Japanese Ethical Philosophy and Algorithm Design

Nicholas Osaka

Introduction

Data has become the focus of a digital-era gold rush. Data collection, colloquially known as "data mining" has become a major point of investment for corporations across various industries. Data Science has become a burgeoning area of study that corporations are increasingly interested in. At the center of this interest, however, is the study of machine learning. Machine learning is fundamentally the study and application of computer algorithms that use data to improve the algorithm's accuracy in predictions. In recent years, the terms "algorithms" or "AI" (Artificial Intelligence) have been used in media references to such machine learning algorithms.

Our problem comes from both the usage of such algorithms and the development of them. As corporations such as Meta Platforms (formerly known as Facebook) have shown, data collection from unconsenting people is the cornerstone of many business practices. Due to public backlash, increased scrutiny and discussion on machine learning practices have been in the public forum. Algorithm design ethics and the according regulations have largely been focused on oversight—human analysis of the results and outcomes from these algorithms. My purpose in this essay is not to argue against this. Others have discussed the benefits and shortcomings of such approaches.[1] My point is to argue that the traditional Western philosophical canon of ethics is insufficiently equipped to deal with the complexities that "Big Data" presents, and that looking to Japanese ethical philosophies can provide useful perspectives to data ethicists. In addition, feminist ethics and Japanese ethical philosophy both share an orientation of intimacy, as described

---

[1]Young et al. survey municipality ordinances regarding algorithmic oversight and advocate for adjustments in policy. (Young, Katell, and Krafft 2019) Contrasting, Ben Green argues that human oversight of algorithms are ineffective and project a false confidence in oversight which leads to harmful algorithms remaining as the seat of decision making. (Green 2021)

---

in Kasulis's *Intimacy or integrity : philosophy and cultural difference.*

I must first ask what the Western ethical canon has come to mean in a privacy context. Following this brief investigation, I will be exploring what differences may be uncovered in various Japanese ethical philosophies. This will allow for a perspective built on collective ethics rather than individual ethics, prioritizing intimacy. While it can be tempting to want to find a way "out" of the problems AI presents, a view inspired by Watsuji Tetsuro's ethics reveals to us that AI cannot be permitted to permeate across our social and political systems. At least, not as AI currently positions individuals in relation to each other. The way "out" is by fundamentally restructuring AI from the ground up—not by implementing oversight programs over existing AI structures.

Western Ethics and Data

Philosophical moralists such as Kant, Mill, and Hobbes (to name a few) have been largely focused on "ethics of rights," argues Carol Gilligan. She claims that such ethics of rights approaches people as "separate but equal individuals." (Darwall 1998, 220) We see this expressed in Kant's Categorical Imperative for example, which places the individual in a position where they must act in a way that implies an equality of all in belonging to moral consideration. No one person is deigned as more morally important than another. This focus on individual rights-driven ethics is riddled with problems for the complexities that data ethics needs to handle.

In a practical application, the Western ethical canon is broadly concerned with the individual. Unquestionably, these ethical frameworks which emphasize an ethics of rights are influential in Western public policy. In California, the California Consumer Privacy Act (CCPA) provides the individual consumer in California with four rights:

(1) the right to know what personal information a business has collected about them and how it is being used; (2) the right to "opt out" of a business selling their personal information; (3) the right to have a business delete their personal information; and

(4) the right to receive equal service and pricing from a business, even if they exercise their privacy rights under the Act. (Pardau 2018, 72)

The CCPA is loosely modeled after the European Union's General Data Protection Regulation (GDPR). The GDPR influences the handling of "the personal data of all EU residents, regardless of the location of the processing" and principles "fairness and lawfulness[,] purpose limitation[,] data minimisation[,] accuracy[,] storage limitation[,] and integrity and confidentiality" as paramount to data privacy. (Goddard 2017, 703) The GDPR and CCPA both place an *individual* right to privacy as paramount and indeed, the centerpiece of regulating algorithmic usage. While there is an implication of *some* collectivist ethics insofar as the GDPR has 'fairness' as a principled ideal (since 'fairness' carries an implicit wariness of how an algorithm treats groups of people), the emphasis on the whole is on the individual.

