

Title: Re-Centralization of AI focusing on Social Justice Author: Adnan Hadzi, Denis Roio

Colophon

A Nourishing Network is a peer-to-peer publishing experiment starting from the feed as a potentially multi-directional circulation device.

A Nourishing Network is initiated by servus.at (Davide Bevilacqua) in collaboration with varia.zone (Alice Strete & Manetta Berends) and is published in the context of AMRO 2020 (Arts Meets Radical Openness).

Editing: Davide Bevilacqua

Design and development: Manetta

Berends, Alice Strete

Paper: xxxx

Typeface: Gnu Unifont, White Rabbit, Ansi Shadow

Print and production: Varia

This project is produced with Free Software tools. The feeds are made with Pelican & Weasyprint.

Davide is an artist and curator working in the blurry area between media and contemporary art.

Manetta Berends is a designer working with forms of networked publishing, situated software and collective infrastructures.

Alice Strete is an artist and researcher interested in the intricate relationship between humans and the technologies they surround themselves with.

Many thanks to our partners, collaborators, authors and the AMRO community.

Published under the CC-BY-SA 4.0 license.

A Nourishing Network - December 2020
Re-Centralization of AI focusing on Social Justice
by Adnan Hadzi, Dennis Roio

-----a-----

-----n-n-n-n-n-n-n-n-----
-----o-o-o-o-o-o-o-----
-----u-u-u-u-u-u-u-----
-----r-r-r-r-r-r-r-----
-----i-i-i-i-i-i-i-----
-----s-s-s-s-s-s-s-----
-----h-h-h-h-h-h-h-----
-----i-i-i-i-i-i-i-----
-----n-n-n-n-n-n-n-----
-----g-g-g-g-g-g-g-----

-----n-n-n-n-n-n-n-----
-----e-e-e-e-e-e-e-----
-----t-t-t-t-t-t-t-----
-----w-w-w-w-w-w-w-----
-----o-o-o-o-o-o-o-----
-----r-r-r-r-r-r-r-----
-----k-k-k-k-k-----



In order to lay the foundations for a discussion around the argument that the adoption of artificial intelligence (AI) technologies benefits the powerful few,¹ focusing on their own existential concerns,² we decided to narrow down our analysis of the argument to social justice (i.e. restorative justice). This paper signifies an edited version of Adnan Hadzi's text on Social Justice and Artificial Intelligence,³ exploring the notion of humanised artificial intelligence⁴ in order to discuss potential challenges society might face in the future. The paper does not discuss current forms and applications of artificial intelligence, as, so far, there is no AI technology, which is self-conscious and self-aware, being able to deal with emotional and social intelligence.⁵ It is a discussion around AI as a speculative hypothetical entity. One could then ask, if such a speculative self-conscious hardware/software system were created, at what point could one talk of personhood? And what criteria could there be in order to say an AI system was capable of committing AI crimes?

Concerning what constitutes AI crimes the paper uses the criteria given in Thomas King et al.'s paper Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions,⁶ where King et al. coin the term "AI crime". We discuss the construction of the legal system through the lens of political involvement of what one may want to consider to be 'powerful elites'.⁷ In doing so we will be demonstrating that it is difficult to prove that the adoption of AI technologies is undertaken in a way, which mainly serves a powerful class in society. Nevertheless, analysing the culture around AI technologies with regard to the nature of law with a philosophical and sociological focus enables us to demonstrate a utilitarian and authoritarian trend in the adoption of AI technologies. Mason argues that "virtue ethics is the only ethics fit for the task of imposing collective human control on thinking machines"⁸ and AI. We will apply virtue ethics to our discourse around artificial intelligence and ethics.

As expert in AI safety Steve Omonhundo believes that AI is "likely to behave in antisocial and harmful ways unless they are very carefully designed."⁹ It is through virtue ethics that this paper will propose for such a design to be centred around restorative justice in order to take control over AI and thinking machines, following

Mason's radical defense of the human and his critique of current thoughts within trans- and post- humanism as a submission to machine logic.

