CONTEXTUALIZING DISCRIMINATION IN AI: MORAL IMAGINATION AND VALUE SENSITIVE DESIGN AS A FRAMEWORK TO STUDY AI DEVELOPMENT IN THE EU

Chiza Chirwa

Student number: 516057

Law and Information Society

University of Turku

Faculty of Law

November 2021

UNIVERSITY OF TURKU

Faculty of law

Chiza Chirwa: Contextualizing Discrimination in AI: Moral imagination and value sensitive design as a framework to study AI development in the EU

AI will continue to play a role in service provision by both public and private sector providers. These services sometimes border on fundamental rights such as the right not to be discriminated against. Commonly, most people hold the prevailing belief that data knows best and that algorithms ensure equality and fairness.

However, algorithms do discriminate and sometimes they perpetuate inequality. The paper is built on the premise that the primary source of discrimination in AI is human input and not the underlying AI technology. Moral imagination, or more accurately, the lack of it, may be responsible for non-technical bias in AI decision-making. Prohibition of discrimination is recognised as a fundamental value of the EU and it follows that AI systems must comply with EU regulations in their decision-making to prevent discrimination and in the process protect human dignity.

As concerns human dignity, algorithmic bias continues to be the main problem regarding automated decision-making. This bias, more often than not, is as a result of reinforcing some institutional and societal discrimination into AI systems in the development phase. This has the effect of continuing to perpetuate bias in the wider society when AI systems are used.

This paper takes a dogmatic approach in analyzing the EU value of prohibition of discrimination as it is interpreted in the design process of AI systems by using moral imagination and value sensitive design as a framework of investigation.

Key Words: Artificial Intelligence (AI), Values, Human Rights, Algorithmic Bias, Discrimination Prohibition, Moral Imagination, Value Sensitive Design.

CONTENTS

REFERENCES

BOOKS

Bengio Y., Goodfellow I. and Courville A. ‘Deep Learning’ (2017). MIT Press

Beranger J. ‘The Algorithmic Code of Ethics: Ethics at the bedside of the Digital Revolution’ (2018). Volume 2. John Wiley and Sons. Hoboken

Bevan D.J, Wolfe W. R. and Werhane P. H., ‘Systems Thinking and Moral Imagination: Rethinking Business Ethics’ (2019). Springer International Publishing AG.

Bodding P. ‘Towards a Code of Ethics for Artificial Intelligence’ (2017).Edt O’Sullivan B and Wooldridge M. Springer International.

Campbell T.D. ‘Adam Smith’s theory of moral imagination’ (1971). George Allen and Unwin Ltd.

Carpanelli E. and Lazzerini N. ‘Use and Misuse of New Technologies: Contemporary Challenges in International and European Law’ (2019). Springer Publishing. New York

Chinen M. ‘Law and Autonomous machines: The co-evolution of legal responsibility and technology’ (2019). Edward Elgar Publishing. Northampton.

Coeckelbergh M. ‘AI Ethics’ (2020). MIT Press. Massachusetts

Coleman F. ‘A Human Algorithm: How AI is Redefining Who We Are’ (2019). Counterpoint. Berkeley

Corrales M., Fenwick M. And Forgo N. ‘New Technology, Big Data and the Law’ (2017). Springer Publishing. Singapore.

Dubber M.D., Paquale F., and Das S. ‘Oxford Handbook of Ethics in AI’ (2020). Oxford University Press. Oxford.

Dearing A. ‘Justice for Victims of crime: Human Dignity as the Foundation of Criminal Justice in Europe’ (2017). Springer International Publishing. Cham

Eamon B. ‘The Condensed Wealth of Nations and The Incredibly Condensed Theory of Moral Sentiments’ (2011). Adam Smith Research Trust.

Edward L, Schafer B. and Harbinja E. 'Future Law: Emerging Technology Regulation and Ethics' (2020). Edinburgh University Press. Edinburgh.

Flanagan M., Howe D. and Nissenbaum H. 'Embodying Values in Technology: Theory and Practice: In Information Technology and Moral Philosophy' (2008). Cambridge University Press.

Friedman B., Kahn P.E and Borning A. in Galleta D and Zhang P. Edt. Human-Computer Interaction and Management Information System: Application (2006). M.E Sharpe. New York.

Friedman B and Hendry D.G. 'Value Sensitive Design: Shaping Technology with Moral Imagination' (2019). Massachusetts Institute of Technology. Massachusetts

Friedman B. and Kahn P.E. 'Human Agency and Responsible Computing: Implications for Computer System Design' (1992). Elsevier Science Publication Co. New York.

Gertrude H. 'The Moral Imagination: From Adam Smith to Lionel Trilling' (2012). Rowman & Littlefield Publishers. Plymouth.

Harari N.Y., 'Homo Deus: A Brief History of Tomorrow' (2016). Signal Publishing.

Haven J.C, 'Heartificial Intelligence: Embracing our Humanity to maximize Machines' (2016). Penguin Books. New York

Haaksonseen K. 'Natural Law Theory. Encyclopedia of Ethics' (1992). Edit Becker L.C and Beker C.B Garland. New York.

Helbing D. 'Thinking Ahead: Essays on big data, digital revolution and participatory market society' (2015). Springer Publishing. Berlin

Johnson M. 'Morality for Humans: Ethical Understanding From the Perspective of Cognitive Science' (2014). University of Chicago Press.

Moran S., Cropley D. and Kaufman James C. 'The Ethics of Creativity' (2014). Palgrave Macmillan. New York.

Miller F. and Wertherimer A, 'The Ethics of Consent Theory and Practice' (2010). Oxford University Press New York.

Kirste S. 'Human Dignity as the Foundation of Freedom'. (2018). Franz Steiner Verlag

Noble S. U. ‘Algorithms of Oppression, How Search Engines Reinforce Racism’ (2018). New York University Press

O’Neal C. ‘Weapons of Math Destruction: How Big Data increases Inequality and threatens democracy’ (2016). Crown Publishers. New York.

Owen R., Stilgoe J.,Gorman M., Fisher E. and Guston D. ‘Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society II: A Framework for Responsible Innovation’. (2013) John Wiley and Sons. New Jersey

Perez C.C. ‘Invincible Women: Data Bias in a World Designed for Men’ (2019). Abrams Press. New York.

Smith A. ‘The Theory of Moral Sentiments’ (1759).18th Ed, Millar A and Kincaid A. London

Woodrow B. ‘Cambridge Handbook on the law of Algorithms’ (2019). Cambridge University Press. Cambridge

ARTICLES

Abney K. ‘Robotics, Ethical Theory and Meta Ethics’ (2014): A guide for the Perplexed. In Robot Ethics: Ethical and Social Implications of Robotics. Edt Lin P., Abney K. And Bekey. G. A. MIT Press.

Albrechtslund, A. ‘Ethics and technology design’ (2007). Ethics and Information Technology. Volume 9. No. 1

Allan R. and Masters D., ‘Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making’ (2019). Europäische Rechtsakademie. Available https://doi.org/10.1007/s12027-019-00582-w

Aizenberg A and Hoven J., ‘Designing for Human rights in AI’ (August 2019). Sage Publication. Available on https://journals.sagepub.com/doi/full/10.1177/2053951720949566

Atkins K.‘Autonomy and the Subject Character of Experience’ (2000). Journal of Applied philosophy Volume 17. No 1.

B

Barn B. S and Barn R.,'Human and value sensitive Aspects of Mobile App Design' (November 2018). A Foucauldian Perspective. Springer International Publishing AG

Baracas S and Selbst D.A'Big Data's Disparate Impact' (June 2016) California Law Review Volume 104, No. 3

Brent M. 'Auditing for Transparency in Content Personalization Systems' (2016). International Journal of Communication. Volume 10.

Brent. M 'Principles Alone Cannot Guarantee Ethical AI' (November 2019). Nature Machine Intelligence

Bird S., Barocas S., Crawford K., Diaz F. and Wallach H.'Exploring Or Exploitation. Social and Ethical Implications of Autonomous Experiments in AI'. Microsoft Research available at https://ssrn.com/abstract=2846909

Bjork A., Paavalo J., Strik T., Tanhua I. And Vainio A. 'Finland in the International Human Rights System' (2019) .Publication of the Government's analysis, Assement and Research activities 50. Prime Minister's Office.

Buhmann, A., Pabmann, J., & Fieseler, C., 'Managing Algorithmic Accountability: Balancing Reputational Concerns, Engagement Strategies, and the Potential of Rational Discourse' (2019). Journal of Business Ethics.

Burton E., Goldsmith J., Koenig S., Kuipers B., Mattei N. And Wash T. 'Ethical Considerations in Artificial Intelligence Courses' (June 2019). AI Magazine. Association for the Advancement of Artificial Intelligence.

Butterworth M. 'The ICO and Artificial Intelligence' (2018): The role of fairness in the GDPR framework, Computer Law and Security Review 34.

Borgesius F.Z., 'Discrimination, Artificial Intelligence and Algorithmic Decision Making' (2018). Council of Europe. Directorate General of Democracy. Strasbourg

Boldyreva E.L. Grishina N.Y., and Duisembina Y. 'Cambridge Analytica: Ethics and Online Manipulation with Decision-making process' (April 2018). Article available and accessed from

https://www.europeanproceedings.com/article/10.15405/epsbs.2018.12.02.10

Bryson, J. ‘Robots Should Be Slaves. In Close Engagements with Artificial Companions’ (2010). : Key Social, Psychological, Ethical and Design Issues. Edited by Yorick Wilks, 63–74. Amsterdam:

C

Cencil A. and Cowthorne D. ‘Refining value sensitive design: A (Capability-based) Procedural Ethics Approach to Technological Design for Wellbeing’ (May 2020). Science and Engineering Ethics Journal. Springer Publishing.

Chris R. ‘Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning’ (March 2018). Queen Mary School of Law Legal Studies Research Paper No. 243/2016) 29 <https://ssrn.com/abstract=2853462> 276 Bygrave (n 165) 22.

Coecklebergh M.‘Regulation or Responsibility? Autonomy, Moral Imagination and Engineering Design.’ (May 2006). Science Technology and Human Values. Volume 31. No.3

Collier M.‘Hume’s Theory of Moral Imagination’(July 2010). History of Philosophy Quarterly, Volume 27 No. 3 University of Illinois Press.

Cowl J. King T.C, Taddeo M. and Floridi L. ‘Designing Artificial intelligence for Social Good: Seven Essential Factors’ (2019). Alan Turing institute. London available at https://philpapers.org/archive/COWDAF.pdf

Citron, D. K., & Pasquale, F. ‘The Scored Society: Due Process For Automated Predictions’ (2014). Washington Law Review. Volume 89. No 1.

Crawford K. and Ryan C. ‘There Is a Blind Spot in AI Research’ (2016). Nature 538:311–313.

D

Daly A, Hasndorff T, Hui L, Mann M, Marda V, Wagner B, Wang W and Witteborn S. ‘Artificial intelligence governance and Ethics: Global Perspectives’(2019). Chinese University of Hong Kong. Research Paper No. 2019-15.

Datta A., Tschantz M.C, and Datta A. ‘Automated experiments on ad privacy settings’ (2015). Proceedings on Privacy Enhancing Technologies. Volume 1.

Davis J.P.‘Legal Dualism, Legal Ethics and Fidelity to Law’(2016). Journal of the Professional Lawyer. University of San Francisco Law Research Paper No. 20.

Diega G.N.‘Against the Dehumanization of decision-making: Algorithmic Decisions at the crossroads of intellectual property, data protection and freedom of information’(2018). JIPITEC 3.

Dignum V. ‘Responsible AI: Designing AI for Human Values’ (September 2017). ITU Journal: ICT Discoveries, Special Issue No. 1, 25

Drew C. ‘Design for data ethics: using service design approaches to operationalize ethical principles on four projects’ (2018).Philosophical Transactions of the Royal Society. A 376

Dove E.S. and Chen J. ‘Should Consent for data processing be privileged in health research? A comparative legal analysis’ (2020). International Data Privacy Law. Volume 10 No. 2

D’ Acquisto G., Domingo-Ferrer J., Kikiras P., Torra V, De Montjoye Y. And Bourka A.‘Privacy by Design In Big Data: An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics’(December 2015). EU Agency for Network and Information Security.

E

Ed B. and Berti S. ‘Legal, Social and Ethical Perspectives on Health & Technology’ (December 2020).

Edwards, L. & Veale, M. ‘Slave to the algorithm? Why a ‘right to an explanation’ is probably not the remedy you are looking for’ (2017). Duke Law & Technology Review, 16

ESPC ‘The Age of Artificial intelligence: Towards a European Strategy for Human-Centric Machines’ (2018). European Political Strategy Centre. Issue 29.

F

Fjeld J., Achten N., Hilligoss H, Nagy A. C and Srikumar M. ‘Principle AI: Mapping Consensus in Critical and Rights-Based Approaches to Principles for AI’ (January 2020). Research Publication No. 1. Berkman Klein center for Internet and Society.

Ford R. and Price W. ‘Privacy and accountability in black box Medicine’ (2016). Michigan Technology Law Review. Volume 23. No. 1

Floridi L. ‘Soft Ethics and the Governance of the Digital’ (2018). Philosophy and Technology Volume 31, No 1.

Friedman, B. ‘Value Sensitive Design. Human Values and the Design of Computer Technology’ (1996). Edited by Batya Friedman. CSLI Publications.

Friedman B. ‘Value Sensitive Design’ (November 1996). Interactions 3

Friedman, B. ‘Social judgments and technological innovation: Adolescents’ understanding of property, privacy, and electronic information’ (1997). Computer. Human. Behavior. Volume 13. No. 3

Friedman B, Kahn P.H., and Borning A.‘Values sensitive design and information systems’(2013). University of Washington.

Friedman B, Kahn P, Jr., and Howe D.C.‘Trust Online’ (December 2000). Communications of The ACM Volume 43. No 12.

G

GazzaneoL, Padovavano A. and Umbrello S. ‘Designing Smart operator 4.0 for Human Values: A value sensitive Design Approach’ (2020). International conference on industry 4.0 and smart manufacturing, published by Elsevier B.V

Grunloh C. ‘Using Technological Frames as an Analytical Tool in Value Sensitive Design’ (June 2018). Springer Publishing.

Guido Noto La Diega. ‘Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information 9’ (2018). JIPITEC 3 paragraph 1.

Gutierrez M.‘The Good, The Bad and the Beauty of Good Enough Data’ (2019). In Daly A., Devitt S.K and Mann M. Eds. Good Data. Institute of Network Cultures. Amsterdam.

H

Hildebrandt M., ‘Law As Computation in the Era of Artificial Legal Intelligence. Speaking Law to the Power of Statistics’. (March 2018)

Hacker P., ‘AI regulation in Europe’ (November 2019). Humboldt University of Berlin

Hacker P. ‘A legal Framework for AI Training Data’ (March 2020). Law, Innovation and Technology 13.

Hoven J. 'ICT and Value Sensitive Design' (2007). IFIP international Federation for Information Processing. Volume 233.

J

Jabłonowska A., Kuziemski M.,, Nowak A.M.,, Micklitz H., Pałka P., and Sartor G. 'Consumer Law And Artificial Intelligence Challenges To The EU Consumer Law And Policy Stemming From The Business' Use Of Artificial Intelligence'(2018). EUI Working Paper Law 2018/11.

Jones J. 'Human Dignity in the EU Charter of Fundamental Rights and its interpretation before the European Court of Justice' (December 2012). Springer Science + Business Media. Dordrecht.

Joshua P. D.'Law Without Mind: AI, Ethics, And Jurisprudence' (May 2018). University of San Francisco Law Research Paper.

K

Kim, P. T. 'Auditing Algorithms for Discrimination' (2017). University of Pennsylvania Law Review Online. Volume 166. No 3.

Kirste S. 'Human Dignity as a foundation of Law' (June 2018). University of Salzburg

Krieger H.L. 'The content of our categories: A cognitive Bias Approach to Discrimination and Equal Employment Opportunity' (1995). Stanford Law Review. Volume 47

Kekes J. 'Moral Imagination, Freedom and the Humanities' (1991). American philosophical Quarterly 28.

Kyung Lee et al. 'WeBuildAI: Participatory Framework for Algorithmic Governance' (November 2019). ACM Human Computer Interactions. Volume 3. CSCW, Article 181.

L

Lanzig M. 'Strongly Recommended: Revisiting Decisional Privacy to judge Hypernurdging in Self-Tracking technologies' (May 2018). Philosophy and Law. Springer Publishing.

Leenes R. and Lucivero F. 'Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design' (2014). Law, Innovation and Technology.

Lehman J.'Big Data and its exclusions'(September 2013). 66 Stanford Law Review Online. Volume 65. No 55.

Lowry S. and Macpherson G., 'A Blot on the Profession' (March 1988) British Medical Journal Volume 296. No 6623.

M

Marelli L., Lievevrouw E. and Van Hoyweghen I. 'Fit for purpose? The GDPR and the governance of European digital health Policy Studies' (February 2020). Available https://doi.org/10.1080/01442872.2020.1724929

McCrudden C. 'Human Dignity and Judicial Interpretation of Human Rights' (2008). The European Journal of International Law, Volume 19. No, 4

Mcnamara D., Graham T., Broad E. and Oong C.H. 'Trade-Offs in Algorithmic Risk Assessment: An Australian Domestic Violence Case Study' in Devitt S.K and Mann M. Edits (2019) Good Data. Institute of Network Cultures. Amsterdam.

Messina D.'Online Platforms, Profiling and Artificial Intelligence: New Challenges for the GDPR and in Particular, for the Informed and Unambigous Data Subjects' Consent' (2019). Saggi Medai Laws.

Meuwese, A. 'Regulating algorithmic decision-making one case at the time. Case note on: District Court of The Hague' 5/02/20, ECLI:NL:RBDHA:2020:865 (NJCM vs the Netherlands (SyRI)). (2020) European Review of Digital Administration & Law,Volume 1.No 1.

Michalczak R. 'Animal's Rights Against the Machines'. In Kurki V.A and Pietrykowski T. Legal Personhood: Animals, Artificial Intelligence and the Unborn (2017). Law and Philosophy Library. Volume 119. Springer Publishing.

Mitrou L., 'Data Protection, AI and cognitive services: Is the GDPR AI Proof?' (April 2019) University of the Aegean.

Morley J., Floridi L., Kinsey L., and Eihalal A. 'From What to How: An overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices' (December 2019). Science and Engineering Ethics.

Morley J. Machador C.C, Burr C, Cowls J., Joshi I. Taddeo M. and Floridi L.‘The Debate on the Ethics of AI in health care: A reconstruction and Critical Review’ (2019).

Monasso, T. ‘I don’t know what I’m doing’ (2006). Mekelessay TU Delft.

Molner P.‘Technology on the margins: AI and global migration management from a human rights perspective’(2019). Cambridge International Law Journal, Volume 8. No. 2.

N

Nida-Rumelin J. ‘Why Human dignity rests upon Freedom’. Available at https://www.researchgate.net/publication/325654586_Human_Dignity_as_a_Foundation_of_Law

Nissenbaum, H. ‘Accountability in a computerized society’ (1996). Science and Engineering Ethics Volume 2.

O

One Hundred Year Study on Artificial Intelligence (AI100),” Stanford University, accessed August 1, (2016)https://ai100.stanford.edu.

P

Pagallo U. ‘Legalize: Tackling The Normative Challenges of Artificial intelligence and Robotics through the Secondary Rules of law’ (2017). In Corrales M., Fenwick M. And Forgo N. New Technology, Big Data and the Law. Springer Publishing. Singapore.

Pauline K. ‘Data-Driven Discrimination at Work’ (2017), 48 Will & Mary Law Review. 857, 90209.

R

Renda, A. ‘Artificial Intelligence Ethics, Governance and Policy Challenges Report of a CEPS Task Force.’ (2019). Retrieved from https://www.ceps.eu/wpcontent/uploads/2019/02/AI_TFR.pdf

Reure M. Mynsberghe A. Janssen M and Poel I. ‘Digital Platforms and responsible Innovation: expanding value sensitive design to overcome ontological uncertainty’ (May 2020). Ethics and Information Technology Journal.

