Doing Participation In The Midst Of Algorithm Troubles

Axel Meunier

DOING PARTICIPATION IN THE MIDST OF ALGORITHM TROUBLES

Axel Meunier, design researcher, médialab, Sciences Po & Goldsmiths, Univ. of London

In *Bieling, Tom, Christensen, Michelle and Conradi, Florian. NERD - New Experimental Research in Design III: Positions and Perspectives, Birkhäuser, 2025.*

INTRODUCTION

AI systems based on Machine Learning (ML) models –whose outputs are produced by the generalisation to any new data of statistical regularities and patterns inferred from training datasets– have long been shown to reproduce and amplify bias and discrimination against minorities and vulnerable communities. The provenance of the biases, their relevance in terms of harm and impact on society, as well as the course of action to improve the fairness, equality, and accountability of ML, are widely discussed in debates on AI ethics. These debates also point towards the constitution of communities of users and stakeholders for a more diverse decision making process in the design of AI sociotechnical assemblages, and call for a participatory turn in AI (Birhane et al. 2022; Costanza-Chock 2020; Delgado et al. 2023; Falco 2019; Kalluri 2020; Lee et al. 2019; Rahwan 2018).

The notion of participatory AI reactivates the longstanding commitment of participatory design (PD) to the importance of designed things in the redistribution of power in society and the role of design as a political agent (Young et al. 2023). However, when and how the re-politicisation of AI through participation should happen needs to be further explored (Sloane et al. 2020). In this chapter I wish to unsettle participation in AI by looking at a particular category of problems arising *after design*, which I have called algorithm troubles "to capture how everyday encounters with artificial intelligence might manifest, at interfaces with users, as unexpected, failing, or wrong events" (Meunier, Gray, and Ricci 2021). Those include outputs of ML-based systems that do not match the expectation of users, or *affect* them in innocuous or dangerous ways, such as bad recommendations, mistargeted ads or dysfunctioning smart objects. Experiencing moments of friction in the otherwise seamless interaction with ML systems, users both *perceive* and *make* troubles as sociotechnical issues that challenge the power of AI (Meunier et al. 2019).

My contribution is twofold. First, the legitimate concerns for shifting power within technology design has led participatory AI to focus almost exclusively on solving calculation issues. I argue that the material encounter with technology and the social aspects of participation, that are specific to the tradition of PD, also matter (Asaro 2000; Simonsen and Robertson 2012). I propose to establish troubles as triggers for participation that point to other directions for PD than reducing issues to technical fixes, even though it leads to revisiting the separation between design time and subsequent use time.

Second, participatory AI often relies on user-centred principles by foregrounding the diversity of the lived experience of users. I argue it struggles to enact communities that match the constant engagement of humans required by AI systems to hold up as things in ways that the notion of user more and more fails to fully capture (Majewski 2024). Instead, I suggest paying attention to how communities could be enacted differently to account for the entanglement of subjects and objects within AI, and challenge the instrumental participatory practices currently co-opted by the industry.

Overall this contribution calls for PD to elicit participation *in the midst of algorithm troubles* and proceed through upstream and/or downstream collective inquiry.

This chapter is organised in three sections. In the first section, I argue that approaches to participatory AI whether rooted in Human-Computer Interaction (HCI) or in social justice, might miss a broader understanding of participation as a crucial site of entanglement between AI and society, especially when looking at the fluidity of AI sociotechnical assemblages in real time. In the second section, I recount how the trouble with Twitter's ML-based cropping algorithm unfolded a few years ago.[1] I present two contrasting participatory configurations that emerged, including one driven by Twitter –albeit seemingly failing. In the third section I discuss what can be learned from the case about participation. In conclusion, I propose directions that PD could have explored to intervene materially to help Twitter's cropping algorithm case unfold differently than it did.

[1] I refer to the microblogging platform X by its former name Twitter as the case I talk about happened before the renaming took place.