The paper will conclude by proposing an alternative practically unattainable, approach to the current legal system by looking into restorative justice for AI crimes,¹⁰ and how the ethics of care could be applied to AI technologies. In conclusion the paper will discuss affect¹¹ and humanised artificial intelligence with regards to the emotion of shame, when dealing with AI crimes. In this paper we will aim at re-centralizing AI ethics through social justice, with focus on restorative justice, allowing for an advanced jurisprudence, where human and machine can work in symbiosis on reaching virtue ethics, rather than being in conflict with each other.

In order to discuss AI in relation to personhood this paper follows the descriptive psychology method¹² of the paradigm case formulation¹³ developed by Peter Ossorio.¹⁴ Similar to how some animal rights activists call for certain animals to be recognised as non-human persons,¹⁵ this paper speculates on the notion of AI as a non-human person being able to reflect on ethical concerns.¹⁶ Here Wynn Schwartz argues that "it is reasonable to include non-humans as persons and to have legitimate grounds for disagreeing where the line is properly drawn. In good faith, competent judges using this formulation can clearly point to where and why they agree or disagree on what is to be included in the category of persons."¹⁷ In the case of AI technologies we ask whether the current vision for the adoption of AI technologies, a vision which is mainly supporting the military-industrial complex through vast investments in army AI,¹⁸ is a vision that benefits mainly powerful elites.

In order to discuss these questions, one has to analyse the history of AI technologies leading to the kind of 'humanised' AI system this paper posits. The old-fashioned approach,¹⁹ some may still say contemporary approach, was to primarily research into 'mind-only'²⁰ AI technologies/systems. Through high level reasoning, researchers were optimistic that AI technology would quickly become a reality.

Those early AI technologies were a disembodied approach using high level logical and abstract symbols. By the end of the 80s researchers found that the disembodied approach was not even achieving low level tasks humans could easily perform.²¹ During that period many researchers stopped working on AI technologies and systems, and the period is often referred to as the "AI winter".²² Rodney Brooks then came forward with the proposition of "Nouvelle AI",²³ arguing that the old-fashioned approach did not take into consideration motor skills and neural networks. Only by the end of the

90s did researchers develop statistical AI systems without the need for any high-level logical reasoning;²⁴ instead AI systems were ‘guessing’ through algorithms and machine learning. This signalled a first step towards humanistic artificial intelligence, as this resembles how humans make intuitive decisions;²⁵ here researchers suggest that embodiment improves cognition.²⁶ With embodiment theory Brooks argued that AI systems would operate best when computing only the data that was absolutely necessary.²⁷ Further in *Developing Embodied Multisensory Dialogue Agents* Michal Paradowski argues that without considering embodiment, e.g. the physics of the brain, it is not possible to create AI technologies/systems capable of comprehension.

Foucault’s theories are especially helpful in discussing how the “rule of truth” has disciplined civilisation, allowing for an adoption of AI technologies which seem to benefit mainly the upper-class. But then should we think of a notion of ‘deep-truth’ as the unwieldy product of deep learning AI algorithms? Discussions around truth, Foucault states, form legislation into something that “decides, transmits and itself extends upon the effects of power”²⁸. Foucault’s theories help to explain how legislation, as an institution, is rolled out throughout society with very little resistance, or “proletarian counter-justice”²⁹. Foucault explains that this has made the justice system and legislation a for-profit system. With this understanding of legislation, and social justice, one does need to reflect further on Foucault’s notion of how disciplinary power seeks to express its distributed nature in the modern state. Namely one has to analyse the distributed nature of those AI technologies, especially through networks and protocols, so that the link can now be made to AI technologies becoming ‘legally’ more profitable, in the hands of the upper-class.

In *Protocol*, Alexander Galloway describes how these protocols changed the notion of power and how “control exists after decentralization”³⁰. Galloway argues that protocol has a close connection to both Deleuze’s concept of control and Foucault’s concept of biopolitics³¹ by claiming that the key to perceiving protocol as power is to acknowledge that “protocol is an affective, aesthetic force that has control over life itself.”³² Galloway suggests that it is important to discuss more than the technologies, and to look into the structures of control within technological systems, which also include underlying codes and protocols, in order to distinguish between methods that can support collective production, e.g. sharing of AI technologies within society, and those that put the AI technologies in the hands of

the powerful few.³³ Galloway's argument in the chapter Hacking is that the existence of protocols "not only installs control into a terrain that on its surface appears actively to resist it"³⁴, but goes on to create the highly controlled network environment. For Galloway hacking is "an index of protocological transformations taking place in the broader world of techno-culture."³⁵

