Essay

The Rise of Particulars: AI and the Ethics of Care

David Weinberger

Berkman Klein Center for Internet & Society, Harvard University, Cambridge, MA 02138, USA; david@weinberger.org

Abstract: Machine learning (ML) trains itself by discovering patterns of correlations that can be applied to new inputs. That is a very powerful form of generalization, but it is also very different from the sort of generalization that the west has valorized as the highest form of truth, such as universal laws in some of the sciences, or ethical principles and frameworks in moral reasoning. Machine learning's generalizations synthesize the general and the particular in a new way, creating a multidimensional model that often retains more of the complex differentiating patterns it has uncovered in the training process than the human mind can grasp. Particulars speak louder in these models than they do in traditional generalizing frameworks. This creates an odd analogy with recent movements in moral philosophy, particularly the feminist ethics of care which rejects the application of general moral frameworks in favor of caring responses to the particular needs and interests of those affected by a moral decision. This paper suggests that our current wide-spread and justified worries about ML's inexplicability—primarily arising from its reliance on staggeringly complex patterns of particulars—may be preparing our culture more broadly for a valorizing of particulars as at least as determinative as generalizations, and that this might help further advance the importance of particulars in ideas such as those put forward by the ethics of care.

Keywords: AI; ethics; ethics of care; morality

1. Introduction

When it comes to caring about us, AI and hammers are tied at zero. Yet, AI may be setting the stage for a broad and deep change that puts caring relationships at the heart of our moral thinking.

For this change to occur, AI would have to do to us what dominant technologies from prior ages have done: give us new models for understanding ourselves and our world. Watches, the peak technology of the 17th century, gave us a universe that looked like a clockwork. The Age of Steam brought us to feel anxiety as pressure. When telephones became common, our neural system started looking like wires connected to switchboards. In the Age of Computing, our brains became information processors. In the Age of the Internet, everything started to look like a network. And now, in the Age of AI—more specifically, the Age of Machine Learning—our ideas about how the world works may be undergoing an even deeper shift: from looking to general rules and laws for explanations and predictions, to emphasizing the importance of stubborn, complicated particulars.

How our tools have this effect on our way of understanding our world and ourselves is something of a mystery—especially since this too is likely interpreted through our relationship with our technology But assuming tech continues to have a transformative power on our thinking and experience, then our engagement with soulless machine learning (ML) may help us embrace the modern feminist philosophy known as the *ethics of care* which asks us to focus on the particulars of moral situations, rather than looking to general moral frameworks to explain what makes right actions right in the first place.

This would be not only a significant evolution in our moral thinking. It would also be evidence that particulars may at last be gaining the respect they deserve.

2. Machine Learning's Weird Generalizations

A traditional computer program can compute your taxes, course-correct a space probe, or keep track of your dog's workout schedules because programmers have determined for each program what the relevant factors are and the logic that connects them.

In contrast, when training a machine learning system, the developer doesn't tell the system anything about what counts and how the pieces interact. Instead, the computer figures all that out on its own by looking for patterns in the mountains of data that these systems are typically trained on.

These patterns can be quite particular. For example, if being trained on images of flowers, an ML system might notice a pattern in the arrangement of pixels with high values for what we see as yellowness. Perhaps with some measurable strength of correlation, they tend to cluster in particular areas of images labeled "tulip". But the patterns can also be so complex that we don't see why they tend to be distinctive of images of tulips. In fact, these patterns can be so multidimensional—that is, they can be patterns of the relationships of multiple factors—that we cannot make sense of them at all because our minds do not work that way.

Let's take as our example a fictitious ML model that predicts the right selling price for a house. It has been trained on tens of thousands of houses recently sold in your region. It has been fed an enormous amount of data about each house, including its selling price and the number of bedrooms and bathrooms, of course, but also the size of the front and back yards, how many cars can park in the driveway if there is one, whether there is a fence in front or in back, the ratings of the public schools, the number of sunny days per year, how many colors the house is painted, the distance to a cozy coffee shop, the percentage of the neighborhood that votes in off-year elections, the mix of local birds, and on and on, perhaps including factors that seem to be wildly irrelevant.