Rickli J.M.‘The Economic, Security and Military Implications of Artificial Intelligence for the Arab Gulf Countries’(November 2018). EDA Insight.

Robert B. And Ellen P. G.‘Algorithmic Transparency for the Smart City’ (2018), 20 Yale Journal of Law and Technology.

Rubistein I. and Good N.‘Privacy by Design; A counterfactual Analysis of Google and Facebook Privacy Incidents’(August 2013). 28 Berkely Technology Law Journal.

S

Selbst A.D., and Powles J. ‘Meaningful information and the right to explanation’ (2017). International Data Privacy Law. Volume 7. No. 4.

Selbst A.D. and Barocas S.‘The intuitive Appeal of Explainable Machines’ (2018). Fordham Law Review Volume 87. No. 3

Sellen, A., Rogers, Y., Harper, R. H. R., & Rodden, T. ‘Reflecting Human Values in the Digital Age’ (2009). Communications of the ACM, 52(3).

Scherer M. U.‘Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, And Strategies’ (2016). Harvard Journal of Law & Technology. Volume 29, No. 2 Spring.

Schmitz, A. J. ‘Secret Consumer Scores And Segmentations: Separating Haves From Have-Nots’ (2014). Michigan State Law Review, 1411.

Scipione J. ‘AI and Europe: Developments, risks, and Implications’ (May 2020). European College of Parma.

Shahriari K. and Shahriari M. ‘IEEE standard review: Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems’ (2017) IEEE Canada International Humanitarian Technology Conference (IHTC)

Stefano S. ‘The EU as A global Standard Setting Actor: The case of Data Transfer to third countries’ (2017). In Carpanelli E. and Lazzerini N. Edts (2019) Use and Misuse of New Technologies: Contemporary Challenges in International and European Law. Springer Publishing

Surden H. ‘Ethics of AI in Law: Basic Questions’ (2020). In Oxford Handbook of Ethics In AI. Edit Dubber M.D., Paquale F., and Das S. Oxford University Press. Oxford.

T

Tschider C.A. and Kennedy K. ‘Data Discrimination: The International regulatory impasse of AI Enabled Medical Wearables’ (2020).Legal, Social and Ethical Perspectives on Health & Technology

Tschider C.A ‘Deux Ex Machina: Regulating Cybersecurity and Artificial intelligence for patients of the future’ (2018). Savannah Law Review. Volume 5, No. 1

Tischbirek A.‘Artificial intelligence and discrimination: Discriminating Against Discriminatory Systems’(2020). In Wischmeyer T. and Rademacher T. Regulating artificial intelligence. Springer Nature

Timmermans J. Zhao Y. and Van de Hoven J.‘Ethics and Nanopharmacy. Value Sensitive design of New Drugs’ (2011). Nanoethics. Volume 5.

Tasioulas J. ‘First Steps Towards an Ethics of Robots and Artificial Intelligence’ (2019). Journal of practical Ethics. Volume 7. No 1.

U

Umbrello S. and De Bellis A.F. ‘A value Sensitive Design Approach to Intelligent Agents’ (2018). In Yampolskiy R. V Edt (2019) AI Safety and Security. CRC Press, Boca Raton.

Urquhart L.D.‘White Noise from White Goods? Privacy by Design for Ambient Domestic Computing’ (2020). In Future Law: Emerging Technology Regulation and Ethics. Edts Edward L, Schafer B and Harbinja E. Edinburgh University Press. Edinburgh.

V

Van den Hoven, J. ‘Innovations, Legitimacy, Ethics and Democracy’ (2007). in IFIP International Federation for Information Processing. Volume 233, The Information Society, eds. P. Goujon, Lavelle, S., Duquenoy, P., Kimppa, K., Laurent, V. Springer. Boston

Van den Hoven, J. ‘Engineering and the Problem of Moral Overload’ (March 2012). Science and Engineering Ethics. Volume 18. No. 1

Van Den Hoven J. and Weckert J. ‘Information Technology and moral philosophy’ (2008). Cambridge University Press.

Varkonyi G.G.‘Operability of the GDPR’s Consent Rule in Intelligent Systems: Evaluating the Transparency Rule and the Right to Be Forgotten’ (2019). Intelligent Environments.

Vallor S. ‘Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character’ (2015). Philosophy and Technology Volume 28, 107–124.

Villarong E., Kieserberg P. And Li T., ‘Humans Forget, Machines Remember: Artificial Intelligence and the Right to be Forgotten’ (2017) Computer Security and Law Review.

W

Wachter S. ‘Normative Challenges of identification in the internet of things: Privacy, Profiling, discrimination and the GDPR’ (2018).Computer Law & Security Review. Volume 34. No.3

Wachter S. Mittlestadt B. and Russel C. ‘Why Fairness Cannot be Automated: Bridging the gap Between EU Non-Discrimination Law and AI’ (March 2020). Computer Law & Security Review 41.

Wachter S., Mittelstadt B, and Floridi L.‘Why a Right to Explanation of Automated Decision-making Does Not Exist in the GDPR’(2017). International Data Privacy Law. Volume 7. No. 2.

Watson, H. J., & Nations, C. ‘Addressing the Growing Need for Algorithmic Transparency’ (2019). Communications of the Association for Information Systems. Volume 45 No. 26.

Wiener N. ‘Some Moral and Technical Consequences of Automation’ (May 1960). Science Magazine Volume 131. No. 3410.

Wilson N. and Kennedy K. ‘The banality of digital aggression: Algorithmic Data Surveillance in Medical Wearables’ (2020). In Digital Ethics: Rhetoric and Responsibility in Online Aggression, Hate Speech, and Harassment. Edited by Reyman J. and Sparby E. Routledge.

Y

Yeung K., Howes A., and Pogrebna G. ‘Oxford Handbook of AI ethics’ (2019). Dubber M and Pasquale F. Eds. Oxford University press.

Yeung K. ‘A study of the Implications of Advanced Digital Technologies (including AI systems) for the Concept of responsibility within a Human Rights Framework’ (2019). Council of Europe.

Z

Zalnieuriute M, Crawford L.B., Boughley J., Moses L.R. and Logan S. 'From Rule of law to Statute Drafting: Legal issues for Algorithms in government decisions making' (2019). In Woodrow B. Eds) Cambridge Handbook on the law of Algorithms. Cambridge University press.

Zhu H., Yu B., Halfaker A., and Terveen L. 'Value-Sensitive Algorithm Design: Method, Case Study, and Lessons' (November 2018). Proc. ACM Human Computer Interactions. 2, CSCW, Article 194

CONVENTIONS AND DIRECTIVES USED

Charter Of Fundamental Rights Of The EU (2012/C 326/02)

Council of Europe Convention No.108 amended by protocol CETS No. 223

European Convention for the Protection of Human Rights and Fundamental Freedoms (Protocol No. 11 and 14)

European Commission, On AI - A European approach to excellence and trust, White Paper, COM (2020) 65 final

European Commission for the Administration of Justice (CEPEJ). (2018). European Ethical Charter on the use of AI in Judicial Systems and their Environment.

European Commission (2019) Ethics Guidelines for Trustworthy AI

European Council Study on Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) And Possible Regulatory Implications. (March 2018)

Genera Data Protection Regulation Act EU 2016/679

Medical Data Devices Directive Regulation (EU) 2017/745

EUROPEAN UNION REPORTS USED

EU Agency for Fundamental Rights Report 2019

EU Agency for Fundamental Rights Report 2020

European Union Commission. High Level on Non-Discrimination, Equality and Diversity: Guidelines on Improving The Collection and Use of Equality Data.

INTERNET SOURCES

Cliff H., ‘Can A.I. Be Taught to Explain Itself?’ The New York Times Magazine (Nov. 21, 2017).

Gavazzi M. S. Cambridge Analytica: The Scandal and the Fallout so far. (June 2020). Article Available at https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

James V. Google “fixes” its racist algorithm by removing gorillas from its image-labeling tech. (January 2018). Available https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai

Joseph R. ‘Are Face Detection Cameras Racist?’ (February 2020) http://content.time.com/time/business/article/0,8599,1954643,00.html

Lum K. and Isaac W. To predict and serve (2016). Available at https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x

Vervloesem K. ‘How Dutch activists got an invasive fraud detection algorithm banned’ (2020). Available at https://algorithmwatch.org/en/syri-netherlands-algorithm/

https://www.yvtltk.fi/en/index/opinionsanddecisions/decisions.html

CASES CITED

Association Belge Des Consommateurs Test-Achats NBL and Others v Conseil Des Ministres (C-236/09)

Calderon and Rocio v Clearview AI 20.Civ 1296

Electronic Privacy Information Center v National Security Commission on Artificial Intelligence No. 1:9 CV 02906

Gill v Whitford 138 S.Ct.1916

Kraus v Cegavske No.82018

Ricci vs. DeStefano No. 07-14-1428.

Serminade Spa v Cassa Conguaglio Zucchero and Others C-106/83

Spokeo Incorporated v Robins 138 S.Ct.931

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AIS	Artificial Intelligence Systems
EU	European Union
CFR	Charter of Fundamental Rights
GDPR	General Data Protection Regulation
MDDD	Medical Data Devices Directives
TEU	Treaty on European Union
TFEU	Treaty on the Functioning of the European Union
UDHR	Universal Declaration of Human Rights
UK	United Kingdom
US	United States of America

1.0 INTRODUCTION

Countries now commonly use AI in welfare systems and are also able to work faster and much more efficiently in the provisions of social services. However, there remains an inherit risk of using data wrongly which may have a negative effect on certain classes of people.
The aim of this paper is to study artificial intelligence (to be referred to as AI from hereafter) at its developmental stage and how it complies with EU values of human rights and fundamental freedoms, particularly, the right to non-discrimination. The research tries to assess whether moral imagination in tandem with value sensitive design is a suitable framework for studying compliance to EU values during the development phase of AI.

Most people are of the opinion that AI is more objective and they also believe that inequality and structural racism is easily removed by deploying AI systems. They imagine that the AI is better suited to handle bias than the real world and that by offloading responsibilities to AI, there is an equal field that is provided and this field eliminates human bias. The problem of historical bias in decision making is an old one indeed and what is new is the probability that these biases are simply being embedded into AI and the systems further strengthen the historical bias.

There is, therefore, a need for a proactive approach to AI systems designs and this need cannot be over emphasized. Much of the debate on AI has, unfortunately, taken a reactive approach, with the focus being on the risks arising from AI systems and their application in uses such as immigration, employment, policing, and several other areas in which there may exist a risk of discrimination.

Employing the use of value sensitive design in the development of AI may install EU values in the design process and may lead to elimination of bias affecting some groups of people. Most importantly, value sensitive design may provide a review mechanism of isolating faults in AI development and remedying the same.

There has not been much consideration to the initial design phase, at least in so far as it concerns the societal and human values of the technology developers. Developmental teams of AI are often of the same age, gender and race, and have on occasions only designed AI with their values to the detriment of certain classes such as elderly people or disabled persons.

Trusting a system that considers all possibilities in AI development is also important in this study. Furthermore, transparency in AI decisions is needed to foster trust in the systems and to see how the system makes decisions. This trust is a benchmark for quantifying dignity of participants who may be subject to an AI decision. Trust in this case is measured by analyzing compliance of AI developers to EU values.

It is the considered view of the author that more research needs to be conducted that focuses on the compliance of AI developers to EU regulations and guidelines when they are designing the technology. Value sensitive design may be one of the available options that could be used to continuously evaluate not only the compliance but the enhancement of human values in AI development. According to the author's knowledge at the time of this writing, no academic paper has been written in the EU specifically on value sensitive design and its application to AI. Thus, I find that there is a strong need for this research approach to studying AI compliance to EU values.

It is very important to study the values that are developed within the AI systems at the inception stage. One main reason is that AI systems have taken a pivotal and center stage in our daily lives and are used to make decisions for us, about us or indeed decisions that concern our private and public lives. The rate at which these decisions will be made by the AI systems will only increase as technology continues to make the use of AI more affordable and readily available.

It has been argued that the more AI technology develops, the less human input will be required in some decisions. Whilst there are safeguards in place to ensure that no one decision is entirely undertaken by AI systems without the consideration of human values by a natural person, a lot more needs to be done. AI makes decisions based on provided data sets, and as this research will show, data sets used in AI training are riddled with bias against minorities or other protected classes. When the data sets are tainted with inaccuracies, AI has been known to perpetuate bias and solidifier the same in digital form.

EU values are sometimes affected by biased AI. The paper shows that EU values should be strongly emphasized at the design phase to prevent discrimination. At the onset, it must be made clear that much of the bias in AI is as a result of the design process not including comprehensive moral imagination and not as a result of the willful exclusion of certain classes by the developers.

The paper studies how discrimination may occur from lack of moral imagination of the AI developers and looks at the value sensitivity of the said technology. In brief, value sensitive design looks at how any technology incorporates values of the society into its design to reflect those values as well as to enhance the enjoyment of those values. A key importance of value sensitive design is moral imagination of the developers which is the ability of the designers to encompass all possibilities when designing a solution so as to arrive at an ethical conclusion.

The thrust of this paper explores the EU value of non-discrimination as a key part of human dignity and how the same has been incorporated into the design of AI systems. The paper borrows heavily from the writings of American author and philosopher Friedman who coined the term value sensitive design. Value sensitive has been defined as a process that may be used to review societal values and how they are reflected in technology, AI included.

Values include freedom from bias and informed consent amongst others. Values, however, are not universal as different individuals will consider their values differently depending on what is important in their life. Consequently, the universality of values are those collective values that society ascribe to. Additionally, there is a potential of these values being either reinforced or eroded through the use of AI systems. The erosion could be either through deliberate design or accidental as a result of faulty technology design.

It is very important to analyze AI at the developmental stage to overcome a significant portion of discriminatory challenges that may arise later. I do not make the claim that a proactive approach using value sensitive design and consideration of moral imagination of developers will eradicate technical bias in AI. A proactive approach will bring awareness to regulators, users and indeed developers that some underlying bias may exist in AI systems.

Most importantly, trust in the AI systems by the public will be strengthened when there is a full awareness that fair human values have been considered in the development of the AI system. The risks and concerns that will be discussed in the thesis concern the potential data subjects of AI systems and decisions that those systems will make. Therefore, the scope of the thesis focus is on AI that has an impact on human rights and in particular human dignity.

The paper is structured in five segments. The first chapter provides an introduction to the subject of AI and its key concepts as well. The reader is also provided with a background to the study. The aims of the study as well as the methods employed in the thesis are provided for in the first chapter. The second chapter analyses EU laws and regulations related to AI. Some general challenges of AI and AI discrimination will also be discussed in the second chapter. The third chapter discusses moral imagination together with an understanding of its relevance as per legal guidelines established in the

EU which chapter two will exhaustively explore. The fourth chapter places emphasis on a value sensitive design approach as a tool for inclusion and enhancement of the EU values of non-discrimination. The fifth Chapter, which is a conclusion, presents qualitative discussion findings related to the concerns generated from adapting value sensitive design in AI and the impact of the AI systems on human values.

1.1 Subject

Often times, when people hear about AI, they think of complex machines which must be approached cautiously. Science fiction has prevailed in the general public's understanding of what AI is or ought to be. Additionally, the media often put out misleading news about AI that make the lay person either worry or be too expectant of the technology's potential. The truth of what AI is now is far from glamorous but given the rate at which AI is being continually developed, we are not so far away from differentiating magic and AI system capabilities.

The paper focuses on the moral imagination of AI developers in relation to issues of non-discrimination in AI. Later in the study, an analysis of the value sensitivity of the said technology shall be undertaken. Value sensitive design incorporates values of the society into technology design to enhance the enjoyment of those values. This is very important because moral imagination of the developers is taken into account in trying to tackle discriminatory AI.

Value sensitive design is concerned with what people value and consider important in their societal cycles. It is therefore key in this study to identify the values of the EU as it relates to non-discrimination before delving into how developers of AI incorporate those values into the development of AI systems.

The fundamental values of the EU will be discussed in the first chapter and the paper goes on to look at the moral imagination and value sensitive design in AI in the EU setup. In discussing values, questions of morality and ethics will arise as the same correlate to what values are.

Legal scholars have tried to look at the impact of AI on society, with a bias towards the inevitable changes in labour laws, data protection and automation of government processes amongst others. At the moment, the European General Data Protection Regulation (hereinafter GDPR) is seen as a landmark law concerning the regulation of data (its privacy, transfer and permitted usage) as well as data service providers. Some scholars, however, do argue that the GDPR is inadequate to govern matters of AI and have proposed more regulation.

Some other scholars, albeit in the minority, have argued for less regulation to allow for the unrestricted growth of AI applications. This, they argue, will see the technology grow without any impediments and will in due course allow public access to any new inventions which will allow for solving many of the problems that continue to hinder human development.

This research, however, focuses on an often overlooked aspect of AI. The moral imagination of the developers and the value sensitivity of the design have unfortunately been neglected. Algorithms may, at certain instances, diminish our freedom of choice, right to privacy and also have a bearing on human values. Of concern to the author is the creative process of moral imagination and value sensitive design that is employed before during the development of AI to safeguard against non-discrimination.

At this juncture, I would like to state this whilst this is not a philosophical paper, it is imperative that we borrow Adam Smith's moral sentiment (to be also referred to as moral imagination) and apply it to the AI development so as to ensure fair legal compliance that does not discriminate or exclude minorities. The justification for having an introductory and cursory glance at the philosophical history is that there cannot be a legal clarification on EU values without understanding how society arrives at values to be adhered to.

Some of the moral imagination philosophies that are considered in this paper are empathy being a guiding principle in the inclusive design of AI. Another aspect of moral imagination which will be considered is mutual pleasure. Smith posits that there is mutual pleasure in sharing empathy, that is, to share the same feelings with others.

It remains difficult to isolate values from technology in any discussion because value enhancement and enjoyment are increasingly reliant on technology including AI. The paper also details an analysis of the ethical considerations when programming human centred AI which ideally, ought to encompass ethical and fair considerations per EU requirement. The paper will show through legal analysis whether these considerations are met and whether the law is adequately prepared to deal with the same and any issues therein. Batya Friedman's value sensitive design approach has been used heavily to relate moral imagination to the modern legal scholar.

Value sensitive design in this paper is used to gauge whether there has been universal and fair representation of human values in AI technology to protect against discrimination. The accountability of AI programmers to avoid infringing human rights or degrading the humanity of certain minorities may be measured by using value sensitive design. Consider the situation were Google's algorithm tagged two black men as gorillas in 2015. After being made aware of the problematic results from the algorithm's image identification, Google simply opted to remove gorillas from the image identification[1]. This presents one question of AI development and how human values like those of inclusion of minorities during data training may not be fully taken into account when developing the said systems.

Yohua Bengio once said, "If we built machines that are as smart as us, they will most likely understand our values as well[2]". In short, we should be able to reasonably prevent unfairness and discrimination in decisions by AI if the programming makes use of moral imagination and value sensitive design. Problems will most likely occur once the AI is making decisions and misuse of AI has already been observed as was the case in Cambridge Analytica were AI was used to influence people's right to vote freely.

Two of the starting speculations of using moral imagination and value sensitive design as a framework for studying AI development in the EU are listed below. Each of these speculations will be discussed in detail in the subsequent chapters and weighed again in the final conclusions.

1. Non-discrimination in AI has primarily been left to two actors, the designers of AI and the authorities assigned to ensure non-discrimination. This problem is twofold, leaving designers responsible and the authority only acting when there is breach. The second problem is that other stakeholders, particularly the targeted data subjects are left out on necessarily processes.
2. Finally, that value sensitive design as a framework is suited to study technology and its compliance with societal values. AI complexities mean that value sensitive design may need to be amplified by addition of other methods to be applied to AI as a unique technology.

[1] James V. Google "fixes" its racist algorithm by removing gorillas from its image-labeling tech. (January 2018). Available https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai

[2] Bengio Y., Goodfellow I. and Courville A. 'Deep Learning' (2017). MIT Press

1.2 Research questions, limitations of the study and structure of the research

Technological developments continue to be the main means in which human rights are exercised as can be seen with the freedom of expression which is primarily enjoyed online. By extension, it can be stated that technological advances are fundamental to enhancing human and societal values.