While many praise the GDPR for its commitment to individual data privacy, there are concerns on its impact to the collective whole. Celeste Cagnazzo argues that there is a valid concern regarding public health research. The GDPR requires researchers to "report the characteristics of future research in detail" for authorization to process sensitive data. (Cagnazzo 2021, 1495) The argument is that future research cannot always be described in detail, and this requirement hampers potentially life-saving public health research. Cagnazzo highlights the danger that slow processes pose to global fast moving threats such as COVID-19. A "rigid and uneven implementation of GDPR" has frustrated research efforts in parts of the EU, and she argues that this is the downside to policy decisions guided by individual ethics. (1496)

I am not arguing for the prioritization of some arbitrary "well-being" of a social whole over the well being of individuals. Rather, I am arguing that the individual rights perspective often overshadows the implications of living in a larger social group. Of course, the Western canon and its contemporaries hold great value in data ethics. In a forthcoming issue of *Daedalus*, Iason Gabriel discusses justice in the context of AI. Drawing from Rawls, Gabriel offers us a liberal view of justice that applies to artificial intelligence and their usage in social structures. Gabriel argues that AI systems ought to be "subject to regulation by principles of distributive justice" in

the Rawlsian sense. (Gabriel 2022, 4) Gabriel is a proponent of the Difference Principle because of its requirement that "for institutional practices to be just, all inequalities in the distribution of 'social primary goods' must work to the greatest advantage of the least advantaged member of society." (9) He argues that this is a stricter guiding principle than just attempting to eliminate injustice in AI systems. Gabriel's proposal interlocks well with contemporary liberal political society, and provides us with a clear argument in favor of distributive justice. While the proposal works well within the current tides of AI, I argue that a more aggressive approach must be taken.

Watsuji and Ethics

Japanese philosopher Watsuji Tetsurō is noted for being one of the prominent ethical philosophers from 20th century Japan. In his seminal work, *Rinrigaku* (ethics), Watsuji objects to the highly individualist ethics of the Western canon. Watsuji approaches ethics through his concept of *ningen* (人間, human). This is the Japanese word used as the analogue of "human being." Watsuji notes the implication of *ningen*—the dual meanings of the paired characters and the tension between them. The first character, 人 (*hito*) can alone be used to mean human being or person. However, the existence of the second character, 間 (*aida*) tells us there is more to the story. The *meaning* of this second character sheds much light on why Watsuji is unmoved by the Western canon's individualistic ethics: *aida* brings forth ideas of space, between—a "gap." This betweenness of people brings about Watsuji's foundational point: "Individual persons do not subsist in themselves." (Watsuji 1937, 101) However, this is not to say that a collective or "social whole" is the solution to the questions posed by *ningen*. Rather, Watsuji focuses on a dual nature of human beings: as the self and as part of a greater whole. One cannot be presupposed before the other, as both occur in their "reciprocal negations." (102)

Watsuji's framework may seem to advocate for the dissolution of individuality, but this is not the case. Watsuji views "human beings not only as individual but also as social in the betweenness (*aidagara* 間柄) among selves in the world." (McCarthy 2010, 20) The ethical identity only takes shape when both are made possible by the other. It's important to note that this does not set

up a dualism. Taigen Dan Leighton writes on the nonduality in Zen Buddhism that Watsuji uses: "nonduality is not about transcending the duality of form and emptiness. This deeper nonduality is not the opposite of duality, but the synthesis of duality and nonduality, with both included, and both seen as ultimately not separate, but as integrated." (Leighton 2007, 79) Further driving the difference Watusji's philosophy has to the Western canon, Erin McCarthy explains that "Japanese philosophy has a nondualistic view of the body," in contrast to the Cartesian dualism which separates and places a hierarchy on body and mind. (McCarthy 2010, 16) The nonduality of Watsuji's ethics is disclosed by the ways in which the "gap" or between-ness of people is treated. There is an intrinsic dynamic quality to this view of identity: "One becomes an individual by negating the social group or by rebelling against various social expectations or requirements." (Carter 2013, 136) This is not a rejection of the social, but an expression of the rich qualities of identity as *ningen* (人間). To Watsuji, ethics (and by extension, humanity) is only possible in this rich between-relation that borrows understanding of negation from Zen Buddhism.