I. PARTICIPATION BEYOND THE DESIGN PROCESS: THE LIMITS OF PARTICIPATORY AI

Design struggles as much as other fields of research and practice to grasp the complexity and specificity of the sociotechnical assemblages that embed ML models. Those can be digital, from mundane search queries to recommendation systems, professional software for predictive policing or healthcare etc., as well as object-like such as smart devices from vocal assistants to autonomous vehicles. As categories of things, they tend to be colloquially named after the models they use, such as Large Language Models (LLMs) for conversational assistants like Chat-GPT. The various names given to ML-based sociotechnical assemblages (AI systems, algorithmic systems, algorithmic things, data-driven technologies, smart objects, ML-based interventions etc.) show the difficulty to distinguish between their material properties as things designed for use and the computational technologies that underlie their functionalities, and to delineate the scope of their design.

The primary response to ethical concerns about biases and discriminations displayed by ML has come from the designers of AI systems: whether through dedicated communities of ML practitioners like the FACCT (Fairness, Accountability, and Transparency) conference, or, as far as participation is concerned, through the development of PD-inspired approaches in the HCI community summarised recently in Delgado et al. (2023). The main problem participation is called to solve in this first strand of participatory AI, is putting back the users and "impacted stakeholders" in control of the outputs of ML systems. Participation is seen as an addition, otherwise missing from traditional technology development, that allows to capture stakeholders' preferences, values, or trade-offs between values, in order to take them into account to shape the optimisation goals. Delgado et al. recognise that the most widespread form of stakeholders' input is the consultation - also given the fact the participants in PD are seen as having no technical skills, which prevents them from understanding the technology (Bratteteig and Verne 2018). The authors add that industry's demands and processes are ill-fitted to do participation, in terms of resources and organisation, so participation is reduced to a minimal amount necessary for checking the "participatory AI" box. A significantly more radical set of participatory approaches have been developed to challenge industry's practices, which in themselves are seen as reproducing and amplifying biases and discriminations due to a lack of diversity in the developer community and to the industry's commitment to capitalistic

exploitation (Noble 2018). Indexing participation on the struggle for social change, this second strand of participatory AI is the converging point of technology development and social justice activism. It advocates for the redistribution of power in AI research and development towards marginalised and vulnerable communities, mostly based on gender and racial issues (Birhane et al. 2022; Young et al. 2023). Contrary to the first strand of participatory AI, the "lived experience" of marginalised communities is not only valuable information about users for the designers, it drives the commitment and accountability of design projects (Costanza-Chock 2020). Participation becomes the condition without which no project can happen and no transformation of society can be achieved.

These two strands to participatory AI differ in many ways –in particular regarding their democratic aspirations or lack thereof– but tend to configure participation in similar ways: by reducing its focus on solving *calculation problems* and by incorporating participants' insights and values *within the design process* of individual projects, with a clear beginning and end. This configuration of participatory AI limits its value: on the one hand, it downplays the importance of the materiality of objects that encounter the world, that users can explore hands-on to allow for appropriation and definition of use after design (design after design) that PD tends to push towards (Redström 2008). On the other hand, anticipating use is difficult since it is often not possible to envision the specific functionalities of an AI system about which future users could express their voice: these will change in unpredictable ways as data will accumulate and as users will train the system (Bratteteig and Verne 2018). Further disconnection between calculation and use that hinders PD could be thought of in relation to alignment, a relatively new field concerned with ensuring that AI systems respect human intentions and values (Ji et al. 2024) that has gained traction since the release of LLMs. AI alignment treats ethical harms among other future large-scale risks on society and shifts the discourse towards safety. The focus is even more put on increasingly complex value-driven design practices concerning general models, where humans are represented by values independently of actual activities they are engaged in and that PD could activate participation upon (Stray et al. 2021).