Amongst emerging technologies, AI has the biggest potential to enhance the enjoyment of these fundamental rights. However, it also has the negative consequence of continuing to alienate certain classes from enjoying those rights if current research focuses on finding solutions to challenges that arise out of the use of AI, as opposed to taking a proactive approach that considers the development process as well as interaction of users with AI.

Some challenges presented by AI can easily be avoided by using a proactive approach, particularly value sensitive design. Some examples of problems of AI that could have been avoided proactively by value sensitive design are failure of some protected classes to use facial recognition software due to non-testing of the facial recognition software on certain races. Another problem that could have been proactively addressed with value sensitive design is that of women involvement being left out in the development of some AI health applications such as was the case when Apple Fitness tracker did not have a provision for tracking menstrual cycles when first released[3].

It is therefore in order to ask the main research question;

Can moral imagination and value sensitive design be used as a framework for ensuring EU AI development complies with the value of non-discrimination?

The research question may further be asked in two secondary parts. The first part of the question would concern how discrimination arises in AI at the developmental stages and the second part would try to answer whether moral imagination and value sensitive design can offer a framework for the identification and elimination of the possible factors that lead to bias towards certain classes of people.

The main question has a limitation to human centered AI and the paper will not discuss general AI or machine learning that does not involve humans and their respective rights and freedoms. I believe this

[3] Duhaime-Ross A. 'Apple promised an expansive health app, so why can't I track my menstruation?' (September 2014) available at https://www.theverge.com/2014/9/25/6844021/apple-promised-an-expansive-health-app-so-why-cant-i-track Accessed August 2021

limitation of study is legitimate and appropriate since AI has various angles and analysing anything outside of this scope would render the research question taciturn and unclear.

The study basis considers ethical considerations that are taken in the designing and programming of AI and the resulting effect on the rights of the concerned users. This study is done against a backdrop of the philosophy of moral imagination and its application to the emergence of the technology through value sensitive design.

With that being said, I should stress that the paper will not focus on the philosophical aspect of moral imagination nor the advanced technicalities of the AI but rather the effect that the amalgamation of the same has on the law as it relates to individual and collective freedoms and rights through the prism of value sensitive design.

The research will start by discussing relevant terms in AI and some background of AI shall be discussed further in the main chapter. The legal problems of ethical handling of data and computation of results by AI systems will also be discussed in the first part of the essay.

Legislation relating to governance of AI is complex and numerous. The EU has many directives aimed at researching, regulating, and improving AI. The challenge with studying AI is not the lack of available information, but rather the multitude of legislature to sift through. With that in mind, the study will confine itself to the most important AI regulations as well as the GDPR which has been said to be one of the most comprehensive AI regulation instrument.

The study will also consider the quality of moral imagination that is considered when designing AI systems. Under this discussion, issues of algorithmic discrimination and how they may be redressed shall also be examined. The study will draw some comparisons from the North American experience to understand the possible extent of algorithmic discrimination. Examples from North America and other jurisdictions are provided so as lay a foundation for a minor comparative study which will provide a learning platform from the mistakes that may have been made.

The study begins with looking at the current legislation related to AI and builds the legal framework. Thereafter, the historical and mostly theoretical theory of moral imagination is looked at and the study moves to the more practical framework of value sensitive design. Value sensitive design as a framework will be used to analyze the effects of AI on non-discrimination. Human values of the EU and how they fit in with AI will also be discussed, that is, whether AI is enhancing those values or diminishing them.

In my concluding chapter, after having analyzed the EU laws governing AI and the often un-scrutinized salient process of moral imagination that is employed in the development of AI, the paper will underscore and stress the importance of incorporating value sensitive design to avoid algorithmic discrimination caused mostly by lack of diversity in developers, use of incorrect data in developing AI as well as failure of properly setting EU human values as a bare minimum, a failure which I believe can be remedied by incorporating value sensitive design as the study will show. This chapter will also provide recommendations for improving AI development in the EU to avoid discriminatory decisions.

1.3 Methods and sources

The research approach is primarily qualitative legal research. Dogmatic and legal doctrinal research are also applied. Additionally, the research in nature will be explorative as it tries to incorporate value sensitive design with AI, a relatively new way of addressing discrimination in AI. Consequently, building on the concept of using value sensitive design as a framework means that the study will take on an instrumental method of study in achieving its goals.

In the research paper, I will analyze the main sources of EU law that govern AI with brief comparative study by quoting some North American law where necessary. AI and its relationship to law remains quite complex and ever evolving and this is evident by the eighty four AI policies that have thus far been enacted globally[4]. The study limits itself to the EU jurisdiction and I will study the relevant laws as well as pay particular attention to the GDPR seeing as it has been touted as the most relevant legislative Act as relates to personal data, the premise of AI.

I will also make use of legal treatises as well as legal research materials that continue to be published in the field of AI and law. The challenge is that there are often new developments in AI within a short space of time which render older information outdated. This does not change EU fundamental rights but calls for ensuring novel application of AI to core values of the EU, a challenge that value sensitive design may address. The books that I will rely on cover human values such as non-discrimination that are affected by AI systems and it is the hope of the author that these books will present a valuable

[4] Jobin, A., Ienca, M. & Vayena, E. ‘Artificial Intelligence: The global landscape of AI ethics guidelines’ (2019). Nature Machine Intelligence Volume 1, at page 3.

foundation for future research into AI systems development and its inclusion of EU values at the developmental stage.

Case law also makes up a small part of the data that is to be collected in reviewing AI, its social ethics (moral imagination) and also some cases that have bordered on human rights discrimination as a result of algorithmic decisions. The few cases highlighted in the essay are meant to show how the European courts have interpreted the GDPR as it relates to freedoms and rights. Case law is reactive, and the focus of the research is on how to prevent future challenges caused by AI decisions when there has been no compliance with EU values. It is in this regard that case law will only be used to show problems that could have been prevented proactively had there been adequate inclusion of EU values in the design process by using value sensitive design as a framework.

2.0 LEGAL FRAMEWORK OF AI AND NON-DISCRIMINATION IN THE EU.

Under this Chapter, the laws that govern AI in the EU will be discussed at length. The EU directives governing AI will be mentioned as will specific Articles in the relevant directives that pertain to individual's rights as it concerns algorithmic data processing. The data processing is one that has an influence on the individual's human values as enshrined under the Charter of Fundamental Rights.

2.1 Introduction to AI

Rouhianen gives a very concise definition of AI in his introductory chapter in which he defines AI as the ability of computers to do things that would normally require human intelligence[5]. Machines use computer programming to learn from a given set of data and make decisions by identifying a preset condition in a data set and to give a corresponding result based on the identified data[6].The accuracy of any automated decisions depends heavily on a large data set of already collected training data which the system will use to calculate a probability and arrive at a decision[7].

The focus of this paper is on the first class of AI which is referred to as narrow AI, this being an intelligence that is designed to perform one specific task like identifying objects in images for facial recognition[8] or indeed sorting out fruit in a factory for packaging. The second class of AI is Artificial General Intelligence (AGI), which is AI comparable to AI operating like the human brain whilst the third is Artificial Super Intelligence (ASI), which is AI with the capacity to exceed the human the brain[9]. The last two classes of AI have not yet been realized and thus will not be discussed in this paper.

One of the strongest arguments for employing AI in human society is that AI can reduce bias, unfair application of the law or indeed corruption. However, it has been shown that AI infers rules from historical patterns including from situations where variables such as race or sex may be used to

[5] Rouhianen L. 'AI: 101 things you must know today about our future' (2018). At page 3

[6] Ibid at page 8

[7] Hildebrandt, M.'Law As Computation in the Era of Artificial Legal Intelligence. Speaking Law to the Power of Statistics' (June 2017). Available at SSRN: https://ssrn.com/abstract=2983045 at page 9

[8] Coleman F. 'A Human Algorithm: How AI is Redefining Who We Are' (2019). Counterpoint. Berkeley at page 104

[9] Rickli J.M.'The Economic, Security and Military Implications of Artificial Intelligence for the Arab Gulf Countries, EDA Insight' (November 2018). At Page 2.

negatively affect certain classes of people[10]. These challenges that AI presents have given rise to a number of international and EU regulations as well as additional national regulations on AI.

As of writing, there are well over 84 directives and guidelines for the regulation of AI globally[11]. In the EU, AI regulations include the White Paper on AI, the European Approach to Artificial Intelligence and Robotics as well as the GDPR which has been touted by some as the best regulation for algorithmic processing.

AI has a lot of positive applications which need no mention but a select few within the scope of this paper are that AI extends learning opportunities to people and provides easier communication. AI is also useful for novel ways of entertainment and provides a better opportunity for working interactively[12].

Improvements in healthcare systems and healthcare management have been observed when AI has been introduced and adopted. Additionally, access to justice has been shown to be increased by use of AI, although it remains problematic in criminal justice[13].

All of the positives of AI are dependent on AI that is fair and impartial as well as value driven. The AI must meet criteria for good data collection and must also be transparent and open for review by third parties[14].

Whoever designs the most effective AI will control the future of not only technology but military warfare, finance as well as acquire economic advantage, the European Commission observed in a 2018 report. The move to dominate AI development is understandable and the EU is distinguished by its

[10] Zalnieuriute M, Crawford L.B., Boughley J., Moses L.R. and Logan S. 'From Rule of law to Statute Drafting: Legal issues for Algorithms in government decisions making' (2019). In Woodrow B. (edt) Cambridge Handbook on the law of Algorithms. Cambridge University press. Pages 16 and 18

[11] Ibid at page3

[12] Hasse, A., Cortesi, S., Lombana-Bermudez, A., & Gasser, U. 'Youth and artificial intelligence: Where we stand. Youth and Media' (2019).Berkman Klein Center for Internet & Society. Retrieved from https://cyber.harvard.edu/publication/2019/youth-andartificial-intelligence/where-we-stand at pages 9, 15 and 20

[13] OECD, Artificial Intelligence in Society (2019), OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en At page 60-71

[14] Broeders D., Schrijvers E., Sloot B., Brakel R., Hoog J. and Ballin E. H. 'Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data' (2017). Computer Law & Security Review. Volume 33. No. 3. 309-32

reluctance to push for AI dominance at the expense of neglecting fundamental rights and freedoms of its citizens and indeed those beyond its borders[15].

This is why the EU has established a number of regulatory frameworks for AI that protect those rights. One of these fundamental aspects of AI development concerns the preservation of EU values in the designing of AI. Values of humans must drive innovation of AI, and AI ought to enhance those values, as opposed to hindering the enjoyment of those values[16].

This brings the discussion to the next section of this paper which is analyzing AI in the context of its compliance with EU values and particularly the right to non-discrimination.

2.2 Values of the EU

Article 1 of the Charter of Fundamental Rights states,

"Human dignity is inviolable. It must be respected and protected"17

Human dignity dates to pre-modern history and has its basis in theology as well as in philosophy. In this study, human dignity as a concept is used to include the prohibition of discrimination. This is justified because the philosophical and historical base of human dignity that will be discussed here could not foresee the emergence of technologies such as AI that would be used to strengthen human dignity.

For one of the most discussed aspects of human values and right, it is rather odd that there is no one standard definition of what dignity is[18]. We can still take comfort in that the general concept of what dignity is remains the same in the academic circles since definitions basically perform two roles, these being to clarify the meaning of words and secondly to increase our understanding of what those words refer to[19].

[15] Renda, A. 'Artificial Intelligence Ethics, Governance and Policy Challenges Report of a CEPS Task Force' (2019). Retrieved from https://www.ceps.eu/wpcontent/uploads/2019/02/AI_TFR.pdf
[16] ibid
[17] Article 1 EU Charter
[18] McCrudden C. 'Human Dignity and Judicial Interpretation of Human Rights' (2008). The European Journal of International Law. 655 655.
[19] Jabłonowska A., Kuziemski M.,, Nowak A.M.,, Micklitz H., Pałka P., and Sartor G. 'Consumer Law And Artificial Intelligence Challenges To The EU Consumer Law And Policy Stemming From The Business' Use Of Artificial Intelligence' (2018). EUI Working Paper Law 2018/11. At Page 4.

This paper will borrow Kant's viewpoint on dignity since he may be said to be the foremost scholar who attempted to define human dignity. Kant argued at length, stating that human dignity accrues to all persons, irrespective of their rank in society, that human dignity was something everyone was born with and therefore could not be taken away by anyone. He further stated that human dignity elevates humans above animals, and as such, humans should be allowed to exercise free will, pursue their own goals and act with morality[20].

To Kant, free will meant that human beings were at liberty to avoid acting on instincts and thus exercise control over the environment. Furthermore, freedom in Kant's definition of human dignity meant being able to make your own laws, as opposed to being subjected to the authority of another person[21]. What follows from Kant's theory of human dignity as freedom and autonomy also places on the person responsibility arising from that person's willful actions[22].

Under the preamble of the EU Charter, human dignity as a concept is always considered with other rights and freedoms and human dignity remains a core aspect of EU values. Furthermore, human dignity remains universal as can be seen by the adoption of human rights by nearly all countries. Thirdly, human dignity as expounded by Kant is legalized in the preamble of the charter wherein there is a consideration of placing responsibility of enhancing human dignity on individuals within society and indeed extended the responsibility to communities[23].

It is common practice for courts to now define what human dignity is once it has been violated. Knowing the consequences of violating human dignity, the definition of human dignity, simply put, is the state of being a person, or the quality of existence that one has as a result of being a person[24]. Note must be taken that human dignity does not equate to human rights but the two complement one another, and may, in most instances be synonymous[25].

[20] https://plato.stanford.edu/entries/kant-moral/ for a lengthy discussion, (accessed on 14th June 2020)

[21] ibid para 2

[22] Loc. cit.

[23] Jones J. 'Human Dignity in the EU Charter of Fundamental Rights and its interpretation before the European Court of Justice' (December 2012). Springer Science + Business Media. Dordrecht at page 284

[24] Dearing A. 'Justice for Victims of crime: Human Dignity as the Foundation of Criminal Justice in Europe' (2017). Springer International Publishing. Cham at page 139

[25] Nida-Rumelin J. 'Why Human dignity rests upon Freedom' at pages 84-92 available at https://www.researchgate.net/publication/325654586_Human_Dignity_as_a_Foundation_of_Law (accessed on 3rd June 2020)

There was a tremendous shift in concepts of human dignity and related rights after World War II in which international organizations begun including human dignity as a fundamental right. An example is the Universal Declaration of Human Rights (UDHR) and how other new countries emerging from colonialism in the 1950s adopted human rights in their Bill of Rights[26].

The UDHR preamble has emphasized the recognition and respect of human dignity. Human dignity under the UDHR includes the right to education and the right to work. The effect that the right to work has on one's ability to contribute to society is also part of human dignity. Prohibitions of inhumane treatment as well as equality before the law are all aspects of human dignity. Suffice it to state that there may be no extent to what human dignity encompasses, as it concerns all human rights, and the freedom to enjoy the said rights.

The legal history of human dignity may also be traced back to the German constitution of 1919 and the 1937 constitution of Ireland which made an explicitly mention of human dignity in the preamble[27].Human dignity has been strongly tied to questions of morality as opposed to law because for the longest time, theorists focused on human dignity as a philosophical and theological course as opposed to a legal once. It remains that it is still difficult to discuss human dignity and its sources without invoking questions of morality and where they come from.

Human dignity is the foundation of all other modern rights such as data privacy[28]. Building on from this EU value, it is pertinent to discuss the other values of the EU in order to understand whether the AI regulation is in line with the values. Aizenberg and Hoven have specified the fundamental values of the EU in relation to human dignity and its relationship to non-discrimination[29]. The authors show that non-discrimination is a component of human dignity and that it ties in with all fundamental freedoms and is especially related to the concepts of equality and freedom[30].

[26]ibid

[27] Kirste S. 'Human Dignity as a foundation of Law' (June 2018). available at https://www.researchgate.net/publication/325654586_Human_Dignity_as_a_Foundation_of_Law page 64

[28]ibid

[29]Aizenberg A, and Hoven J 'Designing for Human Rights in AI' (2018). Sage Publication. At page 6, figure 2.

[30]Loc.Cit.

Values are not universal as the sources of our human values are unique to our communities and most of the questions raised about morals arise because of the differences in our values as well as in instances where our values are in conflict with those of other people[31].

The definition of what values are is fraught with complications, suffice it to state that values involve how we relate to our environment and others within the same environment. Values are tied to our wellbeing, our quest for self-development and also our interpersonal relationship and are also related to our social interactions and institutions[32].

Examples of values are private life, justice coupled with fairness and indeed equality[33]. It may be said that nearly every aspect of EU values are enjoyed by using technology and increasingly relying on AI which presents both a challenge and opportunity to design technology that recognizes these values and enhances them as well. Freedom of expression, freedom of association and freedom of speech are enjoyed online through various technologies including social media whilst right to employment, access to some financial services and other public services have begun to rely heavily on AI.

The EU Charter of Fundamental Rights in the preamble reads,

"…the Union is founded on the indivisible, universal values of human dignity, freedom, equality and solidarity; it is based on the principles of democracy and the rule of law. It places the individual at the heart of its activities, by establishing the citizenship of the Union and by creating an area of freedom, security and justice."[34]

Importantly, as it relates to developments in technology and how the exercise and enjoyment of these freedoms and values can be exercised, the Union acknowledges in the preamble the necessity of strengthening the protection of Fundamental Rights in light of changes in society and technological developments by making the fundamental rights and values more pronounced.AI has proven to be one such technological change that may indeed enhance the enjoyment of these rights. AI must be cautiously streamlined to enhance, and not curtail these values.

[31] Johnson M.'Morality for Humans: Ethical Understanding From the Perspective of Cognitive Science' (2014). University of Chicago Press. At page 52

[32] Ibid at page 65

[33]Gutierrez M: 'The Good, The Bag and the Beauty of Good Enough Data' in Daly A., Devitt S.K and Mann M. Eds (2019) Good Data. Institute of Network Cultures. Amsterdam at page 54

[34]Charter Of Fundamental Rights Of The EU (2012/C 326/02) preamble

Of interest to AI is Article 8, the right to protection of personal data as well as Article 11, the right to freedom of expression. These rights have been the focus of legal scholars as issues of consent to data processing remains problematic as well as the exercise of freedom of expression which is done largely on social media. Social media is liable to be tempered with by foreign elements using AI agents which may have a direct impact on fundamental rights[35].

Article 21 prohibits discrimination on the grounds of race, colour, genetic features, and ethnic minority amongst others. Inadvertently, some AI systems such as facial recognition software may be said to discriminate on the basis of failure to recognize facial features of different races or to identity a person of colour as well as rare cases of intentional discrimination by the AI developers.[36]

Value effects of technology depend on the deliberate design process and technologies such as AI have the capability to enhance enjoyment of values when designed proactively. However, bias in AI systems may still exist in an ambiguous state and the AI systems may amplify the bias. It is therefore important that stakeholders are aware of the inherit risk in AI and how it may unintentionally curtail certain freedoms and rights.

The European Court of Human Rights[37] is a court that has been tasked with interpreting human rights where there are conflicts and the court is an ideal institutional resource for understanding the values of the EU as it relates to human rights, AI and non-discrimination.

Article 14 of the European Convention on human rights binds EU member states to respect right of non-discrimination[38]. This right was cemented in protocol No. 12 which protocol gives member states discretion to look at issues of discrimination more broadly[39]. Furthermore, Article 19 of the Treaty of the Functioning of the EU allows EU laws to prevent discrimination on the basis of sex, race and age

[35] Gavazzi M. S (June 2020) By Monitoring social Media, AI is guiding College Towns through Covid 19, the article states that AI was used to scan conversations with the words covid and corona virus to pinpoint certain areas, what's not clear is whether there was consent from the many users or that data privacy was breached. Consider also the Cambridge Analytica example were there was data misuse of millions of accounts. Articles available at https://www.forbes.com/sites/stephengavazzi/2020/06/17/by-monitoring-social-media-artificial-intelligence-is-guiding-college-towns-through-covid-19/#47c3985c494c and https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

[36] Bourgesius Z.F. 'Discrimination, AI and Algorithmic Decision-making. Directorate General of Democracy' (2018). Council of Europe available https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

[37] Also referred to as the Strasbourg Court established by the EU Convention on Human Rights

[38] ECHR 14

[39] ETS 177-Convention for the Protection of Human Rights (Protocol No.12) 4 XL 2000

amongst other characteristics. Similarly, Article 2 of the TEU acknowledges the inviolability of human dignity as a core value of the EU.