However, this "intimacy" of the social in relation to the individual is not exclusive to Watsuji or Eastern Philosophy. Thomas Kasulis argues that an orientation of intimacy relates the self and other "in a way that does not sharply distinguish the two," (Kasulis 2002, 32) contrasted against orientations of integrity, which emphasizes "external over internal relations." (25) McCarthy argues that "[i]n both Japanese and feminist philosophies we find concepts of self that provide alternatives to the concept of self as the autonomous, isolated individual, and the ethics that results from such a conception of self." (McCarthy 2010, 51) The parallels between the two philosophies are in their orientation of intimacy. An approach to data ethics from a feminist philosophy perspective may yield similar advantages, for the reasons outlined. It is no coincidence that feminist philosophy can bring us to a similar vantage point—after all, feminist ethics of care places importance on "care for the particular others to whom we are related *within* the various different relations of care and concern we share with them." (Darwall 1998, 221)

AI and Latent Space

The major focus on "ethical AI" or "responsible AI" has been in regards to regulating the data collection on individuals. In other words, protecting the individual's right to privacy. I don't argue that this is a misguided project. I do, however, argue that it misses the most insecure point in ethical AI: understanding how the individual and social whole interact and only operate in the context of the other.

Machine learning models are not developed by thinking of how individuals are positioned within the problem domain. Rather, these algorithms are *trained*. Trained on what? Hundreds of thousands of data points—a tech term that de-personifies personal data about individuals in the context of the problem domain. Notably, this means that when the machine learning model makes predictions about whether a person is likely to want a product, or how much to charge a person to maximize profit, the model is actually making an inference based on past data about similar individuals. Therefore, an individual's right to privacy provided through such individualistic ethics is rendered inert once a person who is deemed "similarly located" by such algorithms voluntarily offers their data, for one reason or another. Watsuji's ethics places the spatiality *between* individuals as necessary to understanding the ethical project. I posit that we can view spatiality in algorithms as how models place individuals in digital space. By placing individuals "closer" to each other in this digital space, the group around the individual encompasses those individuals that the algorithm "clusters" together based on the problem domain. Notably here, these individuals likely do not know each other. In the digital age, individualist ethics can be incredibly helpful for a liberal society's endorsement of liberty and a right to privacy. However, I argue that this is insufficient for ethically contextualizing how people are located in digital space.

Indeed, the determining factor in how these individuals are located are their "features." In machine learning and statistics, models use features to predict the resultant outcome. These features may be continuous (e.g. bank account balance), or discrete (e.g. ZIP Code). Through a synthesis of features, where each feature is prioritized or weighted during the training process, we arrive at the final "clustering" in the digital latent space. When the algorithm runs a prediction

given the features derived from an individual, that individual is located in the digital latent space and then the inferred value is based on where the individual was located. Concretely, if one applies to receive a bank loan, a machine learning algorithm may be used to determine whether the bank should provide one or not. The applicant is then reduced into a set of features, and the resulting location in digital latent space determines whether the applicant receives a loan. The individualist ethics would view the individual as largely unrelated to others as it concerns the loan decision. However, Watsuji's *rinrigaku* would demand that the digital latent spatiality be reckoned with when evaluating the ethical situation.

The major difficulty in applying Watsuji here is that the subject of consideration, the loan applicant, cannot actually recognize their relation to those who occupy the nearby latent space. It is, after all, obscured and digital. Re-positioning our perspective to that of one evaluating the ethicality of a system, we can indeed apply these principles. We need to evaluate how the system actually clusters individuals. This can become problematic if not done critically. We ought not to desire to identify the ways individuals are clustered as finding some "hidden" aspect of humanity. Machine learning practices encourage this kind of thought. Data mining—as mentioned in the introduction—as a practice encourages (but does not necessitate) machine learning practitioners to seek to find "hidden features" about the problem domain data. Quickly, we see how the reduction of humans into features and the process of seeking "hidden features" can be seen as modern phrenology. Especially when features can be found to be analogous for race, class, gender, and sexual orientation without explicitly stating so.

Weights and Biases

Statistically, the machine learning model takes features, then applies weights and a bias (almost humorously named) as part of the training process. In practice, machine learning models inherit bias from their training data and reify institutional biases. Members of marginalized groups are more likely to be clustered together than not. The exact reason varies from model to model, but machine learning models mirror our real world biases and social encodings. If a particular ZIP

code encodes a racial and/or class difference implicitly (be it through funding for schools or governmental support), the model will likely encode those same implicit biases into the latent space. Mehrabi et al. survey the kinds of biases in algorithmic processes and find two types of biases especially relevant to Watsuji's ethics. Measurement bias "arises from how we choose, utilize, and measure particular features." (Mehrabi et al. 2021, 5) This is the most plain form of bias. Watsuji's ethics would demand that we consider how particular features functionally treat groups of people, rather than individuals. In fact, Mehrabi et al. already allude to this: a recidivism risk prediction tool used "prior arrests and friend/family arrests ... as proxy variables to measure level of 'riskiness' or 'crime'." (5) The very fact that the designers of this system were already thinking of the social relations of the individual when making this system tells us that thinking in an orientation of intimacy will be valuable. Further, the authors also discuss aggregation bias: "when false conclusions are drawn about individuals from observing the entire population." (5) This kind of bias pulls to the strength of Watsuji's ethics—the double negation from which individual and social identity are formed. This way of viewing identity (and subsequently ethics) as dynamic between self and other enable us to conceptualize aggregation bias more concretely than just as 'misrepresenting individuals'. Understanding that the identity is both individual but also social (and dynamic) forces us to throw away systems which attempt to generalize in such a way.