If you were to draw up a two-dimensional chart, you might plot the number of bedrooms on the horizontal axis vs. the selling price on the vertical, and draw a line showing their relationships. You might awkwardly add a third dimension for how many bathrooms there are. Soon you stop, though, because you're not going to be able to read a chart that plots a fourth, fifth, or two hundredth dimension. Machine learning, however, can take each of the factors and treat it as an additional dimension, and use this to tell you what it thinks is the right selling price for your house.

But it's actually far more complex than that. Not every dimension is equally significant. We may think we can figure out the significance of each of them, but we may well be in for a surprise, especially when dimensions are in complicated relationships with one another. But during its training process, machine learning will churn through the various dimensions of data, looking for correlations and statistical significances (expressed as a numerical weight) that create a model that is satisfactorily accurate in its prediction of sale prices. Who decides what counts as "satisfactorily"? Presumably, the sponsors of the ML project.

At the end of the process, we have an ML model, which is a software application that accepts input (data about the house we're pricing), looks for the patterns that it has learned are significant, applies the weights it has come up with those patterns, and outputs a sales price.

The model as a piece of software is literally a generalization in that we can input the data about any house in our region and it will give us a selling price. But it does not work the way traditional generalizations do, and most important, models-as-generalizations do not do what we have liked generalizations for.

First, we have liked generalizations because they simplify complex situations by showing us what's the same in some salient regard about the different particulars it covers. For example, before machine learning, weather predictions were sometimes generated by applying [1] Newton's laws to data about the current air pressure, movement, moisture, and the like. Once weather prediction started using machine learning, it was able to discern much finer and more complex relationships in the data that were gathered, and its

predictions became far more accurate because it could handle far more particulars expressed as data points, and far more complex interrelationships of patterns Those patterns are, in some way, generalizations, but unlike the generalizations we have traditionally liked, they may not be simple in themselves, and the conjunction of them in a model complicates them further.

Second, we have liked generalizations because we can understand them. That is not surprising since traditionally we humans are the ones who have created or recognized the generalizations. But when you leave generalizing to a machine that can encompass more complexity than we can, the resulting model may not be intelligible to us.

Third, we have liked generalizations because we can apply them. But the magnitude and complexity of these generalizations means that only the model can apply them. That's the sense in which the model *is* the generalization.

That we need to use the model, embodied in hardware, to apply what it's learned is not merely a pragmatic limitation[1]. Since one of the great lures of generalizations is that they give us power and mastery over our environment, this dependency on machine learning models is a slap in the collective human face . . . or, as many of us would say, it is an urgently needed correction to the arrogance that has led to the west's possibly fatal sense of global mastery and entitlement.

Still, it would be flat out wrong to say that machine learning does not generalize. Indeed, there is a word for ML models that fail to generalize: they are *over-fitted*. This occurs when the training process has memorized the particulars about the items it has been trained on, but has failed to find patterns that can be successfully applied to inputs it *has not* been trained on.

So, machine learning models generalize, but they put particularity and generality into a new configuration that we can think of as a Hegelian synthesis of contradictories that raises each of those contradictories to a new level. Like a traditional generalization, an ML model can be applied to many cases. Unlike a traditional generalization, it does not absorb particulars into a single generalization but rather maintains them as necessary for applying the generalization to cases. Likewise, unlike traditional particulars, these new generalized particulars are manifested as very specific patterns among sets of them. The particulars are expressed in their similarities, and the generalizations work because of their refusal to reduce these new particulars to what they have in common.

This new synthesis of the particular and the general is not a neutral change, for in the west we have favored the general both epistemologically and metaphysically: the general has been the truth behind the chaos of particulars.

By wresting benefits from particulars, machine learning gives us in the west a reason to respect particularity. And respect for particularity can be a first step toward an ethics that dismantles the traditional moral frameworks, not by replacing them with a new one, but by undermining the very notion of moral frameworks.

3. Care's Ethics

What is the ethics of care? That is going to take some mansplaining—some cis–white-hetero--mansplaining.

The ethics of care can be traced back to Carol Gilligan's 1982 book *In a Different Voice: Women's Conceptions of Self and of Morality* [2] that shocked that era's predominantly male philosophical discipline by arguing that women and men tend to think about morality differently. Of course, that puts it too simply; gender obviously is complex and fluid; the politics of the patriarchy are pervasive and complicated; cultures vary in this; and intersectionality—considerations of race, economic class, and so forth—has crucial effects. (Intersectionality is a type of multidimensionality [3]).