2.3 EU regulation of AI

There are several policy guidelines and EU commission directive that may be used for AI[40]. The challenge as stated earlier, is having so many regulations and any researcher is hard pressed to isolate the documentation before they begin to analyze the issues of interest to them. Additionally, most data protection laws are compatible with the regulation of AI to some degree as the data protection laws try to provide some level of control to the data subject and the data protection laws also regulate the processing of users' data[41].

The paper focuses on the key directives and regulatory instruments that deal specifically with AI such as the White Paper on AI[42]. Other Commission directives of importance to AI are the European Commission's European Strategy for Data Communication and the European Commission's report on the Safety and Liability of AI, Internet of Things and Robotics[43].

The Convention for the Protection of Individual About Automatic Processing of Personal Data[44] is a very important regulatory instrument that seeks to address the challenges that may be posed by processing of personal data.

The white paper's greatest weakness is in the defining area of application of the said paper because there still is a difficulty amongst scholars in universally agreeing to what amounts to AI. An alternative to this would be to define AI as machine learning and this may ease the problem as most computer scholars agree on what is meant by machine learning. Ideally, simply changing the name to machine learning takes away the fictitious expectation of AI being extremely unpredictable.

The GDPR has an important use in AI, particularly Articles 21 and 22 which concern the right to object to automated decision-making including profiling[45]. Article 22 in particular states that individuals,

[40] Hacker P. 'AI regulation in Europe' (November 2019). Humboldt University of Berlin. Availablehttps://papers.ssrn.com/sol3/papers.cfm?abstract_id=3556532

[41] At page 21 data protection and discrimination

[42] European Commission, On AI - A European approach to excellence and trust, White Paper, COM(2020) 65 final

[43] COM 2020/66 and COM 2020/64

[44] Council of Europe Convention no.108 amended by protocol CETS No. 223 which protocol may be useful for matters involving AI systems.

"shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her[46]".

The above Article assumes that the user is aware that they may be subjected to an automated decision, and further assumes that there may be legal effects that may affect them. The GDPR is noted for its focus on giving users control over their data as it relates to profiling and big data wherein the person provides data to the AI system and the AI system makes a decision[47].

The GDPR, it must be noted is only applicable to AI when personal data is processed[48]. Additionally, the GDPR has a provision for ensuring that there is fairness in the decisions made by AI systems where personal data has been processed[49]. The GDPR has further provided avenues for addressing the outcomes of algorithmic decisions[50]. The rules are in place to avoid discrimination of concerned data subject, whether that discrimination is intentional or not.

Articles 13 to 15 of the GDPR require data subjects to have access to what is termed meaningful information concerning how algorithms work in making decisions that concerns them. This, in practice, is inscrutable. Inscrutability refers to a circumstance in which decision making by an algorithm is dependent on several other interlinked factors which makes it nearly impossible for a logical inspection to be undertaken in reviewing the process of the decision making. The challenge with inscrutability is not that the data subjects or controllers are lack awareness but rather that the process of undertaking a review is complex and extremely difficult even for people with expertise[51].

[45] Guido Noto La Diega.'Against the Dehumanisation of Decision-Making' (2018). Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information 9. JIPITEC 3 para 1.

[46] Regulation (EU) 2016/679.

[47] Mitrou L. 'Data Protection, AI, and cognitive services: Is the GDPR AI Proof?'(April 2019). University of the Aegean at page 23

[48] Ibid at page 74

[49] Butterworth M.'The ICO and Artificial intelligence: The role of fairness in the GDPR framework' (2018). Computer Law and Security Review 34.

[50] Article 28 of the GDRP

[51] Robert Brauneis and Ellen P. G.'Algorithmic Transparency for the Smart City' (2018). 20 Yale Journal of law and technology. 103, 107–08

Guidelines on AI and its use in judicial systems have also been set in the European Commission's Charter on The Use of AI[52].Another important AI regulation tool is the European Ethical Charter on The Use of AI (AI) in Judicial Systems and Their Environment[53] which lists some core principles related to AI and justice such as, the respect of fundamental human rights, that is, the design of AI ought to comply with human rights.

Another core principle of the Charter is non-discrimination of individuals as well as the principle of quality and security. Put simply, processing of judicial decisions and data must be secure and of a very high and traceable quality. Transparency is also a core principle and as such, allowance of external auditors for AI systems must exist[54]. Transparency remains a major concern for AI accountability[55].

There remains multitude of legislation that one has to go through in order to analyze an AI issue and the position of the law on the same[56]. This is so because discrimination does not necessarily occur in a vacuum and cannot be identified minus isolating a law being breached. For example, in cases where a person alleges discrimination on grounds of race, they may have to refer to the Race Equality Directive[57] in addition to any grounds of discrimination. Similarly for one to allege algorithmic discrimination on the grounds of sex, they may have to refer to the Gender Recast Directive[58] in addition to the relevant facts of the algorithmic discrimination.

[52] European Commission for the Administration of Justice (CEPEJ). (2018). European ethical charter on the use of AI in judicial systems and their environment. Retrieved from: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

[53] Loc. Cit.

[54] Mittelstadt B.'Auditing for Transparency in Content Personalization Systems' (2016). International Journal of Communication Volume 10. Page 4995

[55] Buhmann, A., Pabmann, J., & Fieseler, C. 'Managing Algorithmic Accountability: Balancing Reputational Concerns, Engagement Strategies, and the Potential of Rational Discourse' (2019). Journal of Business Ethics. At pages 3 and 4.

[56] Saluzzo S.'The EU as A global Standard Setting Actor: The case of Data Transfer to third countries' (2017). In Carpanelli E. and Lazzerini N. Edts (2019) Use and Misuse of New Technologies: Contemporary Challenges in International and European Law. Springer Publishing.

[57] Directive 2000/43/EC

[58] Directive 2006/54/EC

2.4 AI and Human Rights

In order to safeguard our future as humans, we need to have a humane approach to the intelligent machines that we continue to develop. As simple this appears, there are hurdles on how to practically apply this humanity to the intelligent machines in question[59].

AI, being what it is, requires strict impartation of universal human values. AI is already a major human rights issue in developed countries and it will continue to change our lives as well as how we function as a collective human society across the globe[60]. The recruitment of employees has largely been left to automated processes, and some job applicants never have their applications reviewed by a human and are rejected by AI system[61].

The IEEE Global initiative for ethical considerations in AI and autonomous system may have an answer for incorporating human values into AI systems[62]. The solution, it appears, would be to implant human values into the AI system, taking note that human values are not universal and the said values must reflect those norms and morals that are acceptable in the society in which the AI system will function[63].

A good example for a basic starting point would be the UDHR as the bare minimum of the technology design[64]. Additionally, the High-Level Expert Group on AI[65] proposed these fundamentals for having trustworthy AI; Respect for human autonomy, prevention of harm, fairness, and explicability[66].

[59] Coleman F. 'A Human Algorithm: How AI is redefining who we are' (2019). Library of Congress Cataloguing-in-Publication Data

[60] Aizenberg A and Hoven J. 'Designing for Human rights in AI' (August 2018). Sage Publication. Available on https://journals.sagepub.com/doi/full/10.1177/2053951720949566 at page 11.

[61] Tasioulas J. 'First Steps Towards an Ethics of Robots and Artificial Intelligence' (2019). Journal of practical Ethics. Volume 7. Issue 1. At page 52

[62] Shahriari K. Et al 'Ethically Aligned Design: A Vision For Prioritizing Wellbeing With AI And Autonomous Systems, Version 1. IEEE' (2017). http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html. accessed 13th May,2020

[63] Van den Hoven, J. 'Engineering and the Problem of Moral Overload' (March 2012). Science and Engineering Ethics 18, no. 1 (Note must be taken that AI has the ability for global reach and as such, societal values in which the AI is developed will extend beyond its borders, potentially being in conflict with other values).

[64] Elsayed-Ali S. 'New Human Rights Principles on Artificial Intelligence' available at https://www.openglobalrights.org/new-human-rights-principles-on-artificial-intelligence/

[65] European Commission (2019) Ethics Guidelines for Trustworthy AI

[66] Mittelstadt B. 'Principles Alone Cannot Guarantee Ethical AI' (November 2019). Nature Machine Intelligence, November 2019. Page 9 Available at SSRN: https://ssrn.com/abstract=3391293

Explanation for autonomous decisions made by algorithm is an important tool for ensuring that the AI systems do comply with human values as well as providing some sort of transparency from the proverbial black box[67].

Secrecy is another concern of algorithmic decision-making and ensuring transparency of the same. Secrecy of decision making has two factors with the first being the secrecy of the model's existence. This first part concerns intellectual property rights that developers of algorithms use to protect their financial interest.

The second secrecy also has its foundation in intellectual property protection, particularly the trade secret and the secrecy of its operation[68]. In the second secrecy, the existence of a decision-making process is known in the secrecy of operation, but the inner working of the AI system is kept secret. Potential data subjects are sometimes aware that they may be subjected to an algorithmic decision-making process but are often either ignorant of how the decision-making process operates or have very modest knowledge of the operation[69].

Article 7 of GPDR gives explanation for when a data subject may give consent. Further Article 12 allows for transparency in the intended use of the data, this article, simply put, requires the data controller to inform the user of how the data will be used[70]. Articles 13, 14 and 15 have the effect of trying to define meaningful information. Meaningful information is made out to be information that is provided to the data subject and ought to have the explanations of how, when, and why as it relates to data processing[71]. Different authors have presented different arguments on what could be meant by

[67] Edwards, L. & Veale, M. 'Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for' (2017). Duke Law & Technology Review, 16(1), 18-84. doi:10.2139/ssrn.297285

[68] Selbst A.D. and Barocas S.'The intuitive Appeal of Explainable Machines' (2018). Fordham Law Review Volume 87. At page 1091

[69]Ibid

[70]Varkonyi G.G. 'Operability of the GDPR's Consent Rule in Intelligent Systems: Evaluating the Transparency Rule and the Right to Be Forgotten' (2019). Intelligent Environments. Available athttps://www.academia.edu/39691158/Operability_of_the_GDPR_s_Consent_Rule_in_Intelligent_Systems_Evaluating_the_Transparency_Rule_and_the_Right_to_Be_Forgotten

[71]Selbst A.D, and Powles J.'Meaningful information and the right to explanation' (2017). International Data Privacy Law, 7(4). Available at https://academic.oup.com/idpl/article/7/4/233/4762325

meaningful information[72]. Note must be taken that the GDPR promise of reigning in data controllers who may use their data for AI training is amiss due to the many uncertainties of AI[73].

The use of algorithms is under scrutiny with profiling being a major source of scrutiny[74]. EU regulators are aware of profiling as a problem as well as the negligible information that is provided about how algorithms work and subsequently how they influence the data subject. Data subjects have a right not to be subjected to a decision which affects them in their personal lives if the decision is as a result of automated processing of data which evaluates some of aspects of their personal life[75].

Furthermore, withdrawing consent and trying to make use of the right to be forgotten is a nearly impossible task when it concerns automated systems. The data controller has the responsibility to remove a data set but isolating one individual's data from large datasets is technically a challenging undertaking, it is also a challenge in the legal sense. Data controllers of AI systems will find it difficult to effectively comply with the right to be forgotten because of the complexity of the AI systems[76].

There is also the possibility that limited or no access to information may result in exclusion[77]. The freedom of expression includes the right to access to information, but from the technological perspective, this includes only those who have means and tools to access information and are able to use it[78]. In spite of its promise, AI may increase the risk of amplifying digital inequalities between people with access to technology and certain classes such as people who may not afford access to the

[72]Watson, H. J., & Nations, C. 'Addressing the Growing Need for Algorithmic Transparency' (2019). Communications of the Association for Information Systems Volume 45 No 26, at page 496-497.

[73]Reed C. 'Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning' (March 2018). Queen Mary School of Law Legal Studies Research Paper No. 243/2016) 29 <https://ssrn.com/abstract=2853462> 276 Bygrave (n 165) 22.

[74]Kim, P. T. 'Auditing Algorithms for Discrimination' (2017). University of Pennsylvania Law Review Online, 166(1), 189.

[75]Diega G.N.'Against the Dehumanization of decision-making' (2018). Algorithmic Decisions at the crossroads of intellectual property, data protection and freedom of information. JIPITEC 3. Para 1

[76]Wachter S, Mittelstadt B, and Floridi L. 'Why a Right to Explanation of Automated Decision-making Does Not Exist in the GDPR' (2017). International Data Privacy Law volume 7 No. 2 at pages 81,212

[77]OPTICS Ethics and Technology: (Re)Building Trust in technology(2019). Published by UNESCO, at pages 27-28. Data bias: biased Data produces biased results.

[78] Bjork A., Paavalo J., Strik T., Tanhua I. And Vainio A. 'Finland in the International Human Rights System' (2019). Publication of the Government's analysis, Assement and Research activities 50. Prime Minister's Office. At page 50

internet and have no traceable data to be used in the development of AI systems and AI may further amplify gender and racial biases[79].

Furthermore, most scholars studying the impact of technology on human rights base their studies on developed countries[80], whilst there is an obvious exclusion in this scenario, we can also see that exclusion from data banks also disadvantages certain classes such as the elderly whose use of the internet and potential AI data sets may be negligible as compared to younger demographic[81].It is of course possible to design AI that may recognize unethical acts as well as data exclusion but as is the case with AI, the certainty of their identifying these acts cannot be guaranteed and even if it were to be guaranteed, the problem of data exclusion would still exist[82].

Aside from the negative effects of data exclusion, improper or incomplete data collection affects human rights in a negatively[83]. AI systems relay on provided data sets, and the unavailability of data or wrong data will clearly disadvantage the human rights of the concerned subjects[84].

Human rights are increasingly being subjected to automated decisions that can occur at extremely fast rates that were not previous imagined. Some rights and benefits such as work opportunities or access to social services including health and housing are now being decided within seconds by algorithms. There is need of government involvement in ensuring that AI can be relied upon to not only provide a conveniently fast service but also ensure a fair service that obeys human rights as established by international frameworks[85].

[79] Joshua P. D.'Law Without Mind: AI, Ethics, And Jurisprudence' (May 2018). University of San Francisco Law Research Paper.

[80] Crawford, Kate, and Ryan Calo 'There Is a Blind Spot in AI Research' (2016).Nature 538:311–313.

[81] Burton E., Goldsmith J., Koenig S., Kuipers B., Mattei N. And Wash T. 'Ethical Considerations in AI Courses' (June 2017). AI Magazine. Association for the Advancement of AI at page 25

[82] Pauline K. 'Data-Driven Discrimination at Work' (2017). 48 Will and Mary Law Review 857, 90209.

[83] Molner P.'Technology on the margins: AI and global migration management from a human rights perspective' (2019). Cambridge International Law Journal. Volume 8. No. 2. At pages 315-316

[84] Veale M. and Binns R.'Fairer Machine Learning in The Real World: Mitigating Discrimination without collecting Sensitive data' (November 2017). Available athttps://journals.sagepub.com/doi/10.1177/2053951717743530

[85] Yeung K. Howes A. and Pogrebna G. 'AI governance by Human Rights-centered Design, Deliberation and oversight: an End to Ethics Washing'. In Oxford Handbook of AI ethics (2019). Dubber M and Pasquale F. (Eds). Oxford University press. Page 78

Most scholars have argued time and again, that AI systems are best regulated using already established human rights conventions[86]. This is logical for various reasons, one of which is that international human rights conventions have been accepted by most countries, and with the naturally overreaching AI systems, it appears to be a natural route to take.

For AI systems that concern humans, human rights ought to be the guiding principle. Increasingly, most human rights and freedoms are being expressed and enjoyed though technology such as social media for freedoms of expression and right to free speech. And newer technologies such as AI ought to evolve to protect these existing rights, it is expected that technology ought to mold itself to the laws in place to enhance the protection of those rights[87]. International human rights offer the most universally agreeable standards of what ethical concerns should be adopted in the design of AI systems. The EU, as it concerns human rights and technology, remains committed to a system of valuing human rights above any new technological advances that may occur[88].

Designing human centered AI systems means setting human rights as a priority base in the design process of AI systems[89]**.** There are four principles that may be employed in AI design to prevent discrimination as well as to enforce and enhance other human rights.

1. Design and deliberation

AI systems must be designed in accordance with universal human rights conventions, an example in the EU would be designing AI systems that have the European Convention on Human Rights Articles as a base and they can only enhance those values and not diminish them. Deliberation means a conscious effort during the implementation phase to reflect those values in the AI system.

2. Assessment, testing and evaluation

[86]Scherer M. U.'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, And Strategies' (2016). Harvard Journal of Law & Technology. Volume 29, Number 2 Spring. Pages 363-364

[87]Yeung K. 'A study of the Implications of Advanced Digital Technologies (including AI systems) for the Concept of responsibility within a Human Rights Framework' (2019). Council of Europe. At page 13

[88]ibid

[89] Consider Chapter 4: Value Sensitive Design. Chapter will provide explanation on embedding values like Human Rights into AI.

People would only be willing to subject their rights to an AI system if that system has been tested and evaluated on its ability to favorably consider human rights in any situation. Furthermore, those subject to these systems would want an assurance that the system may be challenged during its course of operation to gauge its ability to act as a human would. This is very important for transparency and limiting and isolating the black box factor that is often attached to algorithmic decisions.

3. Independent investigation and oversight

Closely related to the second principle, is the requirement that these systems ought to be able to be subjected to an independent overnight and investigation if and when one wishes to. There ought to be institutions, preferably from the government who it is assumed, have no monetary interest in the operation of the systems but play a role for the general public and are thus expected to be independent in their investigations of the AI systems. It is further suggested that these independent investigators should have the ability to sanction the companies and the developers of the systems so as to ensure total compliance to human rights values.

4. Traceability, evidence, and proof.

There ought to be mechanisms for meaningful reviews of AI systems that allow an audit. The audit should be logical and measurable by established standards and should be easily available to any authority that is being tasked with the review of the systems. Traceability ought to apply from the development phase throughout the life cycle of development of the systems. The aim of this fourth principle is not to ensure comprehension of the system but rather to provide for adequate human control[90].

2.5 AI and Bias

AI programs depend upon data they collect, and existing societal or structural biases may be subtly embedded in data sets and unwittingly carried onto an AI program which program may continue to be biased against certain protected classes. Much of auditing of algorithms focuses on repeated errors or those errors with severe consequences and may ignore subtle inefficiencies which may still discriminate[91]. Additionally, even where the algorithms is open for public audit, it is often opaque and cannot be interpreted easily even by specialists. In the EU, it has been observed that bias does occur

[90] Yeung K. Howes A. and Pogrebna G. (2019) cited at 85. Pages 89-93

[91] Brent M. 'Auditing for Transparency in Content Personalization Systems' (2016). International Journal of Communication. Volume 10. Page 4995

due to inaccuracy of data which may result in the violation of the right to non-discrimination. This inaccuracy may be caused by the data quality ad well as the risks uneven profiling on good data[92].

Noble[93], speaking of Google, wrote that there is bias in search engine results. The search engine company effectively blocks sites that compete with it and places businesses that it owns on the top results such as YouTube when one searches for a video or Google images over any other photo websites as well as having a preference for its maps when one enters location. There is nothing wrong with this business practice, but it shows the level of control that the company has for top results.

Bias, most of the time, is unintentional because the developers fail to foresee any possible negative effects that may arise against protected classes. The developers may fail course the bias due to not overlooking certain aspects of the AI development or unconsciously designing the AI with their personal bias such as race and sex. Ultimately, the unintentional bias is as a result of lack of imagination on the part of the developers[94].

Another example of institutionalised biased being amplified by technology is how auto completion for searches of women have ended in discriminatory and sexist suggestions. This is problematic in that results for black woman have been sexualized. Google's response has been that the results are because of user input that the algorithm learns. The response by the technology giant shows that they are aware of existing social injustice and perhaps are unwilling or unable to challenge user views.