However, I want to note that it's not possible to wave away problems in algorithmic processes by just applying Watsuji's ethical framework. Because these biases are present in language, as in the case of gender bias, biases must be addressed in our social dimensions. Susan Leavy notes that "men [are] more frequently described in terms of their behavior while women [are] described in terms of their appearance and sexuality" and that gender binary hierarchies became encoded into word embeddings (a kind of word representation used in machine learning models relating to natural language). (Leavy 2018, 15) Birhane, Prabhu, and Kahembwe found that when searching image databases used for computer vision problems, pornographic images and images depicting sexual violence were often in the results for queries unrelated to such depictions. When "even the weakest link to womanhood or some aspect of what is traditionally conceived as feminine" returns such images, it is clear that evaluating gender bias in machine learning cannot be done by addressing how models treat individuals. (Birhane, Prabhu, and Kahembwe 2021, 4)

Adjacent to these incredibly grounded issues of bias (ones that explicitly highlight issues inherent in current AI structures), there is also the issue of how AI is treated by those who develop them as well as fund them. Natural language processing is a burgeoning sub-field with significant advances made recently due to large language models. These large language models push the envelope in what was previously possible—both functionally and architecturally. GPT-3 (Generative Pre-trained Transformer 3) is one such large language model. GPT-3 is trained on massive amounts of text, and when prompted with a question or some sample text, can provide seemingly human-written text with semantic meaning. Bender et al. call these language models "stochastic parrots" pointing to the fact that the statistical nature of these models essentially makes them akin to a parrot. They argue that "human-human communication relies on the interpretation of implicit meaning conveyed between individuals" and therefore when these models are trained on data with these problematic biases, models re-inscribe harmful and dehumanizing language. (Bender et al. 2021, 616). By the very usage of implied biases by these models, such language is not only re-articulated but also validated as legitimate. It is no coincidence that this paper led to the firing of Margaret Mitchell (co-author Bender et al.) from Google. The interest private corporations have in AI outweighs their concern for the explicit harm (i.e. clear and evident bias) as well as more implicit harm (i.e. how AI positions individuals relative to each other). Therefore, I argue that large language models are a point of significant political and social issue, and current regulation and oversight methods are simply unequipped to deal with the engineering culture creating these models.

It is only by an ethics which prioritizes orientations of intimacy (per Kasulis) and balances the individual and social (as in Watsuji and McCarthy) that we can even begin to appropriately address the biases that emerge in these systems. As stated in the Introduction, the way "out" of this ethical dilemma is not through regulation or through oversight programs. The way AI fundamentally positions individuals in relation to each other ignores the very social aspect of our identities, and therefore requires a restructuring from the ground up.

Conclusion

In considering the ways in which we, as both individuals and a social whole, interact with models the question of biopolitics comes forth. While outside of the scope of this paper, there is indubitably a biopolitical component to these interactions. Perhaps the most clear way in which this biopolitical component occurs is the shift in decision making power from manual (and labor intensive) human work to machine learning models. Read most literally and most plainly, when it comes to issues of the state where algorithmic processes are used (e.g. bail recommendations) the biopolitical question becomes forthright.

While the Western canon has been helpful in a liberal society for generating individual rights to privacy and guiding us to create regulations around algorithmic design and implementation, much is left to be desired. I argue that Japanese philosophy from the Kyoto School, specifically from that of Watsuji, allows for us to view people as *ningenn*. As *ningenn*, we are held in a tension of between-ness, where both the individual and between-ness interact together to create a site where both compassion and selfless-ness can occur. Parallels between such ethics and an ethics of care have been drawn, which is a valuable correlation to make. I argue that Watsuji's ethics lets us consider what data and algorithmic design ethics would be like if we imagine how to consider those around us in digital latent space (e.g. how a machine learning model may cluster us). I do not argue that viewing individuals as points in latent space is a *good* way to view people, since it is not. Rather, it is *helpful* to understand how models locate us in a digital latent space. Being able to utilize a framework that is equipped to handle the tension (and indeed, necessity of that tension) between the individual and the between-ness will prove invaluable in identifying how systems fail individuals and society writ large.