In any case, Gilligan found, to take an over-simplified example, that if she asked a moral question such as "Is it morally ok to steal medicine that you can't afford if it will save your parents' lives?", men tended to say it is wrong because stealing is wrong. Women

tended to look at the particulars of the case and the caring relationship of child to parent, concluding that stealing in a case like this is morally ok, if not morally necessary.

In short, and too simply, men looked to a universal principle. Women looked at the particulars of the case and the human relationships involved.

Re-founding interpersonal morality on human relationships is key to the ethics of care. The idea is simple and profound. The ethics of care suggests that rather than trying to come up with moral laws, principles, and frameworks, the application of which we will argue about for thousands of years, suppose we begin with the most common and inarguable *experience* of moral behavior: a mother's caring relationship with their child. Suppose we take that *caring for* as the model and prototype for moral behavior [4–7]. What would we see? (Please note that of course people of all genders can be the primary and loving givers of care to children. But for good, bad, or complexly mixed reasons, mothers are the prototype in the west).

I will not pretend to encapsulate an answer to what this prototypical example of a caring relationship expresses because it is such a rich question on which so much great work has been conducted by feminist ethicists. But aspects of this seem to me to pertain to what machine learning is teaching us . . . while not incidentally subverting much of the western philosophical tradition in which I was brought up.

High among the differences is that the ethics of care start from seeing moral agents not as individuals who enter into a relationship with another, but as being fundamentally relational in their own selfhood. As the philosopher Virginia Held puts this core idea:

> The ethics of care works with a conception of the person as relational. . . [T]he goal for persons in an ethic of care is not the isolated, autonomous rational individual of the dominant, traditional moral theory. It is the person who, with other persons, maintains some and remakes others and creates still other morally admirable relations [5] (p. 135).

The philosopher and psychological researcher Alison Gopnik talks about a caring relationship as an "expansion of self": "[A] parent or a child or a partner, or even a good friend, is a person whose self has been expanded to prioritize the values and interests of another" [4] (p. 59). The cared-for's needs, interests, and personhood become one's own. The relationship comes first; the mother and the child literally only are what they are—mother and child—in that relationship. Importantly, it is an asymmetric relationship in which the parent feels a responsibility for the welfare and the good of the child that the child does not equally feel (or have) for the parent.

But where does this sense of responsibility come from? This enables Nel Noddings, one of the earliest writers about the ethics of care, to reconsider the primacy of the moral *ought*, the power of which to compel action has long puzzled philosophers. The mother's care for their child is experienced not as an *I ought*, says Noddings, but as an *I must*: Not "I ought to attend to my crying child," but "I must" [6] (pages 47–49). The "I must" is based in an undeniable relatedness that is felt and thought in mind and body.

But how does this help us make moral decisions? What are the rules for responding in a caring way? The rule is: attend to the particulars with the cared-for's interests at heart. Can we say something definitive about those interests? Various philosophers of care ethics give different answers to this—aiming at the cared-for's autonomy, being a caring person themselves, and so forth—but there is no single, definitive answer because it depends on the complex interrelation of the people in the caring relationship. This means the ethics of care faces situations arguably more complex even than situated ethics. The answer in practice will emerge from a relationship that is deep, embodied, contextual, multi-dimensional, ever-changing, and mutually interdependent.

Still, if we take caring as the basic moral comportment, how does it apply to strangers, people in countries far from ours in kilometers and customs, other sentient and feeling creatures, and on and on? Further, a caring relationship inevitably affects more than just the dyad of the giver-of-care and cared-for. How do we proceed morally? As Martha Nussbaum says in her 1990 book, *Love's Knowledge*—a work aligned with many of the

themes of the ethics of care—"[W]hat we really want is an account of ethical inquiry that will capture what we actually do when we ask ourselves the most pressing ethical questions"[2] [8] (p. 24).

To a large degree, it comes down to the particulars.

4. Care and AI in Particular

What do you do when the bill comes for a restaurant meal shared by five old friends? The obvious solution might be to divide it equally. But that becomes less obvious as we start to add particulars.

For example, suppose that friends A through D are all making enough money to split the bill without squabbling. But—to add a particular—E has been down on their luck, and footing their share of the bill would require skimping elsewhere. Now, perhaps the right thing would be for each to pay just for what they ordered.