Noble goes on to state that there exists screening and content will only be shown to reflect United States values that society ascribes to, and these unfortunately include racial and stereotypical assumptions about women and minorities. I give this example to show the difficulty of what ought to be fair and equitable human values, and that the developers of technology do have the duty to ensure amongst others, universally acceptable standards of human values.AI development has shown instances where it continues to be trained on data sets that are riddled with data gaps such as the incident when

[92] European Union Agency for fundamental Rights Report. 'Getting the Future Right: Artificial Intelligence and Fundamental Rights'. Publication of the European Union. At pages 57-58

[93] Noble S. U. 'Algorithms of Oppression, How Search Engines Reinforce Racism' (2018). New York University Press.

[94]Coeckelbergh M. 'AI Ethics' (2020). MIT Press. Massachusetts chapter 9

Apple launched their health monitoring which surprisingly lacked a menstrual period tracker for women[95].

Flickr, a popular website for photos and videos once automatically tagged a portrait of a black man as ape. Another notable facial recognition bias example was camera maker Nikon's new technology that was able to detect when eyes were blinking, the technology however always categorised Asian faces as blinking and would subsequently send out a warning[96]. From the above, we can deduce that the technology developers did not test the technology using a diverse pool, which is ironic since the company headquarters is in Japan. We also learn how the underlying reasons of facial recognitions discrimination is faulty training and sampling data and how this may exclude and disadvantage certain classes of people.

In the case of social media companies, automated content moderators allow less restrictions of harmful content so as to avoid appearing very restrictive because then they run the risk of losing users who may argue about restricting of freedom of expression even when the content has the potential to be damaging or injurious[97]. An algorithm used by twitter, which relies heavily on user categorization, places former US President Donald Trump under the category racist and racism[98]. I use this example to show that algorithms continue to learn collective human judgments and that they may be said to only comply with available community standards which feed their data training.

The use of algorithms to manipulate humans has been seen with Cambridge Analytica where algorithms were used to influence voting patterns[99]. Additionally, most AI systems learn from the human interactions that continuously occur and the machine may develop a bias arising from the user

[95] Perez C.C. 'Invisible Women: Data Bias in a World Designed for Men' (2019). Abrams Press. New York at page 99

[96] Joseph R. 'Are Face Detection Cameras Racist?'(February 2020) http://content.time.com/time/business/article/0,8599,1954643,00.html (accessed 18th May, 2020)

[97] Perez C.C Note 95.

[98] Perett C. 'Trump twitter account appears as the top result when users search the word racist. Twitter says an algorithm is behind that recommendation' (Jun 2020). Available at https://www.businessinsider.in/politics/world/news/trumps-twitter-account-appears-as-the-top-result-when-users-search-the-word-racist-twitter-says-an-algorithm-is-behind-that-recommendation-/articleshow/76239330.cms

[99] Boldyreva E.Lm Grishina N.Y., and Duisembina Y.'Cambridge Analytica: Ethics and Online Manipulation with Decision-making Process' (April 2018). Future Academy available at https://www.researchgate.net/publication/330032180_Cambridge_Analytica_Ethics_And_Online_Manipulation_With_Decision-making_Process

input. Granted, this is usually not intentional on the part of the developers but happens when the machine relies on user input for its continued learning.

The above highlighted negative consequences of AI are not to preach doom, but rather to show that it is possible to design machines that obey the law as well as our human values. This is achievable by automating legal practice. However, this does not imply that machines are capable of being endowed with morality, they simply fulfil the condition of being imparted with legal rules that the society has prescribed[100].

AI systems do not have an independent ability to possess morality since they are produced by human developers and only the developers may act morally through their input in the AI. Subsequently, one could say that AI systems, at best, represent human morality provided that morality can be represented by an AI outside of human involvement. It is my view that morality cannot be simply reduced to rules and algorithms that a machine follows, we do need human emotion to be involved at some level.

It may be further argued that machine obedience to law can be used as a means of preventing harm, what constitutes harm is of course defined by the society. In the EU for instance, it is considered harmful to be refused a job opportunity on the basis of race. It is my opinion that machine obedience may be said to be a yardstick for measuring the morality of machines and how that morality is expressed by the AI system. This does not, however, mean that machines are capable of moral decisions, but rather that they may be agents of society's morality.

2.6 How AI discriminates

AI decisions are a black box in that there is no knowing what criteria were used to arrive at a decision. This is especially harder in instances of trying to determine whether there has been any racial or gender influenced decisions.[101]Automated systems, when not carefully designed, do lead to arbitral judgments and indeed amplify discrimination of some groups, this has a broader social implication and may continue to widen the gap between privileged persons and those who are already disadvantaged[102].

[100] Chinen M. 'Law and Autonomous machines: The co-evolution of legal responsibility and technology' (2019). Edward Elgar Publishing. Northampton. At page 147

[101] Baracas S and Selbst D.A 'Big Data's Disparate Impact' (June 2016). California Law Review. Volume 104. No. 3. pages 671-732

[102]Citron, D. K., & Pasquale, F. 'The Scored Society: Due Process For Automated Predictions' (2014). Washington Law Review. Volume 89. No.1

The research by Baracas and Slebst referred to earlier is based on American anti-discrimination law but the contents and examples are applicable to the EU situation as it concerns the potential of AI to discriminate. The authors note that most discrimination is institutional discrimination; this is discrimination that is often ignored in societies or the society is inclined to suppress the discrimination such that the discrimination is considered 'normal'. This is opposed to intentional discrimination and as such it is more of a challenge to fight[103].

AI sometimes reinforces social inequality through its use of consumer data[104]. An example is given of pricing discrimination where an algorithm predicts that online consumers in rural areas are least likely to frequently shop online as compared to people in urban areas. Subsequently, the AI assigns a higher price for online shoppers in rural areas and coincidentally, people in rural areas often earn less money than their counterparts in urban areas and as such, AI reinforces social inequality[105]. There is subtle discrimination in Google as was proven when the search engine offered jobs with a higher salary to candidates who registered as men and offered lower salaries to search engine users who registered as women[106].

The Dutch government used an algorithm called System Risk Indication (SyRi) secretly for analyzing and detecting fraud by using personal data of its citizens. This algorithm gave some false results which targeted persons in possession of double citizenship. There was also a lack of transparency as its use was improperly explained and even when it was explained, the details of its use were published in a gazette that was not widely read[107]. The algorithm was found in breach of Article 8 of the ECHR as it failed to strike a balance between public interest and the right to private Life[108].

[103] Allan R. and Masters D. 'Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making' (2019). Europäische Rechtsakademie. Page 4 and 5 Available https://doi.org/10.1007/s12027-019-00582-w

[104] Schmitz, A. J. 'Secret Consumer Scores And Segmentations: Separating Haves From "Have-Nots' (2014). Michigan State Law Review, 1411 at page 1443.

[105] Borgesius F.Z.'Discrimination, Artificial Intelligence and Algorithmic Decision Making' (2018). Council of Europe. Directorate General of Democracy. Strasbourg at page 68

[106] Datta A., Tschantz M.C, and Datta A. 'Automated experiments on ad privacy settings' (2015). Proceedings on Privacy Enhancing Technologies 1 , 92–112. Specifically at page 94

[107] Vervloesem K. 'How Dutch activists got an invasive fraud detection algorithm banned' (2020). Available at https://algorithmwatch.org/en/syri-netherlands-algorithm/

[108] Meuwese, A. 'Regulating algorithmic decision-making one case at the time' (2020). Case note on: District Court
of The Hague , 5/02/20, ECLI:NL:RBDHA:2020:865 (NJCM vs the Netherlands (SyRI)). European Review of Digital Administration & Law, 1(1). At pages 2010-211

In Australia, a study was undertaken to find the rates of domestic violence in indigenous and non-indigenous Australian, the results over predicted the rates in indigenous Australians and under predicted the rates of domestic violence in non-indigenous Australians because of a focus on previous offenders in the indigenous communities. In this case, the AI was using data sets that were flawed in that they relied heavily on previous convictions of indigenous Australians[109]. It is clear to see that AI discrimination occurs sometimes because AI may targets users who are may be uninformed of the possible outcomes of them volunteering or agreeing to be subjected to an algorithmic decision making process[110]. Such algorithms may continue the perpetuation of bias in societies in which they are deployed

In the case of Sermide[111], which concerned discrimination in price of agriculture products, it was stated that consumers must not be discriminated against on region unless there exists an objective criteria which bridges the gap between privileged individuals and those who are not privileged. This case clearly reinforced no discrimination on region but leaves it open for tech companies to manipulate prices based on location of users.

Algorithms will further make decisions reinforcing the institutional bias and of note, the discrimination may be the result of provided data sets. Most scholars overlook this, and policy makers sometimes show an unwillingness to anticipate human bias when programming AI systems. It is exceptionally difficult to identity institutional discrimination and therefore to address it.

At any given point during the development phase of AI, the designers might make a series of design decisions that might result in unfavourable treatment for different classes when applied broadly in society. Some design biases might be unintentional, such as when an AI programmer makes a personal choice without realizing that such a choice happens to benefit people like themselves (race, age, gender, political inclination)[112].

The following are the types of discrimination that have been identified by Baracas and Slebst:

[109] Mcnamara D., Graham T., Broad E and Oong C.H. 'Trade-Offs in Algorithmic Risk Assessment: An Australian Domestic Violence Case Study' in Devitt S.K and Mann M. Edits (2019) Good Data. Institute of Network Cultures. Amsterdam at page 101

[110] Bird S., Barocas S., Crawford K., Diaz F. and Wallach H: 'Exploring Or Exploitation. Social and Ethical Implications of Autonoumous Experiments in AI'. Microsoft Research available at https://ssrn.com/abstract=2846909 at paragraph 5

[111] Serminade SpA v Cassa Conguaglio Zucchero and Others C-106/83

[112] Surden H. 'Ethics of AI in Law: Basic Questions. In Oxford Handbook of Ethics In AI' (2020). Edit Dubber M.D., Paquale F., and Das S. Oxford University Press. Oxford. At pages 719-736

2.6.1 Defining the target variable and class label

This is where the algorithm attempts to find a relationship in any given data set. In machine learning, a target variable may be defined as the final output you are trying to predict, and the class label are data sets that are assigned labels. An example is given of a message that is repeatedly flagged (labelled) as spam, the system will categorize as spam any similar future messages. Another example could be that of a data scientist who aims to predict the frequency of cancer in smokers and has the target as active smokers under his experiment and the class label will be the number of smokers who get cancer[113].

A real-world instance occurred at a hospital in the UK. Saint George's hospital in the UK had an AI system for job applicants which was biased against racial minorities and women, taking over a previous pattern of hiring. The hospital, unbeknownst to them, was automating bias.[114]

2.6.2 Data Collection

Data collection as an exercise is not without its errors and some errors, faulty or non-representative data may have disadvantageous implications against some of the people concerned. Data collection that has been conducted from incorrect sources might discriminate against some individuals in some cases. The St George's case illustrates this point rather well, there was inaccurate information obtained about black people as well as women and these were discriminated against because of faulty data.

The meaning of data depends on the context in which the data is used. Depending on who feeds the data sets, AI much like humans, may have biases inherently built in them. Having a large data set and being able to identify patterns within a fraction of a second does not mean that AI systems are right[115].

One other aspect of data collection is that some classes of people are excluded from the data collection. Lehman[116]states that the focus has mainly been on how big data poses a threat to privacy and the threat it causes to equality has been ignored. Marginalized classes that do not have their presence on the internet are liable to be disadvantaged, and some other classes are discriminated against because of big data.

[113] My own example used here to simplify the two terms.

[114] Lowry S and Macpherson G. 'A blot on the profession' (5th March 1988) British Medical Journal. Volume 296. No. 6623

[115]Helbing D. 'Thinking Ahead: Essays on big data, digital revolution and participatory market society' (2015). Springer Publishing. Berlin at page 11

[116] Lehman J.'Big Data and its exclusions' (September 2013). 66 Stanford Law Review Online. Volume 65. No 55

Overrepresentation in any given data set can also lead to discrimination for some classes, an example is given of a group of employees whose work data is constantly monitored and as such, mistakes are logged in at a higher rate as compared to another set of employees whose work data is not overly monitored and hence highly unlikely to be so prejudiced.

Data collection has the added risk of a regulatory nature. This is because data collection in itself presents an innovation risk. Regulatory authority and its approach to data collection require a clear and strict process of data policy and this may stifle the growth of innovation[117]. Discrimination risks have occurred as it relates to facial recognition or targeted advertising. Innovation risks further concern blockages to AI development that may be imposed by regulators or that certain data may be subject to intellectual property protection and thus unusable[118].

2.6.3 Proxies

Algorithmic discrimination may occur where a given set of accurate data also has some other protected class of people who may be disadvantaged. Consider a bank which uses an AI system to detect customers who are likely default on their loans, the data set may use postal code to detect customers likely to default. The problem with the proxy could be that the postal code may have people of a racial minority[119]. Proxy discrimination is very hard to solve and the ironically, the only logical way of eliminating this problem is to make the AI system less accurate[120].

2.6.4 Masking

AI developers could intentionally design an algorithm that is prejudicial and then mask its operability by exploiting proxies, data collection as well as the other technical operations of the system. This intentional discrimination could also exist by the developing refusal to acknowledge mistakes in previous data sets from which potentially discriminatory decisions may continue to be made.

Intentional discrimination has not received widespread attention, this is mostly because policy makers and legal scholars have chosen to focus on errors that may result from a faulty algorithm and pay a

[117] Hacker P. ‘A legal Framework for AI Training Data: From First Principles to the Artificial Intelligence Act’ (March 2020). Law, Innovation and Technology 13 at page 3
[118] Ibid at page 4
[119] Borgesius F.Z. ‘Discrimination, ‘Artificial Intelligence and Algorithmic Decision Making’ (2018). At page 21
[120] Ibid at page 22

blind eye to the idea that some developers may intentionally design their systems to discriminate against a certain class of people[121].

2.6.5 Feature selection

Feature selection concerns using an algorithm that focuses on certain features of any given data class, an example by Lehman[122] concerns an organization that uses features such as schools to select employees. Most employers favour applicants that attend famous universities which universities may be expensive and unaffordable for minorities. An algorithm that uses the university feature may be discriminatory to minority races as a result[123].

Without a doubt, AI systems, particularly algorithms for policing have been very helpful in making the workload of the agencies lighter. However, there comes with it some potential violation of specific human rights[124] such as the right to fair trial and due process. National Security wings in the United States deploy mass screening of internet users and there has been an erosion of the presumption of innocence as well as the right to fair hearing. Predictive policing algorithms consider, inter alia, the type of person to commit certain crimes and in some cases have been prejudicial towards racial and ethnic minorities and this may be considered discriminatory.

Additionally, predictive policing algorithms may also affect one's right to defend themselves.[125] The algorithm may be considered to work with some bias and goes on to make the decision which an officer may have to follow, the officers in most cases have no intention to act in a discriminatory manner but are reliant on the algorithm to enforce decisions[126].

[121] Krieger H.L 'The content of our categories: A cognitive Bias Approach to Discrimination and Equal Employment Opportunity' (1995). Stanford Law Review. Volume 47, available at https://www.researchgate.net/publication/271802177_The_Content_of_Our_Categories_A_Cognitive_Bias_Approach_to_Discrimination_and_Equal_Employment_Opportunity
Note: Krieger was writing about discrimination in the American workplace without making mention of any machine learning but she pointed out that discrimination may be intentional by the policy makers.

[122] Lehman J (2013). Note 116. At page 13

[123] Loc.cit.

[124] European Council Study on Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) And Possible Regulatory Implications (March 2018). Available https://ec.europa.eu/futurium/en/european-ai-alliance/study-human-rights-dimensions-automated-data-processing-techniques-particular.html

[125] Ibid.

[126] Lum K and Isaac W.'To predict and serve' (2016). In detail Magazine available at https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2016.00960.x

Predictive models also work against people in that some algorithms may predict that 80 percent of people in a given postal code are less likely to repay their loans due to the fact that they pay their bills late. Such predictive models clearly disadvantage the remaining 20 percent who are well able to repay the loans and are unfortunately bundled in the same risk group by the AI[127].AI will often continue to adapt its own algorithms and develops future processes based on its own development and this makes it difficult for designers to have any meaningful input over its predictive model[128].

The potential for discrimination holds true and extends even when using no sensitive data categories, from which it is possible to infer sensitive information. In addition, algorithms can encourage discriminations through data set cross-checking and through profiling with no prior identification requirement[129].

Even data that may be considered neutral such as a person's address has the potential to lead to inference and discrimination based on ethnicity, gender, or sexual preference especially when these data sets are linked by the algorithm[130].

2.6.6 Normative Unresponsiveness

This is discrimination by AI that combines training data as well as connecting seemingly unrelated data sets to arrive at a decision. An example of normative unresponsiveness may be how sex is used by AI systems to decide the rate of car insurance. Despite sex as an identifying data being removed, the AI systems are still able to deduce the sex of the applicant using other data sets[131]. Equality of the sexes and non-discrimination on the basis of sex as a determining factor for insurance premiums was reinforced in the EU case of Association Belge Des Consommateurs Test-Achats NBL and Others v Conseil Des Ministres[132].

[127] Borgesius F.Z. Note 119. At pages 36-37

[128] Cliff Huang, "Can A.I. Be Taught to Explain Itself?," The New York Times Magazine (Nov. 21, 2017).

[129] Beranger J.'The Algorithmic Code of Ethics: Ethics at the bedside of the Digital Revolution' (2018). Volume 2. John Wiley and Sons. Hoboken at page 29

[130] Van den Hoven, J. 'Innovations, Legitimacy, Ethics and Democracy' (2007). in IFIP International Federation for Information Processing, Volume 233, The Information Society, eds. P. Goujon, Lavelle, S., Duquenoy, P., Kimppa, K., Laurent, V., (Boston: Springer), pp. 67-72.

[131] Tischbirek A. 'AI and discrimination: Discriminating Against Discriminatory Systems' (2020). In Wischmeyer T. and Rademacher T. Regulating AI. Springer nature at page 109

[132] C-236/09

2.7 AI discrimination in medical devices

The E.U has not effectively developed a standardized approach to regulating AI in traditional medical devices, and consumer health devices receive considerably less attention despite how wearables such as Fitbit, smart watches with medical trackers and mobile applications have become popular for most people[133].

Regulation of AI medical devices in the EU is enforced by the Medical Device Data Directive[134]. Under the directive, there is need for the software functionality of the medical device to be verified[135]and it remains unclear as to how this verification may be done, especially given the black box nature of AI systems.

Under the Medical Device directive, risk management is delegated to the manufactures of the devices and not to the regulatory authority. The argument given in support of assigning risk management to the manufacturers is that the manufactures are best suited to plan and identify potential issues that may arise in new technology, particularly the ones they are designing[136].

AI technologies can process unidentifiable data very accurately to determine personal characteristics of users and in the process identify the users of said technologies. Dietary and exercise information found in AI wearables can also be used to identify the state of a user's health, and to identify that user, even when they have opted to use non identified data. This is because of the large amount of other identifiable information that allows the algorithmic system to identify users.

For example, data from the usage of global positioning systems available in medical or exercise wearables is capable of reviewing, through algorithms, a person's home, a potential infringement to right to private life. Location data may also review a person's place of worship, by keeping a record of the location which may be a church or a mosque and also the days and times that a person visits those places allow the algorithms to deduce a person's religious beliefs[137].

[133]Tschider C.A. 'Deux Ex Machina: Regulating Cybersecurity and AI for patients of the future' (2018). Savannah Law Review. Volume 5, No. 1 at page 179

[134] Regulation (EU) 2017/745

[135]Ibid Articles 27 and 29.

[136]See above TSchider C.A Deux Ex Machina (2018) at pages 198-199

[137]Tschider C. and Kennedy K. 'Data Discrimination: the International regulatory impasse of AI Enabled Medical Wearables' (2020). At page 5. In Legal, Social and Ethical Perspectives on Health & Technology. Ed. Bollon & Berti Suman. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3515727

Usually, the data subjects are offered an option to give consent to the tracking of their data. Most scholars have pointed out that the data controllers, in this case the organizations or companies that manufacture medical devices, only need to offer the option of providing details of how data will be used in a privacy notice which the data subject has to consent to. Consent in most cases is ill informed, and the organizations would have met their legal obligation under the law[138]. This is despite the GDPR stating explicitly that consent to data processing must be given with full understanding[139].

