White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/222157/

Version: Published Version

Article:
Jacobsen, Benjamin (2025) Machine Learning, Synthetic Data, and the Politics of Difference. Theory, Culture and Society. ISSN: 0263-2764

https://doi.org/10.1177/02632764241304687

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

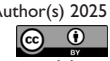
Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Check for updates

Article

Theory Culture & Society

Theory, Culture & Society
1–17
© The Author(s) 2025

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/02632764241304687
journals.sagepub.com/home/tcs
Sage

Machine Learning, Synthetic Data, and the Politics of Difference

Benjamin N. Jacobsen
University of York

Abstract

What is the relationship between ideas of sameness and difference for machine learning and AI? Algorithms are often understood to participate in the continual displacement of the different and heterogeneous in society in favour of sameness, of that which is socio-politically similar and proximate. In contrast to this prevalent emphasis on sameness, however, this paper argues that there is a nascent *heterophilic* logic underpinning the intersection of synthetic data and machine learning, a move towards actively generating differences and heterogeneous data attributes to train, fine-tune, and optimize algorithms. Yet, these synthetic attribute data are nonetheless always machine compatible, devoid of their socio-cultural dynamics and tensions. As such, through a critical examination of three core dimensions of this emergent politics of difference of synthetic data – disentanglement, compositionality, and normativity – the paper argues that this has the potential to ultimately undercut a politics of intervention that seeks to foreground the systemic unfairness and violence of machine learning.

Keywords

algorithms, bias, data, difference, fairness, machine learning, synthetic data

Introduction

In a 2018 interview, CEO of tech company Affectiva and a pioneer of so-called 'Emotional AI', Rana el Kaliouby, was asked about the relationship between bias and AI. She responded that 'It's the data. It's how we're applying this data', which means that 'we need to make sure that the training data is representative of all the different ethnic groups, and that it has gender balance and age balance' (Ford, 2018: 222). Unsurprisingly, issues of bias and representativeness in machine learning and its training data have only become more pressing for computer scientists and researchers since Kaliouby's interview in 2018

Corresponding author: Benjamin N. Jacobsen. Email: benjamin.jacobsen@york.ac.uk
TCS Online Forum: https://www.theoryculturesociety.org/

(e.g. Barocas et al., 2023; Wang et al., 2022). Yet, a recurrent response to such issues has been to build and deploy models that are 'blind to gender, ethnicity and age', that circumvent the issue of representativeness altogether (Drage and Mackereth, 2022: 8). In such cases, Drage and Mackereth (2022) argue, there is a political claim that by not accounting for protected characteristics such as race or gender in input data, machine learning algorithms can be made more representative, removing the grounds for human discrimination and bias.[1]

In contrast to such claims, this paper argues that there is another nascent logic in machine learning – *a logic of heterophily* – towards actively generating socio-cultural differences and heterogeneous attributes as a way to train, fine-tune, and optimize algorithms. This promises to make algorithms less biased, more representative, and more ethical. I take inspiration from Chun (2021), who argues that data analytics and AI systems embody a desire to erase socio-cultural attributes such as race to eradicate longstanding realities of discrimination and inequality in society. The cure to algorithmic biases therefore becomes 'the erasure of all visible markers of difference' (p. 16), crystallized in what Chun calls 'homophily', a conceptual axiom signifying how 'similarity breeds connection'. Homophily implies a fundamental form of violence because it 'assumes and creates segregation' (p. 96). For Chun, therefore, algorithms amplify ideas of sameness and the violent foreclosure of difference, leading to systemic biases and inequalities in wider society.

This paper argues that the emergent intersection of machine learning and synthetic data signals not so much 'the expulsion of the other' (Han, 2018) or the erasure of all visible markers of difference. Instead, computer scientists and machine learning engineers increasingly seek to procure training datasets that are 'representative of all the different ethnic groups' and have 'gender balance and age balance' (Kaliouby cited in Ford, 2018: 222) through the generation of synthetic data attributes on race, gender, and other protected characteristics (Bender et al., 2021; Buolamwini and Gebru, 2019; Crawford and Paglen, 2019). By actively generating and augmenting the different, the diverse, and the underrepresented in algorithms, synthetic data embody a promise to eradicate risks of bias, imbalance, and homogenization in machine learning. But this is a highly problematic promise. As Amoore (2020: 7) argues, the question of ethics does not exist outside of the algorithm precisely because it is 'always already an ethico-political entity by virtue of being immanently formed through the relational attributes of selves and others'. The question this raises, therefore, is not what is an 'ethical use' of algorithms – or how can the algorithm be made 'fairer' – but rather 'how are algorithmic arrangements generating ideas of goodness, transgression, and what society ought to be' (p. 7). Similarly, the logic of heterophily underpinning the intersection of synthetic data and machine learning generates a particular politics of difference, that is, a specific idea of data ethics and representativeness in AI. Yet, it is a claim to representativeness where the subject and the other are nonetheless emptied of their intractable substance and inherent difficulties. The result is models that may be perceived as 'more ethical' or 'more representative' but still do not capture the nuances and variabilities of lived experience.

The paper draws on findings and examples from the emergent field of synthetic data and generative modelling (see also Jacobsen, 2023, 2024; Steinhoff, 2024), exploring how synthetic data attributes are algorithmically generated in order to intervene into the ways in

which algorithms generate new parameters of sameness and differences in society. Synthetic data are used as training data for algorithms, and are most often generated using deep generative models such as large language models, diffusion models, computer graphics pipelines, statistical methods, or a combination of all different approaches (Goodfellow et al., 2016; Nikolenko, 2021; Veselovsky et al., 2023). Synthetic data have become increasingly significant in a number of sectors, including healthcare and government, because they promise to be able to account for data points that are either absent from or underrepresented in the training data, thus making algorithms more representative and placing their outputs beyond the realm of risk (Jacobsen, 2023). While these models rely on the extraction of real-world data to make possible the generation of synthetic data points that approximate the real data distribution (Crawford, 2021; Zuboff, 2019), they are alluring precisely because synthetic data embody a claim to provide a space where more diverse or missing attributes can be generated and incorporated into the training of machine learning algorithms, often as a way of fine-tuning or optimization (Ferrari and McKelvey, 2022).