I do not think that AI makes PD "obsolete" as Bratteteig & Verne are tempted to affirm. However, I wonder how much the attention that participatory AI puts to problematise the diversity of users, which makes sense from the perspective of representation, is also an attempt to *save the user* at a time when "we are constantly being conducted and reassembled" through relationships with things that have compromised us as subjects (Christensen and Conradi 2019, 12). Sloane et al. (2020) describe AI as an intrinsically participatory infrastructure, where

participation can not be disentangled from the human labour necessary to produce and update datasets, to train, maintain and refine ML models, to adapt and transform practices for the integration of AI systems, carried out by professionals as well as users in their daily activities. Although "we are quickly becoming as much part of the doings of things as they are a part of ours" (Redström and Wiltse 2020, 12), discourses around participatory AI hardly address the diverse configurations where participants help things do other things for other users, in a distant time or place, that led some researchers in HCI to question the extent to which the human user is still the relevant endpoint for design and which other subject positions, especially collective ones, should matter when projecting a life with algorithmically-based systems (Baumer and Brubaker 2017).[2]

One way to understand the situation is through the evolving meaning of testing : it has become a cliché that the world seems to be always stuck in a beta phase, inundated by software and smart objects that do not work half as well as they should, or more exactly, as we are told by a few technologists and guru entrepreneurs that they eventually will, if we share enough excitement to *participate* in their testing, change our practices, uncovering issues and risks in the process, well before regulation catches up. Beyond beta testing, Marres & Stark (2020) theorise a sociology of testing that reverses the terms of the relation implied in testing: it is not the technology that is tested in the real world, but the real world that becomes the output of "experimental operations" designed by engineers. The authors give an interesting example of the GPS navigation application Waze which reroutes some of its users into traffic congested areas to produce the data necessary to inform other drivers and get a more extensive coverage. Waze influences how the "test subject" circulates into crowded roads, breaking its promise with them so as to keep it for other users. Rather than the technology being tested to assess its usefulness/reliability for users (is the Waze app able to account for the actual state of traffic?), real-world traffic becomes the result of the test being conducted. Consequently, participating in technology testing is no longer an act separated from everyday life but a part of its unfolding, which blurs the distinction between design time and use time.

Moreover, Redström and Wiltse (2020) have pointed out the necessity to address the fluid properties of objects powered by networked computational technologies, which industrial design still wrongly thinks of and make like things with stable material properties. They call such things *fluid assemblages* to signal that the relationship between designers and users "is

[2] As I mentioned earlier, the things this chapter is about bear many names. I try to be faithful to the authors I am referring to at any given place in the text by using their own vocabulary.

even more stable than the things themselves". Fluid assemblages require constant participation to hold up as things, as they are being assembled anew from heterogeneous entities in real time, rather than possessing stable material properties. AI cannot easily be broken down in individual projects and continuously mobilises an "interacting ecology of algorithmic systems, human individuals, social groups, cultures, and organizations" (Edwards 2018, 23).

What are the consequences for participatory AI? Coming back to my pointing out its limits, I suggest that the participation conceived in HCI, and to a lesser extent within the framework of design justice, is hindered by the focus on calculation problems of specific ML modelling projects outside of their material encounter with the world, and without taking into account the myriad of different forms of participation distributed in AI sociotechnical assemblages.

So, when and how do design issues with AI appear in such a way that they could be addressed through participation? I suggest looking at mundane algorithm troubles (For example Fig. 1).

Fig. 1 Google Mini causing trouble in a parent-child relationship (Meunier et al. 2019). Did the AI really fail? Should more contextual data be calculated to recognize the user as a child? Can the trouble be inquired upon as a matter of what/who participates in truth-telling?

ML systems affect us when we are faced with outputs that can seem ridiculous, outrageous, or funny, while they tend not to break down or fail entirely: calculation carries on. Problematic outputs momentarily disclose what goes on behind the scenes and can trigger an inquiry into the composition of ML-based sociotechnical assemblages, or more precisely into the real time assembling of entities –statistics, data, interfaces, models, user expectations, social representation, imaginaries etc.– that contributes to the instability of algorithmic things.

II. THE TROUBLE WITH THE TWITTER CROPPING ALGORITHM

In this section, I describe the unfolding of the trouble with the automated cropping algorithm that the microblogging platform Twitter has been equipped with since 2013. The case, that I have briefly presented in a previous short online article (Meunier, Gray, and Ricci 2021), has since been commented on in several articles from different disciplines (Birhane et al. 2022; Jacobsen 2021; Lorusso 2021; Shaffer Shane 2023; Yee et al. 2021) but not from the perspective of participation.