References

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021. Accessed December 7, 2021. `https://doi.org/10.1145/3442188.3445922`. `10.1145/3442188.3445922`.

Birhane, Abeba, Vinay Uday Prabhu, and Emmanuel Kahembwe. "Multimodal datasets: misogyny, pornography, and malignant stereotypes." *arXiv:2110.01963 [cs]*, October 5, 2021. Accessed December 1, 2021. `http://arxiv.org/abs/2110.01963`. arXiv: `2110.01963`.

Cagnazzo, Celeste. "The thin border between individual and collective ethics: the downside of GDPR." Publisher: Elsevier. *The Lancet Oncology* 22, no. 11 (November 2021): 1494–1496. Accessed November 2, 2021. `https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(21)00526-X/fulltext`. `10.1016/S1470-2045(21)00526-X`.

Carter, Robert Edgar. *The Kyoto School An Introduction*. Book Title: The Kyoto School An Introduction ISBN: 9781461921363 Place: Albany. N.p.: State University of New York Press, 2013.

Darwall, Stephen. *Philosophical Ethics*. Boulder, CO: Westview Press, 1998.

Gabriel, Iason. "Towards a Theory of Justice for Artificial Intelligence." *Daedelus* 151, no. 2 (2022): 12. `https://arxiv.org/abs/2110.14419`. `https://arxiv.org/abs/2110.14419`.

Goddard, Michelle. "The EU General Data Protection Regulation (GDPR): European Regulation that has a Global Impact." Publisher: SAGE Publications. *International Journal of Market Research* 59, no. 6 (November 1, 2017): 703–705. Accessed November 30, 2021. `https://doi.org/10.2501/IJMR-2017-050`. `10.2501/IJMR-2017-050`.

Green, Ben. *The Flaws of Policies Requiring Human Oversight of Government Algorithms* ID 3921216. Rochester, NY, 2021. Accessed November 30, 2021. `https://papers.ssrn.com/abstract=3921216`. `10.2139/ssrn.3921216`.

Kasulis, Thomas P. *Intimacy or integrity : philosophy and cultural difference*. ISBN: 0824824768 Place: Honolulu Publication Title: Intimacy or integrity : philosophy and cultural difference. Honolulu: University of Hawai'i Press, 2002.

Leavy, Susan. "Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning." In *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, 14–16. GE '18. New York, NY, USA: Association for Computing Machinery, 2018. Accessed December 1, 2021. `https://doi.org/10.1145/3195570.3195580.` `10.1145/3195570.3195580`.

Leighton, Taigen Daniel. *Visions of awakening space and time: Dōgen and the Lotus sutra*. Book Title: Visions of awakening space and time : Dōgen and the Lotus sutra ISBN: 9780195320930 Place: Oxford ; edited by Dōgen. N.p.: Oxford University Press, 2007.

McCarthy, Erin. *Ethics Embodied : Rethinking Selfhood through Continental, Japanese, and Feminist Philosophies*. Lanham, MD: Lexington Books, 2010. Accessed November 3, 2021. `https://ebookcentral.proquest.com/lib/uncc-ebooks/detail.action?pq-origsite=primo&docID=634237`.

Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys* 54, no. 6 (July 13, 2021): 115:1–115:35. Accessed December 1, 2021. `https://doi.org/10.1145/3457607.` `10.1145/3457607`.

Pardau, Stuart L. "The California Consumer Privacy Act: Towards a European-Style Privacy Regime in the United States." *Journal of Technology Law & Policy* 23, no. 1 (2018): 68–114. Accessed November 30, 2021. `https://heinonline.org/HOL/P?h=hein.journals/jtlp23&i=70`.

Watsuji, Tetsurō. *Watsuji Tetsurō's Rinrigaku*. Translated by Yamamoto Seisaku and Robert Carter. N.p.: State University of New York Press, Albany, 1937.

Young, Meg, Michael Katell, and P. M. Krafft. "Municipal surveillance regulation and algorithmic accountability." Publisher: SAGE Publications Ltd. *Big Data & Society* 6, no. 2 (July 2019): 2053951719868492. Accessed November 30, 2021. `https://doi.org/10.1177/2053951719868492.` `10.1177/2053951719868492`.