As such, the logic of heterophily promises to reconfigure this attributive capacity of algorithms. In the computer science literature, attributes are often understood as a quantity that can be further described with a value, where 'the combination of an attribute and a value is a feature' (Abney, 2008: 15; see also Alpaydin, 2010). The attribute is central to how algorithms learn to make differences in society. 'In the ontology of algorithms', Amoore (2020: 169) writes, 'the mechanism that connects people, entities, and events is the attribute.' This means that 'the machine learning algorithm iteratively moves back and forth between the ground truth attributes of a known population' and 'the unknown feature vector that has not yet been encountered', and as such learns to recognize and infer in the world (p. 169). In clustering, for instance, the model learns to find groupings of input data and distributes objects and individuals, whose attributes are more or less similar, into clusters. Machine learning algorithms therefore have an attributive capacity to anticipate and render objects and people knowable in the future based on how they learned from past groupings of data attributes.[2] One crucial question that emerges from this is: what does the attribute make actionable?

The intersection of machine learning and synthetic data constitutes a drive towards generating a wide range of heterogeneous attributes and incorporating these into the training regimes of algorithms to improve generalization. The reason being that adding more data attributes – whether synthetic or real – to the distribution of an algorithm is *productive* for the algorithm. As some machine learning researchers put it, drawing on the example of vehicles, 'how can the learning algorithm know whether the two attributes we are interested in are color and car-versus-truck rather than manufacturer and age?' The answer is: 'having many attributes reduces the burden on the algorithm to guess which single attribute we care about, and allows us to measure similarity between objects in a fine-grained way by comparing many attributes instead of just testing whether one attribute matches' (Goodfellow et al., 2016: 150). Similarly, it has been shown that the addition of synthetic data attributes to the training data set of, say, a skin disease classification algorithm has computational benefits as it increases accuracy rates, lowers error rates, and makes the model more robust to skin images that it has not been previously exposed to (Akrout et al., 2023). But synthetic data also embody a promise to make a normative intervention: to generate and make fine-grained differences matter, to render them attributable, to make

algorithmic models more sensitive to measures of similarity or difference between objects and people. In other words, the logic of heterophily that underpins the enmeshing of machine learning and synthetic data indicates a set of practices and processes towards rendering all differences generatable, attributable, calculable, tractable, and ultimately useful to the algorithm.

In what follows, I foreground three crucial dimensions of the heterophilic logic of machine learning and synthetic data. In the first section – 'Disentanglement' – I outline how data attributes are imagined and generated as radically separable, malleable, and controllable. The second section – 'Compositionality' – discusses what is made possible through the generation of disentangled synthetic data, namely attribute combinations that can be arranged and composed in a plethora of ways so as to produce in algorithms a hypersensitivity to ever more fine-grained differences in the world in order to make algorithms more representative and less biased. The final section – 'Normativity' – examines the normative claim embedded in this emergent conceptualization of difference. It gravitates towards what can be understood as 'the uniformly biased', where the biased, unjust, and imbalanced data distributions deriving from real-world domains can be circumvented or resolved. Lastly, I argue what is at stake in thinking the power and politics of algorithms and synthetic data through the logic of heterophily, and that there is a continuous need to widen the field of critique of machine learning.

Disentanglement

The use of synthetic data for machine learning has become increasingly popular in recent years (Nikolenko, 2021). Algorithmic models are often trained on data distributions that contain an unequal frequency of attributes, such as 'white male' compared to 'black female'. This has implications for ethics because 'notions of fairness often coincide with how underrepresented sensitive attributes are treated by the model' (Hooker, 2021: 3). As a result, a model trained on a large volume of data representing, say, human faces may still fail to recognize faces with rare characteristics (Jacobsen, 2023). It is for this reason synthetic data constitute such a promising development for machine learning: researchers have shown how, in multiple domains, an algorithmic model trained on real data but finetuned on synthetic data that represents rare or sensitive attributes produces a higher accuracy score than a model trained solely on real data. The incorporation of diverse synthetic data is therefore seen as beneficial, constituting a reduction in a model's error rates.[3] Synthetic data constitute an attempt to augment the probability for a model to perceive and recognize a broader range of things more accurately in the world. They embody a promise that different kinds of data (e.g. rare and sensitive attributes) can be generated and incorporated into the training regimes of machine learning algorithms, increasing their capacity to recognize and act in the world.

Yet, this dream to generate a wide variety of data attributes has also transformed the *unit of malleability* of synthetic data. That is, the aim is no longer to generate synthetic training data as such but to reconfigure and tweak its fine-grained features. Algorithmic models can be trained on labelled data or text-image pairs in order to condition the generative process. This affords greater control and specificity in terms of what the model can output. In the case of medical imagining, for example, one can

now input the prompt 'make a photo of a lung X-ray' into a diffusion model and it will generate photorealistic yet synthetic training samples that approximate the distribution of real lung X-ray images (Chambon et al., 2022). In the computer science literature, this conditioning of the generative process is closely related to what has been referred to as *disentanglement.* According to Turing Award recipient Yoshua Bengio (2013: 19), the aim is to develop models capable of learning 'features that are insensitive to variation in the data that are uninformative to the task at hand'. In the case of a model trained on images as inputs:

> An image is composed of the interaction between one or more light sources, the object shapes and the material properties of the various surfaces present in the image. It is important to distinguish between the related but distinct goals of learning invariant features and learning to disentangle explanatory factors. The central difference is the preservation of information. (p. 19)

The aim, in other words, is to build models that learn *disentangled* factors of variation in input data, its core attributes and features. In short, to find out what is really important in the data. 'These factors', Bengio writes, 'interact in a complex web that can complicate AI-related tasks such as object classification. If we could identify and separate out these factors (i.e. disentangle them), we would have almost solved the learning problem' (Bengio, 2013: 19). The underlying assumption behind disentanglement is a particular notion of separability: let an algorithm separate attributes and factors of variations in training data so that these can be subsequently intervened upon. As such, it is a kind of separability that is coupled with a form of generative control. As researchers from DeepMind have put it, disentanglement refers broadly to 'those transformations that change only some properties of the underlying world state, while leaving all other properties invariant' (Higgins et al., 2018: 1). Or as Microsoft researchers state, 'with perfect control, the output image should only change with respect to that attribute' (Kowalski et al., 2020: 307).