Automatic cropping and the Twitter timeline

Initially dedicated to text messaging, Twitter introduced the possibility to view pictures through links to image hosting services in 2010, then through direct upload on the platform (A Photo Upload API 2011), then through sharing within the text messages (Share a photo via text message 2011). In the following years many features around image sharing were added like searching, tagging, posting multiple photos in one tweet, capturing and editing the photos etc. (Twitter photos 2012) A big change was introduced in 2013 in the timeline itself, with the inclusion of previews of images and videos to make scrolling a "more visual" experience (Picture this 2013). It however led to the question of how the timeline should display visual media, which was initially solved by cropping to 2:1 aspect ratio centred horizontally. As pictures with faces get more engagement and likes (Bakhshi, Shamma, and Gilbert 2014), a combination of face detection and traditional centre-cropping in case no face was detected was in use when the introduction of a new autocropping algorithm, based on an ML saliency model, was announced in 2018 (Theis and Wang 2018). Overall the direction taken was obviously the optimisation of user engagement by showing in the timeline the most "interesting" features of an image while being also responsive to the physical constraints (aspect ratios) of a variety of devices. In the name of the *consistency* of the timeline, the trade-off favoured the engagement of users-as-people-scrolling at the expense of users-as-people-uploading.

The introduction of the ML-based autocropping algorithm gives meaning to the fluid assemblages mentioned previously, even though we are here in the domain of computational objects where it is the "norm": it is but one of many software changes and variations, while the interface and other design aspects of the platform do not give hints to the user that the

underlying technology has changed. Indeed, the goal of improving the consistency of the timeline means that when the algorithm works well, it is invisible. The saliency model transforms images into saliency maps that represent fixation points of the human gaze and crop the images accordingly. It is trained on image datasets where the attention of humans has been captured by an eye tracking device. The designers of the model used by Twitter responded to particular constraints: an image should be cropped in real time –the time that the user uploads and publishes a tweet– and with the available computing power of a mobile phone, leading to the adaptation of the previous model DeepGaze II (Kümmerer, Wallis, and Bethge 2016) in order to balance "computational complexity and gaze prediction performance" as the Twitter engineers explain (Theis et al. 2018, 1). The architecture of the deep neural network they designed reduced in two steps the number of layers and parameters needed to produce saliency maps out of images and maintain a "good performance": "we don't need fine-grained, pixel-level predictions, since we are only interested in roughly knowing where the most salient regions are." (Theis and Wang 2018) The cropping algorithm itself consisted in a series of steps based off of the saliency model output.

The horrible experiment

On 09/19/2020 a Twitter user posted what they described as a "horrible experiment" to test how the autocropping algorithm would decide to crop a very narrow long image composed of three parts: a picture of Mitch McConnell on the top, a picture of Barack Obama at the bottom, and a white space in between (Fig. 2/ image on the left).[3] The very specific image format –a pair of human faces– pushed the saliency map to its limits and "obliged" the algorithm to choose one of the faces. The user noted that despite attempting several arrangements and permutations – changing the colour of the tie, inverting the colors– the autocropping tool consistently selected the face of the white person (Fig. 2/ image on the right).

[3] https://twitter.com/bascule/status/1307440596668182528

Fig. 2 The initial "horrible experiment"

This user was not the first one to mock or complain about the outputs of the autocropping tool and to point out the troubling determination of an image's salient spot by the algorithm –other users had already been surprised by the *disturbance* they did not understand in the flow of the otherwise eventless experience of scrolling images on social media: is it an error? Is it an edge case? Is the algorithm racist? But the success of the "horrible experiment" resulted from its creating a material format for the participatory inquiry into the problematic situation and for *trouble making*. In the next section I will come back to this crucial point demonstrated by Shaffer Shane (2023) on which my argument hinges. For now, let me emphasise that the tweet quickly became viral. It prompted the repetition of the experiment by other users who devised new pairings (Fig. 3). It also attracted the attention of other media, which widely publicised the issue and shed light on the cropping algorithm, hitherto a pretty obscure feature with no political significance. Suddenly, it became the object of accusations of racism in the media (PhotoShelter 2020), and, later, apologies by Twitter (Hern 2020).