Having said this, the prospect could be raised that restorative justice might offer "a solution that could deliver more meaningful justice"³⁶. With respect to AI technologies, and the potential inherent in them for AI crimes, instead of following a retributive legislative approach, an ethical discourse,³⁷ with a deeper consideration for the sufferers of AI crimes should be adopted.³⁸ We ask: could restorative justice offer an alternative way of dealing with the occurrence of AI crimes?³⁹

Dale Millar and Neil Vidmar described two psychological perceptions of justice.⁴⁰ One is behavioural control, following the legal code as strictly as possible, punishing any wrongdoer,⁴¹ and second the restorative justice system, which focuses on restoration where harm was done. Thus an alternative approach for the ethical implementation of AI technologies, with respect to legislation, might be to follow restorative justice principles. Restorative justice would allow for AI technologies to learn how to care about ethics.⁴² Julia Fionda describes restorative justice as a conciliation between victim and offender, during which the offence is deliberated upon.⁴³ Both parties try to come to an agreement on how to achieve restoration for the damage done, to the situation before the crime (here an AI crime) happened. Restorative justice advocates compassion for the victim and offender, and a consciousness on the part of the offenders as to the repercussion of their crimes. The victims of AI crimes would not only be placed in front of a court, but also be offered engagement in the process of seeking justice and restoration.⁴⁴

Restorative justice might support victims of AI crimes better than the punitive legal system, as it allows for the sufferers of AI crimes to be heard in a personalised way, which could be adopted to the needs of the victims (and offenders). As victims and offenders represent themselves in restorative conferencing sessions, these become much more affordable,⁴⁵ meaning that the barrier to seeking justice due to the financial costs would be partly eliminated, allowing for poor parties to be able to contribute to the process of justice. This would benefit wider society and AI technologies would not only be defined by a powerful elite. Restorative justice could

hold the potential not only to discuss the AI crimes themselves, but also to get to the root of the problem and discuss the cause of an AI crime. For John Braithwaite restorative justice makes re-offending harder.⁴⁶

In such a scenario, a future AI system capable of committing AI crimes would need to have knowledge of ethics around the particular discourse of restorative justice. The implementation of AI technologies will lead to a discourse around who is responsible for actions taken by AI technologies. Even when considering clearly defined ethical guidelines, these might be difficult to implement,⁴⁷ due to the pressure of competition AI systems find themselves in. That said, this speculation is restricted to humanised artificial intelligence systems. The main hindrance for AI technologies to be part of a restorative justice system might be that of the very human emotion of shame. Without a clear understanding of shame it will be impossible to resolve AI crimes in a restorative manner.⁴⁸

Furthering this perspective, we suggest that reflections brought by new materialism should also be taken into account: not only as a critical perspective on the engendering and anthropomorphic representation of AI, but also to broaden the spectrum of what we consider to be justice in relation to all the living world. Without this new perspective the sort of idealized AI image of a non-living intelligence that deals with enormous amounts of information risks to serve the abstraction of anthropocentric views into a computationalist acceleration, with deafening results. Rather than such an implosive perspective, the application of law and jurisprudence may take advantage of AI's computational and sensorial enhanced capabilities by including all information gathered from the environment, also that produced by plants, animals and soil. Thus one might want to think about a humanised symbiosis between humans and technology,⁴⁹ along the lines of Garry Kasparov's advanced chess,⁵⁰ as in advanced jurisprudence.⁵¹ A legal system where human and machine work together on restoring justice, for social justice.