The logic of heterophily underpinning the intersection of synthetic data and machine learning resonates with this notion of disentanglement. It signals a move towards generating data attributes that are radically separable, isolatable, addable, malleable, and controllable. It embodies a vision of the social world as an array of attributes and features that are simultaneously separable, independent of each other, and easily malleable and controllable. Take the example of synthetic images generated for the training of facial recognition systems. Computer scientists have stated that 'disentangled face generation has become popular, which can provide the precise control of targeted face properties such as identity, pose, expression, and illumination', which in turn has made possible the systematic exploration of the impacts of facial properties on face recognition (Qiu et al., 2022: 10881). What is being discussed here is the development of a 'controllable face synthesis model' that can be used to generate synthetic faces rarely found in the typical training sets of facial recognition systems. With 'precise control of targeted face properties', machine learning engineers and data scientists are now able to collect 'large-scale face images of non-existing identities without the risk of privacy issues' as well as 'analyzing the influences of different facial attributes (e.g., expression, pose, and illumination)' for the algorithmic model (p. 10881).

Synthetic data are therefore transforming the attributive capacity of machine learning algorithms, making it possible to generate increasingly disentangled and granular differences. This impacts both what the algorithm is able to infer in the future and what social relations are engendered in the present. The emergence of synthetic data and machine learning therefore valorises a particular notion of difference as something that is separable, controllable, and malleable in order to make the algorithmic more representative and less biased. The aim is not only to generate facial training data at scale, but to generate 'different facial attributes', tweaking the expression, pose, or illumination of facial images to see how they variously impact the model. As such, the logic of heterophily opens up unto a future where training datasets are not simply well-curated and relatively fixed entities. Instead, they are increasingly heterogeneous and variable, always in a state of becoming; and the unit of malleability shifts from facial images to more fine-grained attributes and features within such images.

This capacity to generate disentangled data attributes to make algorithms more representative, however, reproduces a particular politics of race. The once popular Israeli-based synthetic data company Datagen, which closed down in 2024, provided customers with access to the company's self-service platform in order to build and customize synthetic humans as training data for their computer vision algorithms. These synthetic humans figure as radically isolatable, malleable, and tweakable parameters. According to their 'API Catalog of Attributes' (Datagen, 2023), human data attributes range from the seemingly mundane – such as eyebrows, hair, glasses, and beards – to protected characteristics such as age, gender, and ethnicity. As an example, the list of ethnicity attributes one could choose from is: 'african, east_african, hispanic, mediterranean, north_european, southeast_asian, and south_asian'. Such attribute data are not only radically isolatable and controllable but also divorced from any biological and embodied determinations. On the one hand, these race specifications reflect deeply racialized and stereotypical representations, showcasing how attempts at disentanglement actually produce new entanglements with real-world stereotypes and entrenched representations. Data attributes that were supposed to be divorced from biology become reattached to very particular and narrow understandings of race.

On the other hand, however, it also imagines race as fundamentally separable from wider societal, structural issues of power. The logic of heterophily constitutes, as Amaro (2019: 84) put it, 'a call to make black technical objects compatible to machine learning artificial intelligence algorithms'. Yet, this ground for racialized compatibility nonetheless 'risks the further reduction of the lived potentiality of black life'. Whilst it resonates with historical ideas of race and control (Gilroy, 2000; Mbembe, 2019),[4] the emergence of the malleable and racialized synthetic attribute results in new racial formations that do not map unto any real bodies but only to a controllable feature within the algorithmic model (Phan and Wark, 2021). As such, the emergence of synthetic data embodies a nascent mode of racialized control that has the potential to further evade the already insufficient safeguards of protected characteristics. It is therefore not a question of *removing* or *eradicating* differences from the training of algorithmic models. Rather, we need to take seriously how the increasing generation of separable and controllable racial and gendered data attributes are promising to transform the space of ethics for machine learning and AI.

Compositionality

It is well known that one of the fundamental features of the learning process of neural network algorithms is that they can exploit what has been called the 'compositional hierarchies' in natural signals – that is, 'where higher-level features are obtained by composing lower-level ones' (LeCun et al., 2015: 439). For images, for instance, 'local combinations of edges form motifs, motifs assemble into parts, and parts form objects' (p. 439). Through combining features in one layer and creating more abstract features in the next layer, neural networks are able to learn many complex non-linear tasks and functions from input data, such as detecting objects and people within the relationships of image pixels. The 2018 Turing Award recipients – Yann LeCun, Yoshua Bengio, and Geoffrey Hinton – state that the benefit of this mode of learning is that 'with the composition of enough such transformations, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations' (LeCun et al., 2015: 436). The power of deep learning, in their view, derives from its compositional capacity: to endlessly combine and re-combine data attributes and features into increasingly complex representations.

Crucially, this idea of compositionality feeds into the politics of difference of machine learning and synthetic data. Here it refers to the ways in which synthetic data attributes are not only malleable and controllable but can also be composed and recomposed in a myriad of ways to create something that is ultimately seen as beneficial to the algorithm. Moreover, the notion of compositionality signals a certain relationship between algorithmic recognition and generated combinations of the different and heterogeneous. At the 2021 NVIDIA GTC Conference, Gal Chechik, Director of AI Research at NVIDIA, presented on the challenges of developing machine learning algorithms that fuse perception with reasoning and decision-making. Speaking on the particular challenge of 'compositional recognition', Chechik (2021) stated that 'what people can do is they can understand new combinations of familiar components, but this is really hard for our current deep [learning] methods'. As he put it, a computer vision algorithm may be able to recognize goats and trees in an image, but 'may fail to recognize goats in a tree'. Similarly, it may recognize red tomatoes but 'will struggle to recognize black tomatoes', given it has not learnt a strong correlation between objects such as tomatoes and attributes such as black. In other words, algorithms find it difficult to recognize unfamiliar combinations of familiar objects or attribute-object pairs that they have not been previously exposed to. The aim, Chechik suggests, is to expose algorithms to unseen and increasingly diverse and heterogeneous combinations of objects and attributes in order to make them more robust to different unseen instances in the world.