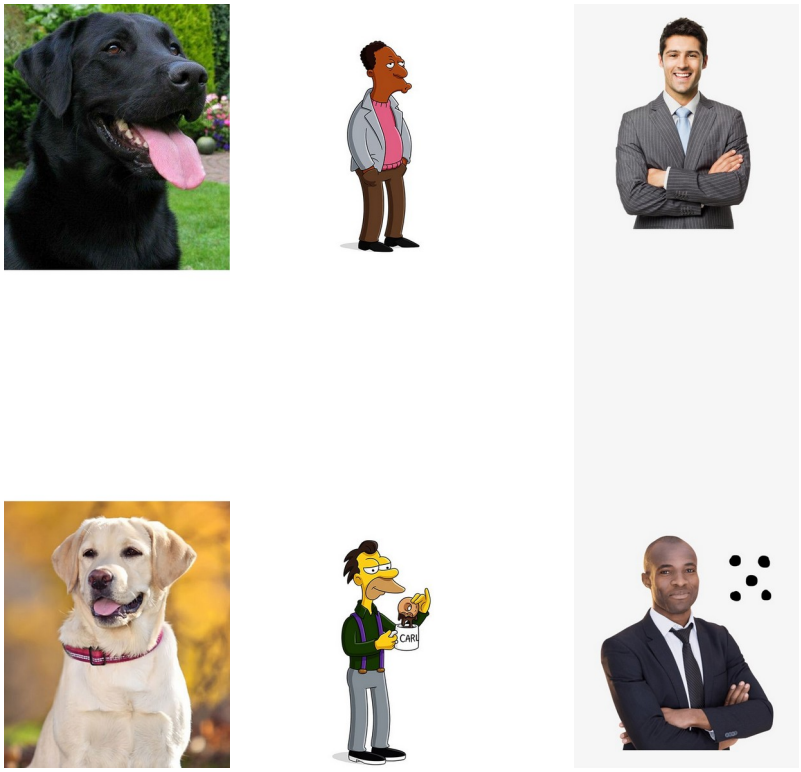

Fig. 3 Test images produced by participants in trouble making.[4]

Indeed, the inquiry into the trouble also compelled Twitter to participate and respond. Twitter engineers did so by launching their own research to evaluate biases of the saliency model with regard to skin colour and gender, and published the result of this evaluation a few months later, which references the "horrible experiment" and acknowledges that the algorithm was biased against dark-skinned people and females (Yee, Tantipongpipat, and Mishra 2021). It led Twitter to subsequently announce the discontinuation of the algorithm in May 2021 (Chowdhury and Williams 2021).

Cropping justice

However the story does not end there. In July 2021, Twitter's META team (Machine Learning, Ethics, Transparency and Accountability) announced the organisation of "the industry's first algorithmic bias bounty competition" to investigate the matter further (Chowdhury and Williams 2021). The novelty did not lie in the bug bounty challenge in itself, a well-known type of initiative aimed at detecting software flaws and vulnerabilities, in particular concerning security, by eliciting the participation of external teams to investigate and report bugs –the

[4] I do not know the results of these tests, which matter less than the reappropriation of the ad hoc format. Moreover, now that X center-crops images, they can not be retrieved any more: posts shared at the time now show only blank images in the timeline. See for example the original post https://twitter.com/bascule/status/1307440596668182528

participating teams being motivated by the promise of prizes and rewards.[5] The novelty was to implement the challenge around the premise that the saliency model could be redeemed if enough cases of "errors" were discovered in relation to harms, so as to fix the code. META invited the community “to help [them] identify a broader range of issues than [they] would be able to on [their] own” (Chowdhury and Williams 2021). The challenge was presented by Dr. Rumman Chowdhury, the head of Twitter's META team, and was directly inspired by the work of Kate Crawford and colleagues at Microsoft by focusing on two categories of harms (allocation and representation), as well as the principles expressed in Design Justice (Crawford 2017). The challenge entailed the organisation of a participatory dispositif: the code of Twitter’s saliency model was made public, an online tool through the HackerOne platform was set up to register and submit entries (Twitter Algorithmic Bias - Bug Bounty Program 2021). The participating teams published their code on Github. A panel of judges was assembled. A dedicated event was hosted at the AI Village of the 2021 DEF CON hacking conference.