-
1. Cp. G. Chaslot, "YouTube's A.I. was divisive in the US presidential election", Medium, November 27, 2016. Available at: <https://medium.com/the-graph/youtubes-ai-is-neutral-towards-clicks-but-is-biased-towards-people-and-ideas-3a2f643dea9a#.tjuusil7d> [accessed February 25, 2018]; E. Morozov, "The Geopolitics Of Artificial Intelligence", FutureFest, London, 2018. Available at: <https://www.youtube.com/watch?v=7g0hx9LPBq8> [accessed October 25, 2019].
 2. Cp. M. Busby, "Use of 'Killer Robots' in Wars Would Breach Law, Say Campaigners", The Guardian, August 21, 2018. Available at : <https://web.archive.org/web/20181203074423/https://www.theguardian.com/science/2018/aug/21/use-of-killer-robots-in-wars-would-breach-law-say-campaigners> [accessed October 25, 2019].
 3. Cp. A. Hadzi, "Social Justice and Artificial Intelligence", Body, Space & Technology, 18 (1), 2019, pp. 145-174. Available at: <https://doi.org/10.16995/bst.318> [accessed October 25, 2019].
 4. Cp. A. Kaplan and M. Haenlein, "Siri, Siri, in my Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence", Business Horizons, 62 (1), 2019, pp. 15-25. <https://doi.org/10.1016/j.bushor.2018.08.004>; S. Legg and M. Hutter, A Collection of Definitions of Intelligence, Lugano, Switzerland, IDSIA, 2007. Available at: <http://arxiv.org/abs/0706.3639> [accessed October 25, 2019].²