What are the implications of this exposure to manufactured heterogeneity for how algorithms generate parameters of sameness and difference in society? The logic of heterophily underpinning synthetic data and machine learning valorises a notion of difference that is disentangled and malleable as well as infinitely recombinable. The fact that models are used to generate disentangled and malleable synthetic attributes makes possible the combination of data attributes into increasingly complex representations – and these representations do not even have to exist in the real world. Crucially, it makes possible a *play of compositionality*, where various synthetic data are generated in order to

create attribute combinations that are previously unseen to the algorithmic model. The ethico-political significance of 'the attribute combination' is well illustrated by work done at the Synthetic Human Lab at Microsoft Cambridge. In 2021, they published work on a model called ConfigNet, the aim of which was to enable the generation of photorealistic synthetic faces and control of various disentangled face attributes such as head pose, expression, and facial hair style (Kowalski et al., 2020). Outlining the learning process of the model, they state:

> ConfigNet learns a factorized latent space, where each part corresponds to a different facial attribute. The first column shows images produced by ConfigNet for certain points in the latent space. The remaining columns show changes to various parts of the latent space vectors, where we can generate attribute combinations outside the distribution of the training set, like children or women with facial hair. (p. 300)

Developing models that learn a factorized latent space representation of the training data is also crucial because 'even when conditional models are trained with detailed labels, they struggle to generalize to out-of-distribution combinations of control parameters such as children with extensive facial hair or young people with grey hair' (Kowalski et al., 2020: 299).

Microsoft Cambridge's ConfigNet model is used to generate different attribute combinations for machine learning models that fall outside of the data distribution. The aim is, again, not to eradicate differences from the model, to make the algorithm 'colour blind', so to speak. Rather, it is to amplify differences for the algorithm through the generation of heterogeneous attribute combinations in order to augment the algorithm's capacity to recognize and infer in the world – even combinations that, at first, appear highly anomalous or perhaps monstrous (such as children with facial hair). For Foucault (2003: 56), the figure of the monster, central to his notion of the abnormal, constitutes a limit point: 'The monster is the limit, both the point at which law is overturned and the exception that is found only in extreme cases.' Foucault argues that the monster is both that which transgresses and that which reinforces the societal boundaries that are transgressed. Similarly, the logic of heterophily promises the endless capacity to generate that which falls outside of the distribution of the training data. But such attribute combinations constitute a form of othering that generates rare and heterogeneous faces and, in turn, reinforces what falls *inside* the distribution, that which is considered normal. This raises the question: what kind of *outside* is the 'outside of the distribution?' The emergence of synthetic data constitutes a move towards the algorithmic generation of differences, rarities, and even the monstrous in the name of making algorithms less biased and more representative. But in so doing it also raises ethico-political questions regarding what is considered normal and abnormal in society.

The question remains: why generate the different, heterogeneous, and monstrous? Microsoft Cambridge's ConfigNet model also helps to unpack this question. As they note in their research paper, the aim of the model is to learn 'a factorized latent space, where each part corresponds to a different facial attribute' (Kowalski et al., 2020: 300). In other words, the model learns a compressed, low-dimensional representation of its training data and, as a result, learns to foreground the salient features and attributes in the

data whilst discarding what it considers irrelevant. In short, the latent space indicates what a model has learned from data (Amoore et al., 2024). The latent space also provides the ground upon which different and fine-grained attribute combinations can be generated. If the model has been trained on a dataset of face images, then by moving between specific points in latent space machine learning engineers are able to change the output of the model, from a generated image of a woman to a man or from a man to a child, from a child with glasses to one without glasses, and so on (Sher, 2021).

The emergence of synthetic data embodies a drive towards the generation of different and heterogeneous attribute combinations for machine learning algorithms. And by incorporating these synthetic attribute combinations into the training regime of the model, it becomes increasingly hypersensitive to differences in new input data. But these synthetic differences are not generated in order to immunise algorithmic models against the real, 'to immunise the actual against the virtual, the probable against the excess of the possible' (Rouvroy, 2018: 100). The logic of heterophily underpinning synthetic data and machine learning is not one of immunisation. As the Microsoft Cambridge researchers have claimed elsewhere, 'training on data with darker skin types leads to a more robust model, perhaps because the task is harder – forcing the model to learn better representations or a more robust attention mechanism' (McDuff et al., 2021: 3746). This is symptomatic of an attitude that algorithms actively benefit from the different, the difficult, the excess of the possible, and the monstrous. The dark skin, for instance, becomes a way to make the task harder, to make the model more robust, and to make the algorithm better at recognizing and inferring different shades of dark skin in the real world.

This means that rather than the erasure of the different and diverse from algorithmic models, the intersection of synthetic data and machine learning is fuelling the generation of increasingly diverse and heterogeneous attribute combinations, some of which may approximate existing social identities and protected characteristics and some of which may not (such as children with facial hair). The aim in either case is to produce algorithms that are increasingly sensitive to fine-grained differences in new data and in the world. Yet, this othering, this mode of generating the different and monstrous, withdraws the ethical obligation to respond to the other. That is because the other is displaced, transformed into a computational problem, a figure of the 'out of distribution'. As such, by making endlessly combinable and recombinable – so there are no fixed reference points, no fixed human bodies – the logic of heterophily opens up a space for those building, tweaking, and deploying algorithmic systems to claim that there is no discrimination, racism, or ageism in their models.