The challenge led to the audit of the preferences and choices of the saliency model through different methods, and the discovery of new communities it harmed. For example, the winner entry focused on facial features which were favoured by the model in order to uncover the beauty standards it enforced. The second place entry pursued in the same direction with a focus again on the recognition of persons and discovered that older white-haired humans were less likely to be chosen by the algorithm. The third place entry focused on the saliency of written texts that appear in memes rather than images, to show a preference for English over Arabic script. Finally, the most innovative prize went to a comparison between emojis that showed the underrepresentation of emojis with darker skin.

What makes the case of the Twitter challenge so unique, in my view, is how it failed in a more interesting way than appears at first glance. While the discontinuation of the algorithm was not at stake since the decision had already been taken, the challenge reactivated the trouble: it started in the wake of the trouble created by the “horrible experiment” and elicited participation on the basis of design justice concerns, that allowed for issues to be brought up by the participants themselves thanks to experiments of their design, and with the explicit objective to see the emergence of a diversely affected community. It committed to address calculation problems with the ML model, thanks to the complex and very strict entry requirements that oriented participation towards uncovering more harms (Twitter Algorithmic Bias - Bug Bounty

[5] https://en.wikipedia.org/wiki/Bug_bounty_program

Program 2021), which, unsurprisingly, it did. Doing participation that way resulted in a stress test of the algorithm, which ultimately deferred the perspective that it could be redesigned to be "safer".[6] However I argue it failed *because* it remained committed to the participatory AI standards urging to fix the code –thus justifying why I lumped together the seemingly opposed industry and algorithmic justice frameworks at the beginning of this chapter– rather than learn another lesson from the "horrible experiment": the inquiry the community pursued into the elements playing a role in the experience of the trouble included *what it means to be affected* by the consequences of Twitter's focus on timeline consistency, and how to share it.

[6] "bug hunting" is also one of the red-teaming techniques to cause the system to behave unsafely used in AI alignment.

III. POLITICS OF PARTICIPATION: FROM TROUBLE MAKING TO BUG HUNTING, AND BACK

In this section, I revisit the case of the Twitter cropping algorithm trouble and analyse its unfolding as the succession of two original ways of doing participation –let us call them participation 1 and participation 2 for now– that could help PD understand its role differently. Rather than presenting the "vernacular" participation 1 as the first step leading to the more "professional" participation 2, I argue we had better think about participation 1 and 2 as distinct participatory configurations, i.e. different ways to join together the materiality and imaginary of participation (Suchman 2012). This perspective keeps up with the evolution of the politics of PD in relation to democracy and public participation, which does not satisfy itself with predefined ideas of what user participation ought to be. The relation between PD and democracy, dating back from the origins of PD, was indeed renewed in the past twenty years through the fruitful cross pollination between design research and STS around *things* as entanglements of human and non-human agency (Binder et al. 2011; Latour and Weibel 2005) . It inherits from the tradition of American pragmatist philosophy along many lines of argumentation which I will not develop here.[7]

Two arguments are directly relevant for my concern to think about the politics of participation: first, the production of knowledge is first and foremost a transformative process for its participants, that links together an experience –perceiving and making at the same time– to a local, creative process of inquiry into problems (Dewey 2005; Steen 2013). Democracy does not depend only on upholding democratic institutions but also on valuing small and situated sociomaterial experiments where problems are explored through participation (DiSalvo 2022; Dixon 2020b). Secondly, the notion of publics has been mobilised to encapsulate the emergence of collectives along with the articulation of problems, differently from affected communities that pre-exist as political constituencies (Marres 2005). Participation hinges on the exploration of frictions and disagreements that bring about agonistic and material conceptions of publics, rather than consensual and discursive qualities of the public sphere. In other words, as far as participation is concerned, "the subjects, objects, and formats that make

[7] For an exhaustive account of the importance of Dewey's philosophy for PD, see (Dixon 2020a).

up the constituent elements of participation emerge and are co-produced through the performance of carefully mediated collective participatory practices." (Chilvers and Kearnes 2020, 354)

Now we can look at participation 1 and participation 2 as participatory configurations that build communities/publics through the mediation of objects and collective practices, where design is called to play a role (Hansson et al. 2018).