- N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford, Oxford University Press, 2014.
- 5.
 6. Cp. T. King, N. Aggarwal, M. Taddeo and L. Floridi, "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions", SSRN Scholarly Paper No. ID 3183238, Rochester, NY, Social Science Research Network, 2018. Available at: <https://papers.ssrn.com/abstract=3183238> [accessed October 25, 2019].
 7. P. Mason, *Clear Bright Future*, London, Allen Lane Publishers, 2019.
 8. Mason, *Clear Bright Future*.
 9. S. Omohundro, "Autonomous Technology and the Greater Human Good", *Journal of Experimental & Theoretical Artificial Intelligence*, 26 (3), 2014, pp. 303-315, here: p. 303.3
 10. Cp. C. Cadwalladr, "Elizabeth Denham: 'Data Crimes are Real Crimes'", *The Guardian*, July 15, 2018. Available at: <https://web.archive.org/web/20181121235057/https://www.theguardian.com/uk-news/2018/jul/15/elizabeth-denham-data-protection-information-commissioner-facebook-cambridge-analytica> [accessed October 25, 2019].
 11. Cp. B. Olivier, "Cyberspace, Simulation, Artificial Intelligence, Affectionate Machines and Being Human", *Communication*, 38 (3), 2012, pp. 261-278. <https://doi.org/10.1080/02500167.2012.716763> [accessed October 25, 2019]; E.A. Wilson, *Affect and Artificial Intelligence*, Washington, University of Washington Press, 2011.
 12. Cp. P.G. Ossorio, *The Behavior of Persons*, Ann Arbor, Descriptive Psychology Press, 2013. Available at: <http://www.sdp.org/sdppubs-publications/the-behavior-of-persons/> [accessed October 25, 2019].
 13. Cp. J. Jeffrey, "Knowledge Engineering: Theory and Practice", *Society for Descriptive Psychology*, 5, 1990, pp. 105-122.
 14. Cp. P.G. Ossorio, *Persons: The Collected Works of Peter G. Ossorio, Volume I*. Ann Arbor, Descriptive Psychology Press, 1995. Available at: <http://www.sdp.org/sdppubs-publications/persons-the-collected-works-of-peter-g-ossorio-volume-1/> [accessed October 25, 2019].
 15. Cp. M. Mountain, "Lawsuit Filed Today on Behalf of Chimpanzee Seeking Legal Personhood", *Nonhuman Rights Blog*, December 2, 2013. Available at: <https://www.nonhumanrights.org/blog/lawsuit-filed-today-on-behalf-of-chimpanzee-seeking-legal-personhood/> [accessed January 8, 2019]; M. Midgley, "Fellow Champions Dolphins as 'Non-Human Persons'", *Oxford Centre for Animal Ethics*, January 10, 2010. Available at: <https://www.oxfordanimalethics.com/2010/01/fellow-champions-dolphins-as-%E2%80%9Cnon-human-persons%E2%80%9D/> [accessed January 8, 2019].
 16. Cp. R. Bergner, "The Tolstoy Dilemma: A Paradigm Case Formulation and Some Therapeutic Interventions", in K.E. Davis, F. Lubguin and W. Schwartz (eds.), *Advances in Descriptive Psychology*, Vol. 9, 2010, pp. 143-160. Available at: <http://www.sdp.org/sdppubs-publications/advances-in-descriptive-psychology-vol-9/>; P. Laungani, "Mindless Psychiatry and Dubious Ethics", *Counselling Psychology Quarterly*, 15 (1), 2002, pp. 23-33. Available at: <https://doi.org/10.1080/09515070110102305> [accessed October 26, 2019].
 17. W. Schwartz, "What Is a Person and How Can We Be Sure? A Paradigm Case Formulation", SSRN Scholarly Paper No. ID 2511486, Rochester, NY: Social Science Research Network, 2014. Available at: <https://papers.ssrn.com/abstract=2511486> [accessed October 25, 2019].
 18. Cp. Mason, *Clear Bright Future*.
 19. Cp. M. Hoffman, and R. Pfeifer, "The Implications of Embodiment for Behavior and Cognition: Animal and Robotic Case Studies", in W. Tschacher and C. Bergomi (eds.), *The Implications of Embodiment: Cognition and Communication*, Exeter, Andrews UK Limited, 2015, pp. 31- 58. Available at: <https://arxiv.org/abs/1202.0440>
 20. N.J. Nilsson, *The Quest for Artificial Intelligence*, Cambridge, Cambridge University Press, 2009.
 21. Cp. R. Brooks, *Cambrian Intelligence: The Early History of the New AI*, Cambridge, MA, A Bradford Book, 1999.
 22. Cp. D. Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence*, New York, Basic Books, 1993; H.P. Newquist, *The Brain Makers*, Indianapolis, Ind: Sams., 1994.
 23. Cp. R. Brooks, "A Robust Layered Control System for a Mobile Robot", *IEEE Journal on Robotics and Automation*, 2 (1), 1986, pp. 14-23. Available at: <https://doi.org/10.1109/JRA.1986.1087032> [accessed October 25, 2019].
 24. Cp. Brooks, *Cambrian Intelligence*.
 25. Cp. R. Pfeifer, "Embodied Artificial Intelligence", presented at the International Interdisciplinary Seminar on New Robotics, Evolution and Embodied Cognition, Lisbon, November, 2002. Available at: <https://www.informatics.indiana.edu/rocha/publications/embrob/pfeifer.html> [accessed October 25, 2019].
 26. Cp. T. Renzenbrink, "Embodiment of Artificial Intelligence Improves Cognition", *Elektormagazine*, February 9, 2012. Available at: <https://www.elektormagazine.com/articles/embodiment-of-artificial-intelligence-improves-cognition> [accessed January 10, 2019]; G. Zarkadakis, "Artificial Intelligence & Embodiment: Does Alexa Have a Body?", *Medium*, May 6, 2018. Available at: <https://medium.com/@georgezarkadakis/artificial-intelligence-embodiment-does-alexa-have-a-body-d5b97521a201> [accessed January 10, 2019].
 27. Cp. L. Steels and R. Brooks, *The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated Agents*, London/New York, Taylor & Francis, 1995.
 28. M. Foucault, "Disciplinary Power and Subjection", in S. Lukes (ed.), *Power*, New York, NYU Press, 1986, pp. 229-242, here: p. 230.
 29. M. Foucault, *Power*, edited by C. Gordon, London, Penguin, 1980, p. 34.6
 30. A.R. Galloway, *Protocol: How Control Exists After Decentralization*, Cambridge, MA, MIT Press, 2004, p. 81.
 31. Cp. M. Foucault, *The Birth of Biopolitics: Lectures at the Collège de France, 1978-1979*, London, Pan Macmillan, 2008.
 32. Galloway, *Protocol*, p. 81.