Normativity

The emergence of synthetic data for machine learning constitutes a drive towards the generation of disentangled and malleable attributes, which can be endlessly composed into attribute combinations. This promises to make algorithms hypersensitive to differences in the world. All differences can be rendered controllable and malleable to the algorithm, decoupled from potential association with specific sorts of bodies or phenotypes. It follows that synthetic data also embody *a normative claim to difference* as such. As Aradau and Blanke (2022: 135) suggest, 'the proliferation of difference in machine

learning has enabled new forms of valorization and new political effects'. Similarly, the heterophilic logic of synthetic data and machine learning embodies a claim to what gets to count as different to an algorithmic model, and what it can be made to achieve, computationally and politically. In a 2022 white paper published by Datagen, titled 'Designing a Synthetic Data Solution', they state that when generating synthetic training data for computer vision algorithms it is crucial to 'reflect the task, not the world'. They explain:

> This is a bit counterintuitive because even though the world may be biased, your models shouldn't be. There are biases that can be caused by the gathering methods, things that are naturally less frequent or are harder to gather, appear less in the data. There are biases that can be caused during the annotation process, so things that are harder to annotate have more annotation mistakes. One of the most widely discussed biases, which poses a serious problem for real-world applications, are biases in demographics that are widely dependent on the geography of the gathering process. This is counterintuitive because you don't want to represent the distributions of the real-world. You want to reflect high-level biases of the domain uniformly, in our training data. Ethnicities, ages, genders, lighting scenarios and smartphone camera type are a few examples. (Datagen, 2022)

Central to Datagen's claim is a desire to *go beyond the world*. As the social world is highly biased and imbalanced, any extracted data will necessarily contain these structural biases and imbalances. The assumption here is that any machine learning algorithm can be fine-tuned with an array of diverse synthetic attributes and thus be made more 'ethical'. Or as Datagen put it, it is about developing algorithmic models that are trained on data distributions that are 'uniformly biased'. This notion evokes a data distribution that is balanced in terms of its sensitive attributes so that the algorithm will not showcase disproportionate biases towards any gender, skin colour, or age group (see Jacobsen, 2024). This notion of the uniformly biased algorithm also resonates with ideas of 'colour blindness' and the AI recruitment companies examined by Drage and Mackereth (2022). Yet, rather than the eradication or removal of differences from the training data distribution of the algorithm, the use of synthetic data authorizes claims to algorithmic fairness and justice, made possible through the generation of disentangled, malleable, and combinable data attributes.

On one level, this normative claim to the uniformly biased is unsettling traditional statistical approaches, with their normal distributions, probability estimations, and the identification of regularities in seemingly stochastic processes (Amoore, 2013). On another level, and more worryingly, this claim to difference has the potential to undercut a politics of intervention that seeks to foreground the systemic unfairness and violence of machine learning models. The synthetic data points, attribute combinations, and subjects that are being algorithmically generated are promising to resolve the ethico-politics of algorithms by going beyond 'the distributions of the real-world' (Datagen, 2022). Here, ethical concepts such as fairness and representation are 'narrowed and instrumentalized', made measurable and easily implementable (Hong, 2022: 936). The fundamental promise is to be able to generate whatever racial or gendered attributes are needed for the training and fine-tuning of a machine learning model. Attributes such as race and gender never need to be insufficient, imbalanced, or wholly missing from algorithmic models.

This normative claim is underpinned by a very specific conceptualization of difference (of race, gender, diversity, and representation): it frames synthetic training data as fundamentally disentangled, malleable, controllable, and radically composable. All generated differences are made amenable and compatible with algorithmic models. And as the notion of the 'uniformly biased' suggests, these synthetic differences may be incorporated into the training of an algorithm, may exist in the model's latent space, but their space for ethico-politics is *flattened.* This is not simply a question of reduction. As the example of Datagen shows, the so-called 'high-level biases' that need to be reflected uniformly include 'ethnicities, ages, genders, lighting scenarios and smartphone camera type' (Datagen, 2022). In other words, to an algorithm trained on such data, the politically significant difference between such categories of attributes is flattened, smoothed over, rendered insignificant whilst they still remain different, heterogeneous, and useful to the model. They exist within what Deleuze and Guattari (1987: 371) call a 'smooth space', one that has 'no homogeneity, except between infinitely proximate points'. Ethnicities, like smartphone camera types, naturally co-exist in this space and are just another set of attributes to be added to the model, along with ages and various lighting scenarios. For the model, it is simply a question of becoming sensitive to their proximities and distances, to their useful differences. It is for this reason that, while synthetic data embody a claim to be able to account for difference, heterogeneity, and the other in AI, they instead engender something that is akin to what Zizek (2013: 12) calls 'the Other deprived its Otherness': an idealized other devoid of any substance, stripped of its tensions and frictions. As he writes, 'On today's market, we find a whole series of products deprived of their malignant properties: coffee without caffeine, cream without fat, beer without alcohol' (p. 12). Similarly, the logic of heterophily underpinning synthetic data and machine learning encapsulates a form of othering that lays claim to fairness and justice, but instead it empties out the subject and the other. It is an 'assumption of difference and heterogeneity', Sarah Ahmed (2009: 12) writes, that 'masks the role of structures of authorisation.' Here, there are no intractabilities or frictions, data distributions devoid of any political potential for resistance, refusal, and change. Rather, there is the endless production of synthetic differences that are nonetheless always compatible with algorithmic processing, emptied of any ethico-political substance. It reinforces and depends upon a notion of difference which does not challenge any power structures nor the underlying logics of algorithmic thinking (Fazi, 2021; Parisi, 2019). In this way, synthetic data justify the continuous use of algorithmic models now made 'uniformly biased', 'without prejudice', 'ethical', 'fair'.