Informed consent from those subjected to AI decisions in nearly impossible to obtain. This is because there are so many complex processes that the AI system undertakes and it is exceedingly difficult to communicate the reasons for a decision to those affected by the decision. The very fast speed at which AI arrives at a decision may be viewed as suspect by most people who may feel that they have not offered consent.

It must be noted that under the GDPR Act, companies are allowed to use personal data for business purposes, including selling that information to third parties, provided that they disclose in their privacy policy how they will use or share that information.

The GDPR Act has also set conditions for how companies must meet consent regulation[140]. It remains that consent for most data subjects in health algorithmic systems is not given with full understanding[141]. This is because AI systems have complicated calculations that blare the line between what is consented to and whether the data subjects have been properly informed of the data being processed[142]. Furthermore, article 22 of the GDPR requires data controllers to offer an option to users of opting out of the company making use of their data to arrive at an automated decision. If a user chooses not to have any automated decisions being made, they may face discrimination related to health service[143].

The GDPR, read in consideration with the Medical Device Directive, has strengthened the position of consumers, without argument. However, there is need for specific address and continued embrace of

[138] Ibid at page 8

[139] GDPR Article 6- Lawfulness of Processing.

[140] GDPR Article 7

[141] Dove E.S. and Chen J. 'Should Consent for data processing be privileged in health research? A comparative legal analysis' (2020). International Data Privacy Law. Volume 10. No. 2 at page 125

[142] Ford R. and Price W. 'Privacy and accountability in black box Medicine' (2016). Michigan Technology Law Review. Volume 23. No. 1 at page 5.

[143] Baracas and Selbst's AI discrimination as a result of data exclusion cited at note 68.

new technology introduced by AI, a feat lacking as values remain the same and technology continues to challenge our enjoyed of values.

Another challenge of AI in health is that it may target disabled minorities who require wearables nearly all the time such as those who wear heart monitors. These are distinguished from people who wear medical devices such as those for exercise. The former has more of their data collected and may be subject to targeted advertising which may border on discrimination. Additionally, people who require compulsory medical wearable often have no option but to consent to all data processing even by automated means[144]. They either consent or have no right to use the wearable. Such consent, as established by the GDPR, is not true consent.

One other ethical consideration of AI employed in medical devices concerns the value system. Values may qualify and define health and often times, privileged individuals or the companies producing medical devices may set different standard for health. Additionally, misdiagnosis risk may also occur if a wearable device is improperly designed and excludes certain groups like wheelchair bound individuals[145].

AI usage in medical devices may further compromise individual freedoms because they violate privacy of decision making because of interference with personal decision making via prediction on the usage of wearable health trackers. Excessive interference occurs when the prediction of the algorithm continues to persist because persuasive technologies also take away personal autonomy[146].

The exclusion of old patient data from the digital era may also work against older patients as they are left out on new and effective medical treatment[147]. Data exclusion continues to permeate in digital health as corporations continue to focus on capitalizing on available data and in the process the

[144] Wilson N. and Kennedy K. ‘The banality of digital aggression: Algorithmic Data Surveillance in Medical Wearables’ (2020). In Digital Ethics: Rhetoric and Responsibility in Online Aggression, Hate Speech, and Harassment. Edited by Reyman J. and Sparby E. Routledge. New York at page 220 and 224

[145] Morley J. Machador C.C, Burr C, Cowls J., Joshi I. Taddeo M. and Floridi L.‘The debate on the Ethics of AI in health care: A reconstruction and Critical Review’ (2019). At page 8

[146] Lanzig M. ‘Strongly Recommended: Revisiting Decisional Privacy to judge Hypernurdging in Self-Tracking technologies’. (May 2018) Philosophy and Law. Springer Publishing at page 2

[147] Floridi L. ‘Soft Ethics and the Governance of the Digital’ (2018). Philosophy and Technology. Volume 31, No. 1 at page 3.

corporation's sidelines individuals who do not own any of their health applications[148]. There is also a wider risk of the health system being restructured differently to reflect a system that may exclude other data subjects

2.8 Conclusions on Legal Framework of AI Regulation and AI Discrimination

This chapter has focused on those important aspects of AI legislation that directly concern fundamental rights and human dignity and in particular the right to non-discrimination. Having established the responsible laws and regulations, the chapter has showed how AI discrimination may occur. Of importance to this study are the factors that are non-technical but still have an ability to cause bias against certain classes of people on the basis of their age, sex, race and indeed residential address amongst some other unclear factors.

Intentional discrimination may occur by the developers and they may mask this discrimination to hide it. It must be emphasized that negligent acts on the part of developers may cause unintentional discrimination as most developers design AI with ethical considerations. The chapter has shown that discrimination most often occurs when there is data exclusion, data over representation, the use of incorrect data or correct data which is incorrectly fed to the AI systems. The takeaway from this observation is that discrimination may be addressed much easier if there is involvement from other stakeholders apart from the developers who may overlook certain things as they tend to focus mostly on the technicalities of AI.

As the paper progresses to discuss moral imagination (ability to imagine all possibilities), it is important to note that most of the ways in which AI discriminates can be proactively prevented by having widespread stakeholder consultation as well as considerations for diversity. Fortunately, the EU has made deliberate steps to include stakeholder involvement in AI as can be seen by the European Union Agency for Fundamental Rights Report. Lacking in the report however, are moral considerations with an explicit and deliberate involvement of potential AI subjects since most of the responsibility is placed on the developers. Considerations on diversity and stakeholder involvements require a comprehensive look at moral imagination and its positive aspects on non-discrimination.

[148] Marelli L., Lievevrouw E. and Van Hoyweghen I. 'Fit for purpose? The GDPR and the governance of European digital health Policy Studies' (February 2020). Available https://doi.org/10.1080/01442872.2020.1724929 at page 5

3.0 MORAL IMAGINATION: A GATEWAY TO NON-DISCRIMINATION IN AI

3.1 An introduction to Moral imagination

Moral imagination is a theory of ethics that derives from the natural law theory. The natural law theories of natural goodness, knowledge of basic good and knowing what is both good and right are all underlying concepts of moral imagination[149].

Moral imagination means envisioning the full range of possibilities in any given situation in order to arrive at an agreeable and ethical conclusion[150]. Moral imaginations stems from the work of Adam Smith's The Theory of Moral sentiments[151] in which Smith postulated that natural empathy ought to be the basis of virtue, that we share empathy, and that we are distressed when people don't necessarily agree with or sympathize with our emotional burdens[152]. Smith argued that there is a shared agreement or understanding of sentiments amongst people as they place themselves into other people situation and ask the question of how they would like to be treated[153].

He further stated that collectively, humans refrain themselves because of the existence of impartial spectators, whom they would want to share their sentiments with and ultimately, humans begin to show genuine concern for others. Human beings, Smith further states, have a natural empathy with each other and do learn what is tolerable within their society[154].

Smith further acknowledged that people in various professions and occupations share a common interest in the cooperation and help that they offer each other. An example is given of how a butcher, a baker and a brewer have more to gain in their business by trading and cooperating amongst each other[155]. It may seem outdated for this day and age but the principle remains the same, that we ought to exercise caution in what we do so as to avoid disadvantaging anyone within the society because ultimately, what benefits others benefits you as an individual.

[149] Haaksonseen K.'Natural Law Theory. Encyclopedia of Ethics' (1992). Edit Becker L.C and Beker C.B Garland. New York. Also available at https://plato.stanford.edu/entries/natural-law-theories/

[150] Coecklebergh M.'Regulation or Responsibility?'(May 2006) Autonomy, Moral Imagination and Engineering Design at page 3

[151] Smith A.'The Theory of Moral Sentiments' (1759). 18th Ed, Millar A and Kincaid A. London

[152] Butler E.'The Condensed Wealth of Nations And The Incredibly Condensed Theory of Moral Sentiments' (2011). Adam Smith Research Trust at page 78

[153] Campbell T.D.'Adam Smith's Theory of Moral Imagination' (1971). George Allen and Unwin Ltd. Page 40

[154] Loc. cit.

[155] Himmelfarb G.'The moral imagination: From Adam Smith to Lionel Trilling' (2012). Rowman & Littlefield Publishers. Plymouth. At page 9

Smith's definition of moral imagination is visionary and far seeing, it may apply to nearly anything that concerns ethics. His analysis forms the basis of modern discussion of moral imagination as the ability to be creative in ethical decision-making including in technology design[156].

Another notable philosophical work on moral imagination is by Hume. Hume's work is relevant in this study because he lists elements that may easily be attributed to moral imagination in AI. These summed up elements are shared sympathy in that human beings care about the emotional state of others and as a result, humans are more willing to understand the situation in which others are in[157].

There are several works on moral imagination which are out of the scope of this study, but I will use the four aspects of moral imagination. Moral imagination has four kinds of imagination which are imaging, problem solving, fantasizing and finally the moral imagination itself[158]. This is a concise and a very helpful guide for developers of AI systems wishing to be ethical in their designs.

The Philosopher Mark Johnson defined moral imagination as considering various possibilities in a challenging situation and envisioning any potential harm in order to arrive at an ethical conclusion[159]. Johnson's definition assumes, in theory at least, that harm will arise if no moral imagination is undertaken, hence the importance of moral imagination in AI to prevent non-discrimination.

3.2 moral imagination in technology law and in AI

"Big data processes codify the past. They do not invent the future. Doing that requires moral imagination, and that's something only humans can provide. We have to explicitly embed better values into our algorithms, creating big data models that follow our ethical lead"[160].

From as far back as 1960, scholars were asking questions of morality and technological advances. Wiener[161] stated that the relationship between humans and machines could be equated to that of a man

[156] Bevan D.J, Wolfe W. R. and Werhane P. H. 'Systems Thinking and Moral Imagination: Rethinking Business Ethics' (2019). Springer International Publishing AG at page 72

[157] Collier M.'Hume's Theory of Moral Imagination' (July 2010). History of Philosophy Quarterly, Volume 27 No. 3 University of Illinois Press at page 255

[158] Kekes J. 'Moral Imagination, Freedom and the Humanities' (1991). American philosophical Quarterly 28. At page 103

[159] Johnson M. 'Morality for Humans: Ethical Understanding from the perspective of cognitive science' (2014). University of Chicago press. Chicago, at page 28

[160] O'Neal C. 'Weapons of Math Destruction: How Big Data Increases Inequality and threatens democracy' (2016). Crown Publishers. New York. At page 169

[161] Wiener N. 'Some Moral and Technical Consequences of Automation' (May 1960). Science Magazine. Volume 131 at page1375

and a slave, wherein the computer is a slave that is expected to be intelligent and to follow instructions without question. The follow up with this relationship is that sometimes, the slave may be much more intelligent than the man. This is problematic for a machine that has become much more intelligent than man. The author goes on to acknowledge that data scientists should be aware that technology will rapidly evolve and must thus, "exert the full strength of their imagination to examine where the full use of our new modalities may lead us"[162].

O Neal elaborated in her book that our algorithms are reflective of the injustices that may already exist in society. She states however, that there still exists an opportunity to fix systematic and institutional bias by allowing a wider and diverse view of potential data subjects by using moral imagination[163].The need for moral imagination in technology maybe the required characteristic of universality in guiding our technological future forward[164]. Technology systems require some emotional input from the developers to prevent the technology system from being corrupt or biased.

This corrupt and biased AI may be partly addressed by applying Hume's considerations of moral imagination. Hume elaborated on other aspects of moral imagination and one concerns sympathizing with strangers with another element being that our care as humans is limited to people within our circles[165]. The two elements expounded by Hume are particularly important for moral imagination as it relates to technology design including AI development. This is so because technology developers are more likely to design their technology in accordance with the values of the society in which they are a part (care being limited to people within their circles). Globalisation, however, ensures that AI technology will extend beyond their society and affect people outside their society (the strangers with whom they must sympathize).

Moral imagination may not be enough for AI, seeing as the advanced of these systems is pushing beyond the traditional expectations of the human mind and moral deskilling occurs in our morality[166]. Moral deskilling refers to the inability to consciously employ oneself to arrive at a moral decision. AI

[162]Ibid at page1358

[163]O'Neal C. (2016) Note 160. At page 81

[164]Moran S., Cropley D. and Kaufman James C. 'The ethics of Creativity' (2014). Palgrave Macmillan. New York. At page 27

[165]Ibid at page 259

[166]Vallor S. 'Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character' (2015). Philosophy of Technology 28: available at: https://link.springer.com/article/10.1007/s13347-014-0156-9

systems take the morality dilemma away from people when the systems have been designed to make decisions[167].

To enhance morality in AI systems, the future of life institute inspired by Isaac Asimov's 1942 law of robotics introduced 23 laws for AI to impart the systems with morality[168]. The Asilomar AI principles include continuance of research issues in AI, ethics, and values that the systems must adhere to as well as an attempt to navigate longer term issues with AI[169].

Risk arising from AI systems may be reduced by imposing external regulation on developers of the technology or by making sure that the technology developers exercise moral imagination in the design of the algorithms so as to make sure there is fairness in the decisions that the system make[170]. Allowing the AI developers and engineers to exercise moral imagination ought to be encouraged as it prevents isolation of the developers from the public for which they design their technology. Classifying developers as part of the society solves half the problem already. Engineers are often isolated from the society whenever negative questions arise because of the technology they develop.

Moral imagination, I opine, would allow for the developers to have a sense of responsibility and this sense of responsibility comes in when we allow regulators (policy makers such as institutions and politicians) only a limited control over the developers[171]. Feelings of responsibility are amplified where there is little external control, and the technology developers are free to work much more independently[172].

Developers should be able to know their own capacity and they do consider personal principles they identify with in designing the algorithms[173]. Notably, it is only possible for developers to act in a

[167] Loc. cit

[168] Hassabis et al, 'Asimolar AI Principles'. Future of life institute. Available at https://futureoflife.org/ai-principles/ (accessed on 14th June, 2020).

[169] Halley G. 'The criminal Liability of Artificial Intelligence Entities'. Available at: http://ssrn.com/abstract=1564096

[170] Coecklebergh (2006) Note 150. At page 10

[171] Atkins K. 'Autonomy and the Subject Character of Experience' (2000). Journal of Applied philosophy. Volume 17. No. 1 At page 74

[172] Loc.Cit.

[173] Michalczak R. 'Animal's Rights Against the Machines' (2017). In Kurki V.A and Pietrykowski T. Legal Personhood: Animals, AI and the Unborn. Law and Philosophy Library. Volume 119. Springer Publishing. At page 99

morally responsible way when they have been given the responsibility to design technology whose functioning is agreeable in the society[174].

Moral imagination in AI design may be amplified by having ethics as part of professional training. This fortunately, is a matter of practice for most computer scientists. Ethics in this instance becomes an integral part of what it means to develop AI systems. Algorithmic ethics may be a considered tool for amplifying ethics in AI. Algorithmic ethics refers to ethics whose perimeter of action concerns the digitization of society. These ethics should be present from a new technology's creation and should evolve over time, taking account existing ethical rules[175].

Furthermore, there should be considerations of several factors of an algorithm such as its margin of error as well as the rate of trust as well as its complexity which developers of AI should bear in mind. The problems of bias in algorithms continue to exist and they do create social discrimination and exclusion and as such developers should not develop algorithms without consideration of AI's societal impact[176]. Put simply, elimination of bias should be a deliberate ethical goal in development of AI[177]. Added to this is the importance of knowing who is involved in developing and implementing codes of ethics for the AI, as well as where responsibility for projects to embed ethical decisions into machines lies[178].

The answer to responsibility and implementation of codes of ethics in AI may lie with the geopolitically powerful countries judging from the GPDR whose data privacy laws apply outside of the EU when EU citizens are concerned[179]. Furthermore, the global nature of in AI requires that AI regulation and its ethics remain global.

[174] Loc. cit.

[175] Beranger J. (2018) 'The Algorithmic Code of Ethics: Ethics at the bedside of the Digital Revolution' (2018). Volume 2. John Wiley and Sons. Hoboken. At page 72

[176] Loc.cit.

[177] Van Den Hoven J. and Weckert J. 'Information Technology and moral philosophy' (2008). Cambridge University Press

[178] Bodding P.'Towards a Code of Ethics for Artificial intelligence' (2017). Edt O'Sullivan B and Wooldridge M. Springer International. At page 16

[179] Daly A, Hasndorff T, Hui L, Mann M, Marda V, Wagner B, Wang W and Watteborn S. 'Artificial intelligence governance and Ethics: Global Perspectives' (2019) . Chinese University of Hong Kong. Research Paper No 2019-15 at page 6.

One may wonder, if AI is expected to become superior at certain tasks, whether the AI will also be better than humans at making ethical decisions[180]. The answer is that there is no standard of knowing whether an automated system is moral, as morality of an AI system is contextual, and context is subjective[181]. Ethics in AI has the ability to impact on individual lives, and also has the power to change society and life as we know it[182]. AI is further changing the social structure of communities and our societies will continue to be influenced by AI[183]. The continued development and widespread use of AI means we have to ensure that the technology continues to comply with EU values.

Finally, development of AI ethics therefore requires that the AI is designed to obey preset values as well as to actively avoid discrimination. However, the functioning of the AI remains complicated and it must be noted that it is difficult for other developers to understand a different AI system and it is nearly impossible for a non-technical person to understand AI[184].

This difficulty may be minimized by using value sensitive design to monitor moral imagination as well as to impart EU values into the design and deployment of AI.

3.3 Conclusions on Moral Imagination as a Gateway to Non-Discrimination in AI

The historical aspects of moral imagination have been discussed in a summarized format with the focus being on the relevant developments in as far as they relate to human dignity, the precursor to non-discrimination. Non-discrimination has been enshrined in the EU laws and breach of the same does not only disadvantage some data subjects, it also calls for action by the respective authorities. Understanding that the principle of non-discrimination is as a result of moral imagination which seeks equality for all helps us understand the required prism with which to view AI values and their compliance to non-discrimination.

[180] Davis J.P 'Legal Dualism, Legal Ethics and Fidelity to Law' (2016). Journal of the Professional Lawyer. University of San Francisco Law Research Paper No. 20 at Page 24

[181] Wachter S. Mittlestadt B. and Russel C. 'Why Fairness Cannot be Automated: Bridging the gap Between EU Non-Discrimination Law and AI' (March 2020). Draft paper artificial intelligence discrimination may have patterns of bias unknown to human beings because of the scale of data they process as well as their ability to find connecting patterns. At page 64

[182] Helbing D. 'Thinking Ahead: Essays on big data, digital revolution, and participatory market society' (2015) Springer Publishing. Berlin at page 11

[183] Coeckelbergh M. 'AI Ethics' (2020). MIT Press. Massachusetts. Chapter 1

[184] Beranger J. (2003) Note 175. At page 142

Furthermore, moral imagination continues to inform our technology development to date. An example of newer frameworks incorporating moral imagination is privacy by design which incorporates values such as right to privacy. Developers are expected to include privacy in their products, catering to the needs of consumers who desire the human value of privacy. Developers do design their products proactively by using moral imagination.

One significant factor to note is that moral imagination is a very theoretical concept that may not be easily applied to technology development. It is near impossible to apply moral imagination to AI. This is made possible, however, by using the theories of moral imagination in a value sensitive design evaluation of AI. Value sensitive design tackles the theoretical aspect of moral imagination by deploying measureable steps to the study of AI values, measuring the AI against values of non-discrimination. A closer look at how to incorporate moral imagination in AI by using value sensitive design is provided in the next chapter.

4.0 VALUE SENSITIVE DESIGN: A FRAMEWORK FOR STUDYING AI DEVELOPMENT IN THE EU

4.1 Value sensitive design introduction

There is an intimate relationship between human centered AI and moral values. AI decisions nowadays have a profound effect on the moral standing of concerned data subjects. Most rights and freedoms such as the right to work, right to privacy as well as freedom of information and movement are being subjected to AI processing at one point or another. If there is a lack of upholding EU values including the right to non-discrimination during the process, it becomes clear that data subjects' rights and freedoms are affected negatively.