But another particular: Friend B is a billionaire and notices E's small signs of discomfort. If they split the bill evenly, after the meal B might privately offer to pay E back. Or perhaps B can make the offer while at the table if the seating arrangement allows them that privacy, which is another particular. Either way, that might make E feel like a "charity case". That depends on E's sense of pride—yet another particular.

Or the billionaire might say to everyone, "You know what? This one's on me", which would be a caring thing to do, although E's pride might take it as another form of charity.

One last particular: suppose that E is the one who ordered the wagyu beef with a quarter pound of fresh truffles wrapped in gold leaf, the 1961 Bordeaux, the five-part flambé dessert, and the ancient sherry. Would those particulars—ones that do not reflect well on E—change what the group should do?

Assuming all five friends want to do the *caring* thing, they won't be guided by universal moral frameworks but by the particularities of the situation and their relation to one another. They will have to make a series of uncertain judgments about the effects, near and far, of their actions, and weigh conflicting values such as honesty, friendship, fairness, and trust[3] [9].

Further, their decisions are unlikely to give them much guidance when next week they face a problem that might be structurally exactly the same—five friends out to dinner, one of whom is down on their luck—but involves a different set of friends in situations that differ even in small particulars from last week's conundrum. They will have "no single metric along which the claims of different good things can be meaningfully considered..." [8] (page 36). Nussbaum calls this "noncommensurability". We might note that in AI's terms, this is another instance of the importance of multidimensionality.

The problem is that the moral frameworks that are supposed to explain and guide moral behavior focus on what moral situations have in common. That is what makes religious ethics, deontological ethics, consequentialist ethics, and other traditional frameworks into frameworks. One framework to rule them all, where the "them" are the concrete situations to which they are to be applied. Instead, the ethics of care says that our moral duty is to care for others, and to do that one must pay close attention to their particular needs, wants, and values in the concrete situation one is considering.

For that reason, some proponents of the ethics of care—as well as other philosophers— suggest we look at fiction to train our moral sensibilities, for great novels and stories often put characters into difficult and quite precise moral circumstances. We then see inside the characters as they consider what can be minute and subtle particularities of the individuals, their relationships and of the situation they find themselves in. For example, Nussbaum unearths the moral thinking in Henry James' *The Golden Bowl* with exquisite skill [8] (pages 85–93), and Richard Rorty [10] discusses Jane Austen as he turns "away from the idea that morality is a matter of applying general principles"[4] [10,11]. Novels may be better sources of moral education than traditional moral philosophers precisely because of the close attention novelists pay to the particulars of each situation.

Both the ethics of care and machine learning models refuse to sand a case down so that it fits a simple, general mold. They want to let the particularities speak, to value the differences and the contradictions, and to do the best they can without denying the complexity and humanity of the situation.

But of course, ML does not care about anything. It is designed to deal with particulars only because that lets its output be more accurate and reliable. It may well miss problems germane to the subsumption of individuals into groups, such as AI's easy and frequent under- and misrepresentation of marginalized groups, including people of color, women, and the non-binary, as well as people in parts of the world that—for many reasons, including benighted capitalism and the self-centeredness of many of the technologically advanced cultures and economies—have not generated much data.

Even so, machine learning's refusal to erase the particulars it has been given in favor of the general may help to advance the acceptance of the ethics of care. The multidimensionality of machine learning models—their ability to relate data across vast numbers of types of connections—could even help us recognize the intersectional complexity of the unfairness imposed on those in multiple marginalized groups [12,13]. But *could* is a long way from *does*.

5. The Fear of AI Leads the Way

It is one thing to point out a formal resemblance between a headline-dominating technology and a philosophy that has not exactly grabbed the public's attention, at least not yet. It is another to suggest that our awareness of the tech might be prepping us to embrace the philosophy.

Yet there are ways it might happen, ironically, through the two main fears and anxieties the technology is raising. Wrapped in those fears is a message about the power of the particular in its new synthesis with the general.

First, there has been, thankfully, a tremendous increase in the awareness of, and research into, machine learning's inherent tendency toward bias due to its reliance on data that, without scrupulous attention, are likely to reflect pernicious social biases. Similarly, the designers of an ML model may not have thought through the social and moral impacts of the objectives they have set. For example, a bus route designed by ML that has successfully decreased *average* transit times has generalized too far if the distribution of transit times skews dramatically shorter for the richer parts of town. The pushback against machine learning's tendency towards bias—its original sin—makes us more aware of the stubborn reality and definitive importance of particulars.