Conclusion

This paper has critically examined the intersection of synthetic data and machine learning through the conceptual lens of heterophily. This intersection is characterized by an increasing drive towards generating differences – various and diverse attributes and features – as additional training inputs for algorithms. I also foregrounded three core dimensions of the heterophilic logic of synthetic data and machine learning. Firstly, disentanglement expresses how attributes have become radically separable, malleable, and controllable. Secondly, this results in the endless play of compositionality by which heterogeneous attribute combinations are generated in order to create 'more difficult' and productive training datasets. Lastly,

the drive towards the disentangled, controllable, and radically compositional also embodies a normative vision of the world. In other words, the emergence of synthetic data embodies a promise where controllable and endlessly modifiable and recombinable data attributes can be generated and incorporated into the training of algorithms as a way to bypass their ethico-political limitations and constraints whilst making them (seem) more representative. Synthetic data also constitute a claim to a particular notion of difference where the subject and the other are emptied of their intractable substance.

Whilst these synthetic attributes and their underlying conceptualization of difference are evidently reductive – they do not capture the complex and nuanced variability of the social world – there is still a danger that social science critiques that evoke reduction rely too firmly on what Ramon Amaro (2022: 46) has called 'the problematic of representation'. That is, using the example of Joy Buolamwini's Aspire Mirror project,[5] Amaro observes how issues of bias, risk, harm, and violence in machine learning are too often reduced to a question of racial and gendered representation: a lack of diversity in the training data as well as a lack of diversity of those that build the models. Such critiques are still valuable and necessary – especially given that issues of representation remain entangled with different forms of participatory injustice (see Noble, 2017). Yet, they are by themselves insufficient. 'Coders like Buolamwini', Amaro writes, 'speak directly to the problem of erasure, more specifically the erasure of being, yet the act folds seamlessly into a desire for representation' which is 'devoid of the dynamisms of Black life' (Amaro, 2022: 48). Amaro's work foregrounds the limitations of operating solely within a framework of representation, because it does not fundamentally challenge its underlying sociocultural and structural conditions. Nor does it take into account how machine learning algorithms, in engaging with the world, necessarily transform and generate it. The danger of the representational framework is that it opens up a problem space where all possible critical responses and interventions inevitably gravitate towards the solution vector of either more representation or better representation – neither of which challenges the fundamental politics and power of algorithms. Indeed, the heterophilic logic of machine learning and synthetic data relies on and fuels this problematic of representation. Here, all differences – whether that of race, age, gender, lighting conditions, or smartphone types – can be generated and incorporated into the training or fine-tuning of a model with the aim of making it more representative.

There is also a danger that the intersection of synthetic data and machine learning participates in the steady erosion of what Louise Amoore (2020) has called 'the unattributable'. For Amoore, the unattributable is 'a potentiality that cannot be attributed to a unitary subject' as well as 'a refusal to be governed by the attribute' (p. 171). Yet, she asks towards the end of her book *Cloud Ethics*, 'amid the technologies of the attribute, what remains of that which is unattributable in the scene?' (p. 157). The danger is that this vision of the unattributable becomes increasingly impossible. The algorithmic generation of synthetic differences becomes the 'justificatory scaffolding' (Hong, 2022) for the erosion of the unattributable. Because when everything can be algorithmically generated, everything can be attributed to, accounted for, calculated and processed by the algorithm. It matters not if it is reductive or flattened. Nothing is left out. Where, then, is the space for resistance and refusal? In a society where 'everything must be attributed, even the outliers understood as distant gradients from the curve of normality' (Amoore, 2024: 170), where is space for the emergence of the new, unpredictable, incalculable, and different? For as Amoore has also highlighted, 'a person can flee from genocide, may

seek refuge or claim asylum, but their attributes will be recognisable in advance, before even a claim can be made' (p. 5). This, therefore, is the real risk of synthetic data and machine learning as well as the persistent problematic of representation: it reduces the space of ethico-politics and critique to propositions that algorithms just need to become more accurate, more balanced, and better at representing the diversity of the social world. In short, we just need to algorithmically generate diverse and heterogeneous samples and attributes. Yet, algorithms remain hungry for difference. The heterophilic logic of synthetic data and machine learning embodies a normative vision where all differences are fundamentally generatable, tractable, and compatible with algorithms. A vision where the subject and the other are emptied of all intractabilities, where conditions of the unattributable are increasingly eroded.

There is therefore a need to problematize this new politics of difference generated by synthetic data and machine learning. There is also a need to rethink the notion of difference in relation to contemporary algorithms. To redraw the contours of a social critique that does not rely solely on representation. As Rosi Braidotti (1991: 177) once asked, 'Can we formulate otherness, difference without devaluing it? Can we think of the other not as an other-than, but as a positively other entity?' Part of the answer may be opening up for a more agonistic reading of algorithmic culture (Crawford, 2016: 87), where 'algorithmic decision making is always a contest', where differences are never settled or without friction. Another part of the answer may lie in, as Edward Said (1985: 43) put it, not thinking what difference or representations can do *for* machine learning, but rather tracing 'where its politics can lead'. This is crucial if sites of refusal and competing claims are to persist in our algorithmic societies. Could such a notion of difference maybe emphasize the need of contradiction and incompatibility? Of intractability, of difficulty? Of the need for the unattributable and to not be governed by the power of the attribute? With the emergence of synthetic data, is such a conception of difference still possible?

Acknowledgements

I want to thank Louise Amoore, Ludovico Rella, and Alexander Campolo for many productive discussions on topics such as machine learning, synthetic data, and ethics, which ultimately produced the ideas for this paper. My thanks also go to the anonymous peer reviewers as well as *TCS* board members for their careful engagement and comments.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research has received funding from the European Research Council (ERC) under Horizon 2020, Advanced Investigator Grant ERC-2019-ADG-883107-ALGOSOC.

ORCID iD

Benjamin N. Jacobsen https://orcid.org/0000-0002-6656-8892

Notes

1. While claims to 'color blindness' and 'gender blindness' are not new (Benjamin, 2019; Chun, 2021), they resonate with ongoing debates about the power of algorithms (Amoore, 2020; Beer, 2023; Bucher, 2018) and, more specifically, their exclusionary politics (Eubanks, 2018; Noble, 2017).