Unequal distribution, access and availability of AI may enlarge inequality by excluding certain classes from data sets and this has the consequence of denying services to these classes. A study on the future of AI states that public policy must be formulated in the present time to preserve human dignity and other human freedoms and rights[185].

Value sensitive design provides a well-rounded approach to dealing with the challenges that arise from data exclusion as well as major problems that occur in the developmental phase of AI by providing a proactive solution to prevent common problems from arising[186].

Value sensitive design is regarded as one of the best ethical approaches in technological innovation[187]. Value sensitive design may be applied to AI regulation and review of AI performance in as far as it relates to evaluation of compliance with EU values.

Value sensitive design draws from the works of Batya Friedman who argues that moral values may not be universal and that it is especially difficult to account for human values in technology design[188]. In simpler words, technological systems are either intentionally or are unintentionally informed by the moral values of their developers[189].

[185] One Hundred Year Study on Artificial Intelligence (AI100)," Stanford University, accessed August 1, 2016, https://ai100.stanford.edu. At pages 42-44

[186] Hoven J. 'ICT and Value Sensitive Design' (2007). IFIP international Federation for Information Processing. Volume 233. At page 67.

[187] Umbrello S. and De Bellis A.F. 'A value Sensitive Design Approach to Intelligent Agents' (2018). In Yampolskiy R. V Edt (2019) AI Safety and Security. CRC Press, Boca Raton. At page 3.

[188] Friedman B. 'Value sensitive design' (November 1996). Interactions 3 at page 17

[189] Monasso, T. 'I don't know what I'm doing', (2006) Mekelessay. TU Delft.

Friedman states that there are two issues that concern values in technology, these being user autonomy and freedom from bias. My focus is on the latter, freedom from bias as it relates to AI systems and their algorithmic results in relation to non-discrimination.

AI systems may be said to be discriminatory if they assign results based on grounds that may be unreasonable or indeed unexplainable[190]. Discrimination in AI technology may be threefold, pre-existing discrimination that already exists in society and the developers simply engineer systems that carry the discrimination[191]. The second and most discussed in AI is technical discrimination where the technical operations of the algorithm may be problematic, and the third discrimination is an emergent discrimination. Emergent discrimination refers to a discriminatory bias that only becomes evident after the technology has been made available to the public. This could be because of a shift in cultural values or societal knowledge[192].

Technology designers including AI developers have influence on the technology they develop but it remains uncertain as to whether they can influence the use of the technology[193]. Furthermore, technology developers cannot predict the future because of complexity in the deployment of the technology[194].Developers must require stakeholder input in the design phase of AI, at least when the technicalities of AI design are excluded and only the end result is considered[195].

I remain doubtful whether this addresses the challenge posed by AI's opacity and a result may not necessarily satisfy explanation especially when the determining factors remain unclear. This doubt therefore calls for stakeholder involvement and application of value sensitive design to ensure some semblance of control over technology at the very least, and control over the ongoing usage of the technology at the very best.

[190] Morley J., Floridi L., Kinsey L., and Eihalal A. 'From What to How: An overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices' (December 2019). Science and Engineering Ethics. At page 13

[191] See also Note 114, St George's Hospital case. 'A blot on the Profession' (1998). Wherein bias was automated.

[192] ibid

[193] Albrechtslund, A. 'Ethics and technology design' (2007). Ethics and Information Technology 9 (1): 63-72.

[194] Bimber, B. 'Three faces of technological determinism. In Does technology drive history?'(1994) MIT Press. London. At Page 83

[195] Kyung Lee et al 'WeBuildAI: Participatory Framework for Algorithmic Governance' (November 2019). ACM Human Computer Interactions Volume 3. CSCW, Article181. At pages 28-29

A practical application of value sensitive design is in the hiring of new employees. This can be observed in the case of Ricci[196] which concerned the use of examination results to promote city workers within the council of New Haven in the United States. The majority of those who passed the test happened to be white employees and the city refused to promote them as this would be counterproductive in the city's quest for inclusiveness of minorities.

Of interest to this paper is the company that was contracted to provide the examination. IOS, the company in question had employed value sensitive design in its exam preparation as it included all relevant stakeholders and indeed had minority representation as well. The Supreme Court upheld the lower court ruling, taking into account the inclusive design of the examination and ruling that there could be no discrimination when the exam was a reflection of fairness.

The case establishes, inter alia, that value sensitive design is an approach that ought to be applied as it encourages hiring based on qualifications and tends to overlook aspects such as race or gender when applied in the hiring process.

4.2 Suitability of value sensitive design in reviewing AI development

Value sensitive design as a methodology takes place at all development phases of the technology as opposed to simply being a onetime review of developing technology[197].Note must be taken that value sensitive design relies heavily on moral law theory in the development of technology. It is the modern application of moral law to technological development with a heavy focus on human values such as non-discrimination being incorporated into the technology[198].

First of all, to use value sensitive design as a framework for AI review, the proverbial black box that is AI has to be opened. Some of the advantageous aspects of attempting to open the black box of AI have been codified into three main issues. The first is that explanation is inherently good and necessary for the respect of autonomy and furthering the protection and enhancement of fundamental rights. Secondly, an explanation of the workings of the AI system is an enabling action which allows an individual affected by an AI decision to learn how they might challenge that decision and perhaps achieve a different result and whether it is possible do so.

[196] Ricci vs. DeStefano No. 07-14-1428

[197] Friedman B. (November 1996). Note 188. At Page 21

[198] Friedman B, Kahn P.H., Borning A and Huldtgren A. 'Value sensitive design and information systems. Early Engagement and New Technologies: Opening up the laboratory' (2013). Edit Doorn N. Schuubiers D., and Gorman M.E. At page 59

Thirdly, and important for value sensitive design, is that the explanation may be used as a basis for evaluation of the AI system. Using explanation as an evaluation criterion allows policy makers and others to evaluate the compliance of the AI design with EU values.

This approach provides criteria for analysing the basis of decision-making and it forces the basis of decision-making into the public sphere, providing a way to question the validity of the decisions made[199].

Developing AI without consideration of diversity by using a framework such as value sensitive design may inevitably replicate existing societal bias. An example of lack of diversity is when the first air bags where designed and tested only on male subjects in the early 1970s. The result was that airbags excluded women who are usually smaller than men and thus were harmed more by the deployment of the airbags[200].

Writing on nano technology, Timmermans et al[201], states that the field of nanotechnology is complex and dynamic and thus the suitability of value sensitive design as it reduces the complexities by narrowing them to clearly defined targets and goals.

It is the my opinion that AI presents the same or more challenges in terms of complexity as compared to nanotechnology and that the application of value sensitive design to AI development is adequate for inclusion of moral and value input into the AI system. This is because value sensitive design places emphasis on stakeholder involvement in the design process. Additionally, the fact that value sensitive design calls for openness means it can be easily aligned under the empirical framework within value sensitive design.

Writing concerning robots, the authors argued that robot design and production ought to comply with criteria and requirements set by authorities with stakeholder involvement including the designers of the robots as well as consideration of the end users. This ensures that the robots not only comply with societal values but also enhance those values[202].

[199] Selbst S. and Baracas A. (2018). Note 68. At pages 1118-1126

[200] ESPC (2018) The Age of Artificial intelligence: Towards a European Strategy for Human-Centric Machines. European Political Strategy Centre. Issue 29. At page 14.

[201] Timmermans J. Zhao Yingshuan and Van de Hoven J.'Ethics and Nanopharmacy. Value Sensitive design of New Drugs' (2011). Nano ethics Volume 5. At page 279

[202] Leenes R. and Lucivero F. 'Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design' (2014). Law, Innovation and Technology. At page 6

Similarly and as is the case with AI, value sensitive design offers a framework that is designed to address societal concerns and addition of values into design as well as how those values can be reviewed along the development and publication chain[203]. Value sensitive design also allows for concerns to be represented as values, which values can be assimilated into the design of technology and it also allows for stakeholders to monitor the implication of the values.

Value sensitive design may be the tool needed to assist both developers and end users in the development of fair and ethical AI. Technical problems with AI may continue to still exist but there will presumably be wider acceptance of AI and its usage if people trust the systems from its development phrase.

Technology is a result of human imagination, and human beings use their imagination to address challenges through technology, as well as to enhance the enjoyment of their lives. It is important that the moral imagination of the designers considers universal values that are ethical and fair to the human society. Some of the values of value sensitive design include the welfare of the technology users, the level of privacy and freedom from bias that the system provides. Universal usability is becoming standard as a value since these technologies mostly have global usage[204]. Autonomy, trust, and informed consent[205] as well as accountability of the technology for actions arising out of their use is also a value. Environmental sustainability is also a common design value as efforts to curb global warming are increasingly being deployed across various sectors.

People trust people, and not the technology. They trust that the AI technology developed will have the same shared value as them. People are willing to trust that AI will make decisions without prejudice and in some instances, make them better than people might. Discussing interactions on the internet Friedman et al stated that there is multifaceted blend of human actors and technological systems. The question then is who users should trust amongst the many players involved, inter alia, the system,

[203] Cowl J. King T.C, Taddeo M. and Floridi L. 'Designing Artificial intelligence for Social Good: Seven Essential Factors' (2019). Alan Turing institute. London available at https://philpapers.org/archive/COWDAF.pdf

[204] Consider the ability of AI to be global in nature and the difficulty it may present.

[205] Consent with full understanding see Wachter S. Normative Challenges of identification in the internet of things: Privacy, Profiling, discrimination and the GDPR (2018). Elsevier Ltd at page15 and also Kleinig J. The Nature of consent (2010). In Miller F and Wertherimer A. Edit, The Ethics of Consent theory and practice. Oxford University Press. New York at page 19.

computer, and service providers or indeed the people who developed the technology[206]. Ultimately, the authors conclude that people trust people and not the technology that people design. Users hold faith, albeit unconsciously, in the developers of the technology, trusting them to design technology that enhances the users experience. In return, the implication of trusting the developers is that there is an expectation that the developers of technology will adhere to good societal expectation[207].

The list of values is not exhaustive and depends, in part, on regional and national laws as well as objectives of the company developing the technology. It appears that the user has the least input in the design of the technology and can only demand for changes once the technology is already released in the public.Therefore, it is expected that questions may be asked concerning the application of value sensitive design and one of the most obvious ones would be who gets to decide the good values that are to be incorporated and how are they to be measured[208]. Thankfully, the EU has already established basic and additional human values in the forms of Human Rights and Fundamental Freedoms.

As a desired goal, value sensitive design functions on the assumption that technologies are reflective of the society they are developed in and should thus involve the stakeholders (developers, users and intermediaries) in the development process to ensure that the technology complies with the social standards. Above all, there ought to be a proactive approach to the design of technology, to consider all possibilities before releasing the product to the public.

Value sensitive design can be enhanced by procedural ethics and the deliberate approach to the values and the welfare of end users. Value sensitive design in its basic form may prove inadequate for AI since the technology's ability to learn and simultaneously evolve may prove too much of a challenge for traditional value sensitive design approach. I would argue that there is need to review value sensitive design for AI by adding value construction, expanding the scope of value sensitivity by adding another framework onto the third that value sensitive design advocates and lastly by borrowing from the theory of technological framework.

[206]Friedman B., Kahn P., and Daniel C. H (December 2000) Volume. 43, No. 12 Communications Of The ACM at page 36

[207] Nissenbaum, H. 'Accountability in a computerized society' (1996). Science and Engineering Ethics. Volume 2. 25–42. Particularly at page 37

[208]Cenci A. and Cowthorne D. 'Refining value sensitive design: A (Capability-based) Procedural Ethics Approach to Technological Design for Wellbeing' (May 2020). Science and Engineering Ethics Journal. Springer Publishing at Page 5

Value construction in AI design involves having a genuine expression and appreciation of different stakeholders' values and acting in a socially accepted manner[209]. The methodological, procedural, and empirical implication for value construction in AI are that they not only enhance wellbeing and agency but self-determination as well. This allows for the accommodation of diversity and the desired ethical values and ideas can also be suitably translated into the design requirement once the values have been constructed[210].

Expanding value sensitive design to AI will involve the addition of new tools to value sensitive design. Such tools include moral sandboxing and moral prototyping[211]. Cencil and Cowthorne state that emerging technologies present ontological uncertainty, that is, uncertainty that cannot be foreseen even with full information and enough overview. They propose the addition of a fourth investigative concept in value sensitive design, reflexivity. Reflexivity involves a reflective inquiry into how current applications help realize important values or how to raise new value issues or if there is need to raise the new value issues[212].

Attached to the investigative concept of reflexivity is a new method for supporting it. Some authors further advocate for moral sandboxing as one such method. Moral Sandboxing is defined as a mechanism used to identify values and their influence in a new technology within a controlled environment[213]. In practice, this would simply mean involving different design teams to test the functionality of a system and how it enhances human values before the technology is made available for public use[214].

Incorporating the concept of Technological Frames to value sensitive design may, at the very least, help to measure the ever-growing intelligence of AI systems and how and if they enhance our collective human values. Technological Frames refer to examination of people's interpretation of a

[209] Ibid at page 22

[210] Loc. cit.

[211] Reuvar M., Wynberghe A., Janssen M. and Poel I (2020). 'Digital Platforms and Responsible Innovation: Expanding Value Sensitive Design to Overcome Ontological Uncertainty'. Ethics and Information technology 22. At page 5

[212] Loc. cit.

[213] Reuvar et. al (2020). At page 6

[214] Consider the Nikon Digital Camera which was released to the public and was biased against Asian faces. Cited at note 96

particular technology and it concerns the assumptions, expectations and knowledge that people use to understand technology[215].

Technological Frames may be used in AI and its value sensitive design to support the identification of values in the AI system and its operation. There are three domains of Technological Frames and they are well suited when used in combination with standard value sensitive design and reflexivity. The three domains of Technological Frames are nature of the technology, technology strategy and technology in use. The first refers to the type of technology at hand, the second refers to the main reason why the technology exists and the third concerns the proper or improper usage of the technology[216]. All these three domains are important when analyzing a fast-evolving technology like AI. Technological Frame as an analytical tool prioritizes the investigation from the stakeholder's point of view, and this is very important because it provides an opportunity to enhance our understanding of the human values in a given technology such as AI[217].

As a side note, the approach used in privacy by design is a very good example for installing values in the AI system from inception[218]. Privacy by design is related to privacy in design which argues for raising an awareness of values and norms before beginning the technology development[219]. A value by design for AI has been recognised by the EU as being an approach that should be used in the design of AI systems. Values for Design for AI approach was presented in a report by the Finnish Ministry of economic affairs in their submission to the EU[220].

Law, unfortunately, has not been designed for easy accessibility to non-lawyers and privacy by design, as the name suggests, places the responsibility of compliance to regulatory requirements and

[215]Grunloh C. 'Using Technological Frames as an Analytical Tool in Value Sensitive Design' (June 2018). Springer Publishing. At page 3

[216]ibid

[217]Sellen, A., Rogers, Y., Harper, R. H. R., & Rodden, T. 'Reflecting Human Values in the Digital Age' (2009). Communications of the ACM. Volume 52. No. 3. At page 64.

[218]Privacy by design advocates that privacy has to be designed into the system from the development phase. See Rubistein I and Good N. (August 2013) Privacy by Design; A counterfactual Analysis of Google and Facebook Privacy Incidents. 28 Berkely Technology Law Journal at page 1336. See also Article 25 of the GDPR which codifies privacy by designer, putting the onus on data controllers (developers of AI included) to ensure considerations of privacy in the development phase.

[219] D' Acquisto G., Domingo-Ferrer J., Kikiras P., Torra V, De Montjoye Y. And Bourka A 'Privacy by Design In Big Data: An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (December 2015). EU Agency for Network and Information Security. At page 21

[220] Available at http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf (accessed on 15th June 2020)

regulations to the developers. There is a need to have proper guidelines that developers can understand on their own terms[221].

Article 25[222] of the 2016 GDPR simply places the onus on the data controllers by requiring them to have safeguards during data processing by having measures that inter alia, outline the nature of data processing as well as how that processing may affect rights and freedoms of the data subjects[223].

Developers of AI system have control over the functioning of the algorithmic system. It follows therefore, that the developers have an influence on the future functioning of the system, and it is very important to ensure that the developers' input into the system complies with established values, particularly those of the EU, as regards principles of non-discrimination[224].

The main idea reasoning behind placing a burden on data controllers and developers of AI is to ensure responsibility in the development of the technology, primarily in the developmental stages. It may further be argued that placing the responsibility on the developers provides them with an opportunity to act creatively in the absence of strict legislative requirements[225].

There can be no assurance that AI developers will strictly adhere to policy regulations if it harms the profit of their companies but they can, however, use fairness as a standard for developing ethical algorithms[226]. It must be noted that the designers of AI systems may incorporate their commercial interests in their algorithms as well as any bias that may come with it unconsciously.

There remains a need to take human values into the design stages of AI technology because it helps to eliminate societal problems which would become much more difficult to solve when later discovered. Considering Human values at inception of AI development has the dual effect of negating bias that

[221] Urquhart L.D, 'Towards User Centric Regulation: Exploring the Interface between IT Law and HCI' (University of Nottingham/ PhD Thesis 2016).
[222]Article 4(2), GDPR, 2016
[223] Messina D.'Online Platforms, Profiling and AI: New Challenges for the GDPR and in Particular, for the Informed and Unambigous Data Subjects' Consent'(2019). Saggi Medai Laws. 159-173
[224]Flanagan M., Howe D and Nissenbaum H.'Embodying Values in Technology: Theory and Practice: In Information Technolgy and Moral Philosophy'(2008). Cambridge University Press at page 329
[225]Urquhart L.D 'White Noise from White Goods? Privacy by Design for Ambient Domestic Computing. In Future Law: Emerging Technology Regulation and Ethics' (2020). Edts Edward L, Schafer B and Harbinja E. Edinburgh University Press. Edinburgh. At page 65
[226]Drew C. 'Design for data ethics: using service design approaches to operationalize ethical principles on four projects'(2018).Phil.Trans.R.Soc.A376

arises in the future as well as placing the responsibility for any future mishaps on the developers of the technology[227].

The intent of the law should be to regulate human producers of AI through liability laws for any wrongful actions of the AI system. Additionally, there ought to be a responsibility placed on the designers of the AI systems that prevents the system from being manipulated to avoid them carrying out unlawful actions. Lastly, Pagallo states that the AI systems should be embedded with community values[228]. I am of the view that the first two requirements as propounded by Pagallo may be too much of a requirement for AI designers. However, I am in agreement that the human values must be embedded in the AI from the onset, this does require the burden to fall on developers of the technology, and ultimately prevents many future problems that the AI systems may have.

The current drawbacks of EU legal regulation such as not adequately providing to cater for transparency for AI decisions, the complexities of AI regulation as well as the same legislation being spread out over different Acts should be taken note of. However, as has been highlighted in the research several times, the inadequacy of the law to govern the decision-making process of AI does not suggest that we have to embrace the claims of popular authors such as Yoah Noah who famously claims in his book Homo Deus that AI technology is finite and so powerful that it cannot be deterred by legal means[229]. Rather, we must take steps today, to ensure that AI is enhancing our rights and is not acting detrimental towards them. This is done by prioritising human values in the design process and placing the responsibility on the developers. This has the double effect of ensuring further research in AI (for the general good of the public) and giving developers the freedom to act[230].

However, it is not always possible to prophetically predict effects of new technologies when they are being developed, and so innovating responsibly is not limited to including ethics in the design but will also require considering the opinions and interests of various stakeholders[231]. This may be helpful when challenges arise in that the fall back is always the values that have been identified and agreed upon.

[227] Coeckelbergh M. ‘AI Ethics’ (2020). MIT Press. Massachusetts chapter 9.

[228] Pagallo U. ‘Legalize: Tackling The Normative Challenges of AI and Robotics through the Secondary Rules of law’ (2017). In Corrales M., Fenwick M. And Forgo N. New Technology, Big Data and the Law. Springer Publishing. Singapore. At page 284

[229] Harari N.Y., ‘Homo Deus: A Brief History of Tomorrow’ (2016). Signal Publishing. At page 317.