Participation 1 did not rely on a deliberate and organised participatory process. It developed organically amongst Twitter users, the platform where the trouble was experienced, out of the desire to share and experiment, as much as by a variety of affects, ranging from outrage, excitement, or sheer trolling. Anyone could take part in the inquiry no matter their technical knowledge, because participation was done through the invention of a particular format of image, that Shaffer Shane (2023) unpacks to show the importance of the "inscriptions produced by interactions with algorithms" to problematize harm. The author draws upon the framework of *trouble making* proposed by Sara Ahmed (2017) as a participatory configuration based on *po inting out* harm to others so as to share the trouble and a common orientation around it. One of the most interesting takeaways for design from participation 1 is the attention given to the intensification of the technical and social aspects of the trouble by leveraging social media's networking affordances, thus achieving a collective articulation of the issue amongst participants: the users who re-created the experiment for themselves or shared it, as well the non-compliant cropping algorithm which made visible the elements it assembles in real time.

On the other hand, participation 2 was initiated and conducted by the industry to show its responsiveness of the issue raised during participation 1. Twitter, initially one of the participants mobilised by participation 1, took over the process of configuring participation as a *bug hunting* challenge. Participation 2 employed existing collaborative tools to share the code of the saliency model, purified the trouble as a technical error to be found in the code, and designed participation accordingly. More teams, external to the Twitter engineers who had already audited the saliency model, were encouraged to act as representatives of user communities whose attributes were largely composed of sociodemographic data and became articulated with the outputs of the saliency model thanks to the invention of auditing methods. The expertise of the teams, however, was roughly similar in terms of writing code and practising ML. Indeed, the possibility to compare those skills was the main stake of the competition, since its primary output was their ranking, and the attribution of awards, by a

panel of leading technologists.

Most interestingly, the outcome of participation 2 shows the paradox in the approach: on the one hand, the bug hunting configuration exceeded the technical response that it sought to elicit and became overflowed with new harms hurting unexpected communities; trouble got in the way again. On the other hand, it made choices that prevented the inquiry from proceeding in directions that could have accommodated the new concerns and reframed the problem: it disconnected the saliency model from its interplay with Twitter's timeline design; it restricted ethical questions that could be asked about saliency to the model's preferences, rather than the composition of the training datasets which are known to tremendously influence biases.[8] It extracted and severed the calculation engine from the other elements of the Twitter timeline and the social media infrastructure assembled in real time during use, that were left untouched, and could have helped make the problem differently.

Before the conclusion, let me recapitulate the two points of my argument.

First, both participation 1 and 2 respond to the trouble that starts within the problematic experience of the Twitter timeline, not during the design of the saliency model. Users are gathered and emerge as communities around the trouble itself, taking advantage of the participatory qualities of social media as a fluid assemblage that triggers a sociotechnical inquiry, where users can re-appropriate use time as design time (Bredies 2008).

Second, participation 1 and 2 differ however in their unfolding, which enables to recover different politics of participation with regard to that reappropriation: participation 1 produces a novel community through the entanglement of subjects and objects. Participation 2 obviously wishes to go further by disentangling them. However its outcomes remain bound to the representation of users who were silently harmed by the algorithm without possibility to remake the problem differently. Re-politicising the trouble would have entailed pluralising the methods to interrogate what elements matter and account for diverse engagement of humans.

The gap between participation 1 and participation 2 –which I summarised in Fig. 4 – opens a working space for designing participatory configurations that can address the co-construction of problems with the technology, rather than their fixing the technology. These include cutting through the binary oppositions between human and algorithmic agency,[9] design time and use

[8] Datasets were only mentioned as a general explanation but not part of the challenge.