2. The ethico-political connotations of the attribute are also echoed in Amoore's (2023: 22) more recent work: 'With the use of "attribute" I am foregrounding the slippage between computational ideas of the properties of a cluster of features in data, and political notions of what can be attributed to a person or to a social group.'
3. In a 2018 paper, for instance, medical researchers found that using a generative adversarial network algorithm to generate synthetic data and adding this data into the training of a neural network classifying liver lesions in CT scans improved the accuracy score of the classifier: from 78.6% accuracy to 85.7% accuracy (Frid-Adar et al., 2018). Similarly, and more recently, Google researchers argued that adding synthetic images to their algorithmic model improved the fairness metrics for sensitive attributes such as 'East Asian Male' from 0.4% in CLIP (industry standard) to 22.8% (Tian et al., 2023). There is, in other words, a slippage between improved computational accuracy rates and a certain ethical project of recognition and diversity.
4. For Achille Mbembe (2019: 71), for instance, racial control is central to Foucault's notion of biopower, a control which 'defines itself in relation to the biological field' and, more specifically, 'presupposes a distribution of human species into groups, a subdivision of the population into subgroups, and the establishment of a biological caesura between these subgroups'.
5. Joy Buolamwini developed the Aspire Mirror project in 2016 at MIT Media Lab. Comprised of facial detection and tracking software, Buolamwini defined it as 'a device that enables you to look at yourself and see a reflection on your face based on what inspires you or what you hope to empathize with'. Yet, when testing the model, Buolamwini found that it failed to recognize her face but successfully detected the faces of her white colleagues. In fact, the model was only able to detect her 'face' while she was wearing a white facial Halloween mask, thus demonstrating underlying racial biases in computer vision systems and their training datasets (see Aspire Mirror website: https://www.aspiremirror.com/).

References

Abney, Steven (2008) *Semisupervised Learning for Computational Linguistics*. Boca Raton, FL: Chapman and Hall.

Ahmed, Sara (2009) *Differences That Matter: Feminist Theory and Postmodernism*. Cambridge: Cambridge University Press.

Akrout, Mohamed, Gyepesi, Balint, Hollo, Peter, et al. (2023) Diffusion-based data augmentation for skin disease classification: Impact across original medical datasets to fully synthetic images. *Deep Generative Models: Third MICCAI Workshop*, Vancouver, Canada, 8 October 2023, pp. 99–109.

Alpaydin, Ethem (2010) *Introduction to Machine Learning*, 2nd edn. Cambridge, MA: MIT Press.

Amaro, Ramon (2019) Artificial intelligence: Warped, colorful forms and their unclear geometries. In: Danae Io and Callum Copley (eds) *Schemas of Uncertainty: Soothsayers and Soft AI*. Amsterdam: PUB/Sandberg Institute, pp. 69–90.

Amaro, Ramon (2022) *The Black Technical Object: On Machine Learning and the Aspiration of Black Being*. London: Sternberg Press.

Amoore, Louise (2013) *The Politics of Possibility: Risk and Security beyond Probability*. Durham, NC: Duke University Press.

Amoore, Louise (2020) *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham, NC: Duke University Press.

Amoore, Louise (2023) Machine learning political orders. *Review of International Studies* 49(1): 20–36.

Amoore, Louise (2024) The deep border. *Political Geography* 109: 102547.

Amoore, Louise, Campolo, Alexander, Jacobsen, Benjamin N., et al. (2024) A world model: On the political logics of generative AI. *Political Geography* 113: 103134.

Aradau, Claudia and Blanke, Tobias (2022) *Algorithmic Reason: The New Government of Self and Other*. Oxford: Oxford University Press.

Barocas, Solon, Hardt, Moritz and Narayanan, Arvind (2023) *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge, MA: MIT Press.

Beer, David (2023) *The Tensions of Algorithmic Thinking: Automation, Intelligence, and the Politics of Knowing*. Bristol: Bristol University Press.

Bender, Emily M., Gebru, Timnit, McMillan-Major, Angelina, et al. (2021) On the dangers of stochastic parrots: Can language models be too big? In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* New York, NY: Association for Computing Machinery, pp. 610–623.

Bengio, Yoshua (2013) Deep learning of representations: Looking forward. *arXiv Preprint arXiv*: 1305.0445.

Benjamin, Ruha (2019) *Race After Technology: Abolitionist Tools for the New Jim Crow Code*. Cambridge: Polity Press.

Braidotti, Rosi (1991) *Patterns of Disonance: A Study of Women and Contemporary Philosophy*. Cambridge: Polity Press.

Bucher, Taina (2018) *If . . . Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.

Buolamwini, Joy and Gebru, Timnit (2019) Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research* 81(1): 1–15.

Chambon, Pierre, Bluethgen, Christian, Langlotz, Curtis P., et al. (2022) Adapting pre-trained vision-language foundational models to medical imaging domains. *arXiv Preprint arXiv*:2210.04133.

Chechik, Gal (2021) Perceive, reason, act. In: *NVIDIA GTC Conference 2021*, virtual event, 9 November 2021.

Chun, Wendy Hui Kyong (2021) *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA: MIT Press.

Crawford, Kate (2016) Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, Technology & Human Values* 41(1): 77–92.

Crawford, Kate (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Crawford, Kate and Paglen, Trevor (2019) Excavating AI: The politics of images in machine learning training sets. Available at: https://excavating.ai/ (accessed 1 December 2024).

Datagen (2022) Designing a synthetic data solution. Available at: https://datagen.tech/ai/designing-a-synthetic-data-solution/. (accessed 12 September 2023)

Datagen (2023) API catalog attributes. *Datagen Documentation*. Available at: https://docs.datagen.tech/en/latest/index.html (accessed 12 September 2023).

Deleuze, Gilles and Guattari, Felix (1987) *A Thousand Plateaus: Capitalism and Schizophrenia*, trans. Brian Massumi. Minneapolis, MN: University of Minnesota Press.

Drage, Eleanor and Mackereth, Kerry (2022) Does AI debias recruitment? Race, gender, and AI's 'eradication of difference'. *Philosophy & Technology* 35: 89.