Second, every time we hear the anxiously uttered refrain about a model that "We don't know how it works," we also hear an acknowledgement that *it works*. ML's fastidious attendance to more particulars than we could count or imagine is the source of both its power and its tendency to be inexplicable.

A turn toward the particular does not guarantee that the ethics of care will become more prominent. Nor does it mean that it needs the cultural attention that AI brings: It is becoming more discussed and prominent on its own. But if AI's presence helps valorize the stubborn particulars—which we sometimes refer to as "the real world"—perhaps we will let ourselves hear the caring that, alert to particulars, is at the heart of morality as well as at the heart of human experience itself.

Notes

[1] The use of words like "learned" falsely anthropomorphize ML. But, talking about ML without using such words is, alas, virtually impossible. So, I have chosen to use the expected words rather than interrupt the reader's concentration with awkward jargon in order to avoid any implication that AI is conscious or has intentions.

[2] Nussbaum in a footnote acknowledges that "The argument of this book [Love's Knowledge] has many connections with" Gilligan's In a Different Voice" and "other related work in feminism. . .". [8] (pp. 42–43, footnote 76).

[3] Not even the values of a person deciding on an ethical issue necessarily generalize well. As Miya Perry writes: "Any consciousness that thinks it has its values fully understood will be surprised by its own behavior in a sufficiently new environment." [9]

[4] (p. 398). [10], cited in [11] (page 398) which points to Rorty's valuing of novels overall because they provide a way to enlarge the circle of those we count as worthy of dignity and respect, and thus serve as a democratic tool.

References

1.   Edwards, P. *A Vast Machine*; MIT Press: Cambridge, MA, USA, 2013.
2.   Gilligan, C. *In a Different Voice: Psychological Theory and Women's Development*; Harvard University Press: Cambridge, MA, USA, 1982.
3.   D'Ignazio, C.; Klein, L. *Data Feminism*; MIT Press: Cambridge, MA, USA, 2020. Available online: https://direct.mit.edu/books/oa-monograph/4660/Data-Feminism (accessed on 26 December 2023).
4.   Gopnik, A. Caregiving in Philosophy, Biology & Political Economy. *Dædalus* **2023**, *152*, 58–69. Available online: https://direct.mit.edu/daed/article/152/1/58/114998/Caregiving-in-Philosophy-Biology-amp-Political (accessed on 26 December 2023).
5.   Held, V. *The Ethics of Care*; Oxford University Press: Oxford, UK, 2006.
6.   Noddings, N. *Caring: A Feminist Approach to Ethics and Moral Education*, 2nd ed.; University of California Press: Berkeley, CA, USA, 1984; pp. 47–49.
7.   Slaughter, A. Care Is a Relationship. *Daedalus* **2023**, *152*, 70–76. [CrossRef]
8.   Nussbaum, M. *Love's Knowledge*; Oxford University Press: Oxford, UK, 1990.
9.   Perry, M. Benevolent AI Is a Bad Idea. Palladium, 10 November 2023. Available online: https://www.palladiummag.com/2023/11/10/benevolent-ai-is-a-bad-idea/ (accessed on 26 December 2023).
10. Rorty, R. Redemption from Egotism: James and Proust as Spiritual Exercises. In *The Rorty Reader*; Voparil, C., Bernstein, R., Eds.; Wiley-Blackwell: Malden, MA, USA, 2010; p. 398.
11. Voparil, C.J. Rorty and the Democratic Power of the Novel. Eurozone, 21 November 2012. Available online: https://www.eurozine.com/rorty-and-the-democratic-power-of-the-novel (accessed on 26 December 2023).
12. Nelson, L. Leveraging the alignment between machine learning and intersectionality: Using word embeddings to measure intersectional experiences of the nineteenth century U.S. South. *Poetics* **2021**, *88*, 101539. [CrossRef]
13. Roy, A.; Horstmann, J.; Ntoutsi, E. Multi-dimensional Discrimination in Law and Machine Learning—A Comparative Overview. In Proceedings of the FAccT '23: The 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023; pp. 89–100.