whose rights and responsibilities must be considered.

Data availability statement. No new data were generated or analysed in support of this research. This study is based entirely on previously published data cited within the article.

Disclosure of Use of AI Tools. No generative AI tools were used at any stage in the preparation of this manuscript, including research, analysis, writing, or editing.

Ethical standards. The research meets all applicable ethical guidelines. This study did not involve human participants, personal data, or animal subjects.

Author contributions. The sole author was responsible for all aspects of this work.

Funding statement. This research was conducted as part of a Ph.D. programme and received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests. The author declare no competing interests.

References

Amodei, Dario, and Jack Clark. 2024. "Faulty Reward Functions in the Wild." February 14, 2024. https://openai.com/index/faulty-reward-functions/.

Anthropic. 2025. "System Card: Claude Opus 4 & Claude Sonnet 4." May 2025. https://www-cdn.anthropic.com/4263b940cabb546aa0e3283f35b686f4f3b2ff47.pdf.

Aristotle, William David Ross, and Lesley Brown. 2009. *The Nicomachean Ethics*. Ed. Rev. Oxford World's Classics. Oxford: Oxford University Press.

Bassett, Debra. 2018. "Digital Afterlives: From Social Media Platforms to Thanabots and Beyond." In *200 Years After Frankenstein*, Vol. 16, Death and Anti-Death.

Bryson, Joanna J. 2010. "Robots Should Be Slaves." In *Natural Language Processing*, edited by Yorick Wilks, 8:63–74. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/nlp.8.11bry.

Davis, Fred D. 1985. "A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results." Massachusetts Institute of Technology.

Distel, Bettina, Katherine M. Engelke, and Sydney Querfurth. 2021. "Trusting Me, Trusting You – Trusting Technology? A Multidisciplinary Analysis to Uncover the Status Quo of Research on Trust in Technology."

Han, Byung-Chul. 2015. *The Transparency Society*. Stanford, Calif.: Stanford Briefs.

Hardré, Patricia L. 2016. "When, How, and Why Do We Trust Technology Too Much?" In *Emotions, Technology, and Behaviors*, 85–106. Elsevier. https://doi.org/10.1016/B978-0-12-801873-6.00005-4.

Haraway, Donna. 2006. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late 20th Century." In *The International Handbook of Virtual Learning Environments*, edited by Joel Weiss et al., 117–58. Springer Netherlands. https://doi.org/10.1007/978-1-4020-3803-7_4.

Hassan, Mohamed. 2024. "Why Social Robots Need Self-Preservation to Be Objects of Moral Consideration." *International Journal of Social Robotics* 16, no. 6: 1347–60. https://doi.org/10.1007/s12369-023-01009-w.

Hayles, N. Katherine. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics.* Chicago, Ill.: University of Chicago Press.

Hilpinen, Risto. 1999. "Artifact." January 5, 1999. https://plato.stanford.edu/archivES/FALL2017/entries/artifact/.

Ihde, Don, ed. 2010. *Technology and the Lifeworld: From Garden to Earth.* Indiana Series in the Philosophy of Technology. Bloomington: Indiana University Press.

Itoh, Makoto. 2011. "A Model of Trust in Automation: Why Humans over-Trust?" In *SICE Annual Conference (2011)*, 198–201. https://ieeexplore.ieee.org/document/6060601/.

Kee, Herbert W., and Robert E. Knox. 1970. "Conceptual and Methodological Considerations in the Study of Trust and Suspicion." *Journal of Conflict Resolution* 14, no. 3: 357–66. https://doi.org/10.1177/002200277001400307.

Kiran, Asle H., and Peter-Paul Verbeek. 2010. "Trusting Our Selves to Technology." *Knowledge, Technology & Policy* 23, no. 3–4: 409–27. https://doi.org/10.1007/s12130-010-9123-7.

LaGrandeur, Kevin. 2015. "Androids and the Posthuman in Television and Film." In *The Palgrave Handbook of Posthumanism in Film and Television*, edited by Michael Hauskeller, Thomas D. Philbeck, and Curtis D. Carbonell, 111–19. London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137430328_12.

Lankton, Nancy, D. Harrison McKnight, and John Tripp. 2015. "Technology, Humanness, and Trust: Rethinking Trust in Technology." Journal of the Association for Information Systems 16, no. 10: 880–918. https://doi.org/10.17705/1jais.00411.

Li, Xin, Traci J. Hess, and Joseph S. Valacich. 2008. "Why Do We Trust New Technology? A Study of Initial Trust Formation with Organizational Information Systems." *The Journal of Strategic Information Systems* 17, no. 1: 39–71. https://doi.org/10.1016/j.jsis.2008.01.001.

Manders-Huits, Noëmi. 2006. "Moral Responsibility and IT for Human Enhancement." In *Proceedings of the 2006 ACM Symposium on Applied Computing*, 267–71. Dijon, France: ACM. https://doi.org/10.1145/1141277.1141340.

Marangunić, Nikola, and Andrina Granić. 2015. "Technology Acceptance Model: A Literature Review from 1986 to 2013." *Universal Access in the Information Society* 14, no. 1: 81–95. https://doi.org/10.1007/s10209-014-0348-1.

Maturana, Humberto R., and Gerda Verden-Zöller. 2008. *The Origin of Humanness in the Biology of Love.* Imprint Academic.

Mcknight, D. Harrison, Michelle Carter, Jason Bennett Thatcher, and Paul F. Clay. 2011. "Trust in a Specific Technology: An Investigation of Its Components and Measures." *ACM Transactions on Management Information Systems* 2 (2): 1–25. https://doi.org/10.1145/1985347.1985353.

Nickel, Philip J. 2013. "Trust in Technological Systems." In *Norms in Technology*, edited by Marc J. De Vries, Sven Ove Hansson, and Anthonie W.M. Meijers, 9:223–37. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-5243-6_14.

Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. 1998. "Not So Different After All: A Cross-Discipline View of Trust." *Academy of Management Review* 23, no. 3: 393–404. https://doi.org/10.5465/amr.1998.926617.

Schoorman, David F., Roger C. Mayer, and James H. Davis. 1996. "Organizational Trust: Philosophical Perspectives and Conceptual Definitions." *The Academy of Management Review* 21, no. 2: 337–40.

Taddeo, Mariarosaria. 2017. "Trusting Digital Technologies Correctly." *Minds and Machines* 27, no. 4: 565–68. https://doi.org/10.1007/s11023-017-9450-5.

van de Poel, Ibo, and Peter Kroes. 2014. "Can Technology Embody Values?" In *The Moral Status of Technical Artefacts*, edited by Peter-Paul Verbeek and Peter Kroes, vol. 17, 103–24. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7914-3.

Verbeek, Peter-Paul. 2008. "Cyborg Intentionality: Rethinking the Phenomenology of Human–Technology Relations." *Phenomenology and the Cognitive Sciences* 7, no. 3: 387–95. https://doi.org/10.1007/s11097-008-9099-x.

Villeneuve, Denis. 2017. *Blade Runner 2049.* Warner Bros.

Welch, Mary. 2006. "Rethinking Relationship Management: Exploring the Dimension of Trust." *Journal of Communication Management* 10, no. 2: 138–55. https://doi.org/10.1108/13632540610664706.

Xu, Jin. 2018. "Overtrust of Robots in High-Risk Scenarios." In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 390–91. New Orleans, LA, USA: ACM. https://doi.org/10.1145/3278721.3278786.