Eubanks, Virginia (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

Fazi, Beatrice (2021) Introduction: Algorithmic thinking. *Theory, Culture & Society* 38(7–8): 5–11.

Ferrari, Fabian and McKelvey, Fenwick (2022) Hyperproduction: A social theory of deep generative models. *Distinktion: Journal of Social Theory* 24: 338–360.

Ford, Martin (2018) *Architects of AI: The Truth About AI From the People Building It*. Birmingham: Packt Publishing.

Foucault, Michel (2003) *Abnormal: Lectures at the Collège de France 1974–1975*, trans. Graham Burchell. London: Verso.

Frid-Adar, Maayan, Diamant, Idit, Klang, Eyal, et al. (2018) GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 321(10): 321–331.

Gilroy, Paul (2000) *Against Race: Imagining Political Culture Beyond the Color Line*. Cambridge, MA: The Belknap Press of Harvard University Press.

Goodfellow, Ian, Bengio, Yoshua, and Courville, Aaron (2016) *Deep Learning*. Cambridge, MA: MIT Press.

Han, Byung-Chul (2018) *The Expulsion of the Other: Society, Perception and Communication Today*. Cambridge: Polity Press.

Higgins, Irina, Amos, David, Pfau, David, et al. (2018) Towards a definition of disentangled representations. *arXiv Preprint arXiv*:1812.02230.

Hong, Sun-Ha (2022) Prediction as extraction of discretion. In: *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT 2022)*. Seoul, Republic of Korea, 21–24 June, pp. 925–934.

Hooker, Sara (2021) Moving beyond 'algorithmic bias is a data problem'. *Patterns* 2: 1–4.

Jacobsen, Benjamin N. (2023) Machine learning and the politics of synthetic data. *Big Data & Society* 10(1): 1–12.

Jacobsen, Benjamin N. (2024) The logic of the synthetic supplement in algorithmic societies. *Theory, Culture & Society* 41(4): 41–56.

Kowalski, Marek, Garbin, Stephan J., Estellers, Virginia, et al. (2020) CONFIG: Controllable neural face image generation. In: *2020 European Conference on Computer Vision: Online*, 23–28 August, pp. 299–315. Available at: https://www.ecva.net/papers/eccv_2020/papers_ECCV/papers/123560290.pdf (accessed 25 November 2024).

LeCun, Yann, Bengio, Yoshua, and Hinton, Geoffrey (2015) Deep learning. *Nature* 521: 436–444.

Lury, Celia and Day, Sophie (2019) Algorithmic personalization as a mode of individuation. *Theory, Culture & Society* 36(2): 17–37.

Mbembe, Achille (2019) *Necropolitics*. Durham, NC: Duke University Press.

McDuff, Daniel, Liu, Xin, Hernandez, Javier, et al. (2021) Synthetic data for multi-parameter camera-based physiological sensing. In: *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 31 October–4 November 2021, pp. 3742–3748. Available at: https://ubicomplab.cs.washington.edu/pdfs/embc_synthetic.pdf (accessed 25 November 2024).

Nikolenko, Sergey I. (2021) *Synthetic Data for Deep Learning*. Cham: Springer International.

Noble, Safiya Umoja (2017) *Algorithmics of Oppression: How Search Engines Reinforce Racism*. New York, NY: NYU Press.

Parisi, Luciana (2019) Critical computation: Digital automata and general artificial thinking. *Theory, Culture & Society*: 36(2): 89–121.

Phan, Thao and Wark, Scott (2021) Racial formations as data formations. *Big Data & Society* 8(2). Available at: https://journals.sagepub.com/doi/epub/10.1177/20539517211046377 (accessed 25 November 2024).

Qiu, Haibo, Yu, Baosheng, Gong, Dihong, et al. (2022) SynFace: Face recognition with synthetic data. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision 2021*, Montreal, Canada, pp. 10880–10890.

Rouvroy, Antoinette (2018) Governing without norms: Algorithmic governmentality. *Psychoanalytical Notebooks* 32: 99–102.

Said, Edward (1985) An ideology of difference. *Critical Inquiry* 12(1): 38–58.
Sher, Varshita (2021) Keywords to know before you start reading papers on GANs. *Towards Data Science*. Available at: https://towardsdatascience.com/keywords-to-know-before-you-start-reading-papers-on-gans-8a08a665b40c (accessed 3 December 2024).
Steinhoff, James (2024) Toward a political economy of synthetic data: A data-intensive capitalism that is not a surveillance capitalism? *New Media & Society* 26(6): 3290–3306.
Tian, Yonglong, Fan, Lijie, Isola, Phillip, et al. (2023) StableRep: Synthetic images from text-to-image models make strong visual representation learners. In: *37th Conference on Neural Information Processing Systems (NeurIPS 2023)*, New Orleans, USA. pp, 1–21.
Veselovsky, Veniamin, Ribeiro, Manoel Horta, Arora, Akhil, et al. (2023) Generating faithful synthetic data with large language models: A case study in computational social science. *ArXiv* 1–8.
Wang, Angelina, Ramaswamy, Vikram V. and Russakovsky, Olga (2022) Towards intersectionality in machine learning: Including more identities, handling underrepresentation, and performing evaluation. *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, 21–24, June 2022, Seoul, Republic of Korea, pp. 1–14.
Wood, Erroll and Baltrusaitis, Tadas (2021) Synthetic data with digital humans. Webinar 'Implementing Data-Centric Methodology with Synthetic Data', *Datagen*. Observed 17 November 2021.
Zizek, Slavoj (2013) *Welcome to the Desert of the Real: Five Essays on September 11 and Related Dates*. London: Verso.
Zuboff, Shoshanna (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Benjamin N. Jacobsen is a Lecturer in Sociology at the University of York as well as a Visiting Fellow on Professor Louise Amoore's 'Algorithmic Societies' project at Durham University. His current research explores the ethico-political implications of generative modelling and synthetic data on society and culture.