[230] Pagallo U, note 228. At page 286

[231] Flanagan M., Howe D.C. and Nissenbaum H. ‘Embodying values in technology: Theory and practice’ (2008). In: Van Den Hoven J and Weckert J. (eds) Information Technology and Moral Philosophy. Cambridge: Cambridge University Press. 322–353

Consequently, there is need for a bottom up approach starting with the developers and this is done by paying more attention to research work on the subject matter as well as to developers of AI with the inclusion of those protected classes who may be disadvantaged by AI[232].

4.3 Considerations in Value sensitive design

There are three investigative frameworks to consider when applying value sensitive design and these are: conceptual, empirical, and technical.

4.3. 1 Conceptual approach

The conceptual approach is more theoretical than the other two and involves the designers of technology incorporating human values during the design phase of the technology[233]. The question of what human values are is left to the developers and they are supposed to analyse the relevant scope of the values as well as their limitation.

The conceptual approach is faulted perhaps on account of it being theoretical and likely therefore to be misunderstood or ignored. Conceptual investigations may also leave some questions unanswered as they may be considered unnecessary, and even when asked, some questions present half-truths. An example is given of consent when agreeing to cookies, the consent is, in most cases, uninformed[234].

4.3.2 Empirical Investigation

Under empirical investigation, the developers are tasked with generating data from other stakeholders to see whether the said human values are reflected in the technology and if so, how easy they are to spot and how identifiable they are. The investigation may be by way of surveys or requests for reviews of products that are yet to be released (beta testing for software as an example).

Variables of interest in empirical investigation, like the name suggests should be quantifiable in statistics and include things like describable data patterns as well as accurately isolating user needs and cares. In other words, the empirical investigation is concerned with measuring human activities that may be documented.

[232] Aizenberg A and Hoven J., ‘Designing for Human rights in AI’ (August 2019). Sage Publication. At page 15

[233] Friedman B. ‘Value sensitive design’ (1997). Human Values and the Design of Computer Technology. Edited by Batya Friedman. CSLI Publications. At page 3

[234] Friedman B., Kahn P.E and Borning A. (2006) in Galleta D and Zhang P. Edt. Human-Computer Interaction and Management Information System: Application. M.E Sharpe. New York at page 111

4.3.3 Technical investigations

Technical investigation looks at the technology itself and how the functionality affects the human values that are important to the developers and other stakeholders. An example I would give would be analysing data usage of users and how that may infringe on their human right of privacy. The human right of privacy in this instance is a human value that the technology must accommodate and enhance.

Technical investigations are twofold with the first aspect focusing on how the technology itself supports or hinder human values and the second form focuses on how proactive the design is. The second aspect tries to answer whether there is consideration of how the system design supports the values identified in the conceptual investigation stage[235].

AI development may be fraught with some negative consequences and the onus to ensure that technological advancements in AI comply with community values falls squarely on the developers[236].

The strength of value sensitive design is found in the combination of the three frameworks and it is advisable to add other measures as it relates to AI. These could be taking into consideration the ethical practices of the developers and users of the system as well as the rules the systems are designed with and lastly the learning abilities of the systems and the subsequent decisions they make as a result of that learning process[237].

The best example of the deployment of value sensitive design that took into account all three concept, is the construction of campus under Aalto University in Finland. The architects when building the campus considered conceptual framework of incorporating ecologically friendly designs and economic cost as values. They then used empirical investigation by expert interviews and finally deployed the technical investigation by building the campus[238].

[235] Friedman B, Kahn P.H., and Borning A.'Values sensitive design and information systems'(2013). University of Washington. At page 4

[236] Umbrello S. and De Bellis A.F. Note 187. At page 395

[237] Abney K. 'Robotics, Ethical Theory and Metaethics: A guide for the Perplexed' (2014). In Robot Ethics: Ethical and Social Implications of Robotics. Edit Lin P., Abney K. and Bekey. G. A. MIT Press at page 39.

[238] Friedman B. and Hendry D.G 'Value Sensitive Design: Shaping Technology with Moral Imagination' (2019). Massachusetts Institute of Technology. At page 2.

Whilst this example does not involve AI, it is important to use examples close to home where value sensitive design has been used to preserve the values of the community in which the campus was constructed.

Additionally, there are four dimensions of responsible innovation which may be applied to value sensitive design for AI. The commitments to be followed are that the designers, in line with EU values, must have an anticipatory mind-set when design the technology. This means includes analysing any potential impacts be they environmental or that may indeed be potentially discriminatory. The second considerable is that the designers ought to be reflective of the risk that have been faced before whilst the third calls for deliberative approach that includes listening to the public about what they desire as well as providing audience to intended users.

The fourth consideration is that of being responsive in this instance the designers allow a participatory and open process of learning to be adapted, with room for change so as to respond to potential changes in law or indeed societal demands[239].

4.4 Value sensitive design for AI

It has been said that only 20 thousand people in the world could design AI systems[240]. However, AI systems and their applications reach millions more people who may be disadvantaged if the designers fail to take into account certain attributes when designing the said system. Thus, it is important to rethink the process of AI development to include various stakeholders by applying a value sensitive design[241].

Value sensitive design is based on the assumption that technology cannot be value neutral but is reflective of the human values of the stakeholders involved, primarily the designers of the technology and the responsibility for bad AI systems begins with the programmer[242]. Of course, there are also end users who may use such systems wrongly and there are other third parties like governmental regulators who may equally use technologies wrongfully. One focus of this research paper is on how unfair

[239] Owen R., Stilgoe J.,Gorman M., Fisher E. and Guston D. 'Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society II'(2013) : A Framework for Responsible Innovation. at page 38

[240] Mittlesdat B. 'Principles alone cannot guarantee ethical AI' (November 2019) cited at note 66.

[241] Gazzane L., Padovavano A. and Umbrello S. 'Designing Smart operator 4.0 for Human Values: A value sensitive Design Approach' (2020). International conference on industry 4.0 and smart manufacturing, published by Elsevier B.V at page 220

[242] Friedman B. See note 188.

decisions by algorithms can be reduced at the programming stage by ensuring a grounded approach in human values.

It remains that very little is said about human values in the designing of technology and it may be argued that the idea of what constitutes correct values is problematic seeing as morality is plural and what it may also represent is not easy to quantify, at least in science as it relates to the application of the law[243].

There are several human values that users of technology consider to be important and the list is not exhaustive and as it relates to AI and it includes privacy, non-bias, trust and informed consent. A review of some of these values can only be undertaken once the technology has been availed to the public.

Value Sensitive design's pro-activeness and anticipatory framework incorporates stakeholder involvement from the early development stage allows for continuous improvement of the technology throughout the design process and can also adapt for new values which may arise[244]. This makes value sensitive design a suitable framework for AI.

Value sensitive design is sometimes neglected by software developers and this neglect may lead to potential failure of the technology being accepted by society. There is a special need to preserve modern human values such as the right to privacy of data, security and indeed a very measurable degree of autonomy. Whilst the authors here were writing concerning mobile application, the need for value sensitive design extends to the development of AI to ensure the preservation, protection as well as enhancement of the said values[245].

Value sensitive design also considers the values of the developers as they design the AI systems, hoping to install community values into the technology being developed. It is important that the said human values have a universal approval because of the global nature of AI systems. Universal values,

[243] Friedman B. and D. G. Hendry (2019). Note 238. At page 24

[244] Gazzane L. et al (2019). At page 226

[245] Barn B. S and Barn R.'Human and value sensitive Aspects of Mobile App Design. A Foucauldian Perspective' (November 2018). Springer International Publishing AG At page 2

admittedly, may be difficult to gauge but there are similar core values that are shared universally, and these may be taken into consideration when designing AI systems[246].

The need for technology that accounts for human values has been the main contention of the previous chapter and, and this accountability should be reflected in an ordinary language that an ordinary person should be able to understand[247].Granted, this may not easily apply in understanding AI as even specialist may struggle to understand certain aspects of AI functioning[248].

Computers and AI systems may be said to be moral agents, whether intentional or not[249]. The question to ponder is whether such a system can be held morally liable for the decisions it makes[250]. The answer, states Friedman, is that AI systems may be said to be moral agents in that they take intentional actions to perform a function[251].

There ought to be promotion of human values in AI design as well as that AI system having the ability to benefit society[252]. This promotion of human values by AI system means that the system incorporates human flourishing, access to the said technology and the aptness to be useful to the society. Human flourishing is considered as the use of the system that enhances the societal values as well as cultural beliefs, in other words, the AI system should ease humanity in all its endeavours[253].

Access to the technology as a human value means that the technology must be accessible to the majority of people seeing as lack to the said technology may end up discriminating against those that do not have access to it[254]. Finally, the technology itself must benefit the society as a whole, as opposed

[246]'Draft Ethics Guidelines for trustworthy AI' (December 2018). The European Commission High-level expert group on artificial intelligence. At page 2

[247]Friedman, B. 'Social judgments and technological innovation: Adolescents' understanding of property, privacy, and electronic information. Computer'. Human. Behavior. 13, 3 (1997), 327–351

[248] Villarong E., Kieserberg P. And Li T., 'Humans Forget, Machines Remember: Artificial Intelligence and the Right to be Forgotten' (2017) Computer Security and Law Review. At page 8

[249]Friedman B. and Kahn P.E. 'Human Agency and Responsible Computing: Implications for Computer System Design' (1992). Elsevier Science Publication Co. New York at page 7

[250]Ibid at page 8

[251]Loc. Cit. Contrast with Wachter S. Mittlestadt B. and Russel C. 'Why Fairness Cannot be Automated' (March 2020) at footnote 176, the authors argue that machines cannot be ethical nor can they be moral agents.

[252] Fjeld J., Achten N., Hilligoss H, Nagy A. C and Srikumar M. 'Principle AI: Mapping Consensus in Critical and Rights-Based Approaches to Principles for AI' (January 2020). Research Publication No. 1. Berkman Klein center for Internet and Society at page 33

[253]ibid

[254]Lehman cited at note 116. At page 14

to being a detriment. The IEEE makes a strong case that AI need not only obey societal values, but that the AI system amplifies those values to the benefit of society[255].

Additionally, there are other aims of value sensitive design and these include ensuring that human values, whatever they may be, are considered whenever there is a design of new technology. The scope of the same human values are to be enlarged, as technology evolves, it becomes a necessity as opposed to a luxury and with it there is change of human values, technological change should be reflected of this change.

The EU has actively sought the alignment of human values to AI and this includes human control of the AI systems themselves[256]. This human centric approach to AI could be found in the liability law set up by European commission[257]. The commission has ensured that any harm arising from AI systems should ultimately be attributed to someone, whether they are a natural or legal person.

Human values ought to be reflected in the technological tools that we use, tools are an inseparable part of what makes us human and our tools help us explore the world around us by amplifying our humanity. An example of this is how we use social media to express ourselves and also to communicate, things we would still do without the new technology.

Increased autonomy of AI will reduce the amount of human supervision required, and the question remains of who to blame once something goes wrong. To provide some answers for this, we may consider the three pillars of responsibility in AI systems.

Responsibility in AI relies on three keystones with the first being that society must be ready for any responsibility as a result of AI impact[258]. Essentially, this translates to the developers being aware of the impact of the AI on the society and also that governments and their citizens should decide the liability arising from AI.

[255] Shahriari K. Et al 'Ethically Aligned Design: A Vision For Prioritizing Wellbeing With AI And Autonomous Systems, Version 1. IEEE' (2017). At page 22

[256] Scipione J. 'AI and Europe: Developments, risks, and Implications' (2020). European College of Parma. At page 13

[257] https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608 (accessed on 17th June 2020)

[258] Dignum V. 'Responsible AI: Designing AI for Human Values' (September 2017). ITU Journal: ICT Discoveries, Special Issue No. 1, 25 at page 4

The second keystone would be the need for programmable mechanisms in the AI that allows the system to reason according to human values. These human values can be embedded in the modelling and algorithm of the AI and the decisions made by the said system should reflect human values and justification.

The third keystone concerns the participation of the community in which the AI system is understood to operate as part of the society as opposed to a separate or alien concept of machines that present worry for the community[259]. In the same vein, governance of AI needs to ensure that a progressive view of AI is developed that ensures that AI advances societal good[260].

Most of the concepts of the three keystones of responsible AI development have been adopted by all but three countries in the EU. Fundamental rights continue to be at the centre of AI development in the EU. For example, Finland had proposed the enactment of legislature to ease immigration processes which legislation was found to be a risk factor for fundamental rights upon review. This was so because the proposed law would allow decisions to be made without need of hearing the data subjects[261]. Furthermore, in Finland's domestic review of AI systems, the Data Protection Ombudsman found discrimination by credit companies who used systems that determined credit worthiness based non prohibited classes such as language, gender and sex[262].

The recommendations by the EU agency for fundamental Rights rightly encourages EU member states to ensure protection and enhancement of these rights by AI. The strategy recommended leaves out the lay person and focuses only on experts in computer science[263].

The EU is trying to find a solution of how protected classes may be accorded equality in AI. This is evident by several Directives that aim to promote equality of all individuals within its borders. In particular, the concept of equality data is very useful for tackling discrimination in AI. Equality data refers to data that is free from bias[264]. Of the relevance to this paper, is the acknowledgement by the

[259]Ibid at page 7

[260]Loc. cit.

[261] EU Agency for Fundamental Rights Report 2020. Available at https://fra.europa.eu/en/publication/2020/fundamental-rights-report-2020. At page148.

[262] EU Agency for Fundamental Rights Report 2019. Available at https://fra.europa.eu/en/publication/2019/fundamental-rights-report-2019. At page 160. See also https://www.yvtltk.fi/en/index/opinionsanddecisions/decisions.html

[263] Ibid. At page 166

[264] European Union Commission. High Level on Non-Discrimination, Equality and Diversity: Guidelines on Improving The Collection and Use of Equality Data. At page 8

Commission that stakeholder involvement in data equality processes are negligible and this still poses a challenge as it concerns equality and recognition of diversity in data[265]. I opine that employing value sensitive design will alleviate some of these concerns as the framework allows stakeholder input in the design process of AI including the quality of data being used to train the AI.

Another challenge apart from the inadequacy of stakeholder involvement is that the available EU laws as well as national legislation either has language that is too technical or composed of complicated legal language that disadvantages the greater majority of those who may be subjected to potential AI decisions at some point. There can be no effective stakeholder involvement when the language adopted by authorities is not understood by the biggest stakeholders. A concentrated value sensitive design would identify this challenge and redress it during the development process of AI.

4.5 Conclusions on Value Sensitive Design as a Framework for studying AI Development in the EU.

Non-discrimination remains a core of EU fundamental freedoms and values. The European Union has maintained its stance to safeguard fundamental rights by way of strong regulation and penalties for breach of the same. Previously, EU responses were reactive and remedial but there has been a shift to focusing on wider stakeholder involvement. It must be noted that the stakeholder involvement does not necessarily mean that other end users fully comprehend AI systems as the emphasis continues to be placed heavily on AI developers through regulation by design practice.

Value sensitive design offers a proactive and preventive approach to tackling algorithmic discrimination. Value sensitive design is suited for human cantered AI because of its basis in moral considerations of technology. Value sensitive's three conceptual frameworks of investigation include conceptual investigation, empirical investigation and finally technical investigation. These are the basic components of the framework and they are used for most technologies. The paper has shown that the challenging nature of AI may require additional frameworks to not only prepend human values to AI but to also strengthen the same.

The additional requirements to make value sensitive design a viable framework for AI development and review are designed to tackle societal challenges and concerns with AI. Particularly, the inclusion of moral sandboxing and prototyping tries to address the issues of AI being seen as lacking morality.

[265] Ibid. At page 11

Coupled with the concept of reflexivity, value sensitive design allows for the inclusion of value construction in AI including non-discrimination.

5.0 CONCLUSION

Alan Turing, the famous scientist known for decoding the German Enigma code during World War II, stated that there was need for autonomous systems to be embedded with human values. Granted, at the time that he made this statement, intelligent systems were not as intelligent as they are now and could not influence public and private life as they do now. Turin stated this out of concern for what he presumed would be a slow but sure decay of humanity as AI took on an increasingly larger role in our societies. This ability of AI to overwhelm our humanity makes it much more important to ensure that human values are complied with in the development of AI.

The EU, in its plethora of AI regulation, has managed to strike a balance between the competing interests of AI developers and the values that the union holds dear. However, there should be further consideration of European values in the design of AI systems and currently, the onus happens to be on the developers of the systems.

AI, it has been observed, should not be a risk to human values, but an amplifier of those values. AI collectively makes our lives easier. However, AI does make certain classes of people disadvantaged and discriminated against when they are excluded from data sets. The fact that AI sometimes disadvantages certain classes of people was pivotal in the study being undertaken, to find the means by which algorithmic bias and discrimination occur and also how the same may be addressed proactively.

In the beginning of the study, the paper made two speculations using moral imagination and value sensitive design as it concerns AI. The thesis has expanded on the speculations and the findings from the same will be presented individually. The first speculation was that the ensuring of non-discrimination in AI has been left to the developers and the regulators. The concerned data subjects have been being sidelined and made spectators. The biggest problem that the research has identified is the inadequacy of stakeholder involvement in AI development. The stakeholders in question are those persons who may be subjected to an AI decision at some point in their lives. Regulation of AI has focused on two aspects, these being the developers being tasked with a responsibility and the second being a remedial and reactive approach to incidents of algorithmic discrimination when they occur. Data subjects who should have the biggest voice in AI development have largely been ignored and only resurface when their human values have been infringed upon.

The second hypothesis assumed in the beginning of the study was that value sensitive design could be used to not only study AI development in the EU but also to enhance values of the EU in AI such as the value of non-discrimination. In this research, the relevance of appreciating human values from a moral, legal and philosophical background, using value sensitive design, have been analyzed and applied to appraise the development of EU societal values into the AI technology systems.

The research has shown that studying AI from this framework provides some important observations that may be useful in fostering trust in AI. Such trust also helps in positively effecting widespread acceptance and usage of AI within societies. A moral imagination approach by the developers of AI systems coupled with value sensitive design may offer an answer to inclusive consideration of potential data subjects.

Furthermore, it must be noted that choosing what values to install in AI is not as easy as it may appear. Using EU fundamental rights as a standard for human values is an ideal model for what EU values are to be incorporated in AI design. To a great extent, value sensitive design may be one of the best frameworks for public involvement in the design of AI.

5.1 Suggestions for further research

Moving forward, we ought to be constantly aware that despite the ability of AI systems to make their own decisions, they are tools designed by humans. Humans impart their values into these systems and as AI becomes more advanced, they speedily execute decisions. If the values that the AI systems have been trained on are biased against certain classes, the bias is equally speedily executed in the decision.

Human input remains essential for the development of AI systems that and it is the emphasis of value sensitive design that at the development stage, developers ought to implant EU values and other responsible design into the system such that the system will consolidate those values in its execution of decisions.

Further research is needed on how to incorporate value sensitive design in the continued development of AI. The tripartite approach of value sensitive design can be easily used to monitor the adherence of intelligence systems to the values that the EU holds dear but value sensitive design’s three approaches may prove inadequate on their own. I offer two additional approaches to weighing AI compliance with EU values, these being reflexivity and technological frame. Moral sandboxing and moral prototyping can also be used to assign a sense of morality to AI.

A continued research would focus on those responsible for research and innovation acknowledging the presence of unfairness present in the AI development process and combating this. The research would spotlight on value algorithmic sensitive design with a heavy focus on moral sandboxing and related concepts. Hopefully, the stakeholder involvement would legitimatize AI usage within the public sphere.

Regulation over AI ironically continues to push the myth that AI is a danger which ought to be cautiously avoided. It is easy to observe the EU safeguards as it relates to human rights and AI in the many Directives and Guidelines. However, the lack of end user involvement in the development process is what hinders trust in AI, not the inadequacy of laws. In fact, the many laws regulating AI in the EU have been observed to constrain AI development, at least as compared to China and the US who don't have very rigid AI regulation. A continued value sensitive study would focus on improving end user involvement in the AI development process as the EU prepares to enact the Equal Treatment Directive.