- Cp. Galloway, Protocol, p. 147.
- 33.
 34. Galloway, Protocol, p. 146.
 35. Galloway, Protocol, p. 157.
 36. Crook, Comparative Media Law and Ethics, p. 310.7
 37. Cp. R. Courtland, "Bias Detectives: The Researchers Striving to Make Algorithms Fair", *Nature*, 558, 2018, pp. 357-360. Available at: <https://doi.org/10.1038/d41586-018-05469-3> [accessed October 25, 2019].
 38. Cp. H. Fry, "We Hold People With Power to Account. Why Not Algorithms?" *The Guardian*, September 17, 2018. Available at: <https://web.archive.org/web/20190102194739/https://www.theguardian.com/commentisfree/2018/sep/17/power-algorithms-technology-regulate> [accessed October 25, 2019].
 39. Cp. O. Etzioni, "How to Regulate Artificial Intelligence", *The New York Times*, January 20, 2018. Available at: <https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html> [accessed October 25, 2019]; A. Goel, "Ethics and Artificial Intelligence", *The New York Times*, December 22, 2017. Available at: <https://www.nytimes.com/2017/09/14/opinion/artificial-intelligence.html> [accessed October 25, 2019].
 40. Cp. N. Vidmar and D.T. Miller, "Socialpsychological Processes Underlying Attitudes Toward Legal Punishment", *Law and Society Review*, 1980, pp. 565-602.
 41. Cp. M. Wenzel and T.G. Okimoto, "How Acts of Forgiveness Restore a Sense of Justice: Addressing Status/Power and Value Concerns Raised by Transgressions", *European Journal of Social Psychology*, 40 (3), 2010, pp. 401-417.
 42. Cp. N. Bostrom and E. Yudkowsky, "The Ethics of Artificial Intelligence", in K. Frankish and W.M. Ramsey (ed.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge, Cambridge University Press, 2014, pp. 316-334; Frankish and Ramsey, *The Cambridge Handbook of Artificial Intelligence*.
 43. Cp. J. Fionda, *Devils and Angels: Youth Policy and Crime*, London, Hart, 2005.8
 44. Cp. Nils Christie, "Conflicts as Property", *The British Journal of Criminology*, 17 (1), 1977, pp. 1-15.
 45. Cp. J. Braithwaite, "Restorative Justice and a Better Future", in E. McLaughlin and G. Hughes (eds.), *Restorative Justice: Critical Issues*, London, SAGE, 2003, pp. 54-67.
 46. Cp. J. Braithwaite, *Crime, Shame and Reintegration*, Cambridge, Cambridge University Press, 1989.
 47. Cp. A. Conn, "Podcast: Law and Ethics of Artificial Intelligence", *Future of Life*, March 31, 2017. Available at: <https://futureoflife.org/2017/03/31/podcast-law-ethics-artificial-intelligence/> [accessed September, 22 2018].
 48. Cp. A. Rawnsley, "Madeleine Albright: 'The Things that are Happening are Genuinely, Seriously Bad'", *The Guardian*, July 8, 2018. Available at: <https://web.archive.org/web/20190106193657/https://www.theguardian.com/books/2018/jul/08/madeleine-albright-fascism-is-not-an-ideology-its-a-method-interview-fascism-a-warning> [accessed October 25, 2019].
 49. Cp. D. Haraway, "A Cyborg Manifesto", *Socialist Review*, 15 (2), 1985. Available at: <http://www.stanford.edu/dept/HPS/Haraway/CyborgManifesto.html> [accessed October 25, 2019]; C. Thompson, "The Cyborg Advantage", *Wired*, March 22, 2010. Available at: <https://www.wired.com/2010/03/st-thompson-cyborgs/> [accessed October 25, 2019].
 50. Cp. J. Hipp et al., "Computer Aided Diagnostic Tools Aim to Empower Rather than Replace Pathologists: Lessons Learned from Computational Chess", *Journal of Pathology Informatics*, 2, 2011. Available at: <https://doi.org/10.4103/2153-3539.82050> [accessed October 25, 2019].
 51. Cp. J. Baggini, "Memo to Those Seeking to Live for Ever: Eternal Life Would be Deathly Dull", *The Guardian*, July 8, 2018. Available at: <https://web.archive.org/web/20181225111455/https://www.theguardian.com/commentisfree/2018/jul/08/live-for-ever-eternal-life-deathly-dull-immortality> [accessed October 25, 2019].