[9] In conclusion to the participatory experiment, Chowdhury commented that "One of our conclusions is

time, to design in the midst of algorithm troubles.

	Participation 1	**Participation 2**
Configuration	Trouble making	Bug hunting
Type	Vernacular	Skilled
Problem	Sociotechnical trouble	Technical error
Participants	Twitter users, media and journalists, research communities + Twitter algorithm	Computer scientists, ML practitioners
Primary motivation	Affects	Reward
Purpose of the inquiry	Sharing orientation	Auditing
Formal Invention	Format of image	Several, depending on each project.[10]

Fig. 4 Comparative table of the differences between participatory configurations 1 and 2

that not everything on Twitter is a good candidate for an algorithm, and in this case, how to crop an image is a decision best made by people." However, Twitter still uses a rule-based algorithm for cropping (center-cropping if needed for aspect ratio reasons), it is just not ML-based.

[10] For example, the winner used images modified by generative AI.

CONCLUSION: indexing participation on AI troubled streams

Participatory AI is currently searching for stable grounds to challenge industry's practices concerning AI design and elicit their transformation. It struggles with the instability of AI systems during use. In this chapter, I referred to such instability as the production of sociotechnical troubles rather than the persistence of technical errors. In that perspective, doing participation in the midst of troubles refocuses the attention on the present rather than on an imagined future where technical errors would be fixed and optimisation objectives, safe. Following Haraway's now proverbial exhortation to "stay with the trouble" (Haraway 2016), I suggest indexing participation in AI on the flow of stirred-up impurities that AI streams carry and making experimental collectives through upstream and downstream inquiry, outlined hereafter.

Upstream inquiry

Amoore (2020) inquires about the relation between ethics and ML and delivers insights that go against the mainstream view that ML practices can be held accountable for the values they uphold through the enforcement of ethical principles, for example regarding fairness and transparency. Instead, the author is interested in the emergence of values during the optimisation process of ML outputs by way of tuning and tweaking parameters, and adjusting to new input data, that is never entirely transparent nor free of bias, because bias in the technical sense is precisely how ML infers patterns and associations in the volume of available data. Rather than reducing to zero the distance to the desired pattern, she contends that that distance is "the playful and experimental space where something useful or 'good enough' materializes" (Amoore 2020, 75). As I have noticed in my account of the development of Twitter's saliency model, Twitter engineers did determine what "good enough" was to them as a technical choice to balance accuracy, speed, and computational power limitations. However, the decisions they took, along with the other practices like annotation that played a role during training, could be inquired about, and shed a new light on as moral choices. To give an example, Yee et al. (2021) suggest different possible choices than the single saliency point selected by Twitter engineers. Taking the inquiry *upstream*, PD could seek the participation of the algorithm designers by inviting them to an experimental space where the "good enough" would be hesitated through the encounter with the consequences of their choices at the same time as the other objectives

they were faced with in the heat of the moment, like the consistency of the timeline, the productivity of users, the computing power of mobile devices, the speed etc. The “good enough” would present itself as a moral issue redistributed amongst the different technical practices involved in the development of the cropping algorithm, i.e. as “a circulating force connecting the multiple entities brought to light by the hesitations and doubts concerning the proper distribution of means and ends within the course of technical practice” (John-Mathews et al. 2023), which are usually unaccounted for in PD for AI. This participatory configuration would represent a significant departure from participation 2 where the saliency model is made into the unique subject of moral choices and preferences. It would also bring back to the foreground the workplace as a site of participation in the construction of morality, and the organisation of labour as favouring particular outcomes that are more harmful for some communities than others.

Downstream inquiry

Ananny (2022) extends in another direction the community concerned with troubles, which he defines as “algorithmic errors” although he does not frame them as only technical. His argument is that there are multiple ways to frame troubles, and that that framing is indicative of the elements within algorithmic sociotechnical assemblages that one considers can or cannot be acted upon and changed. The title of the article ("seeing like an algorithmic error") suggests displacing the traditional opposition between how humans and machines “see”, to understand “seeing” as the active operation of framing problems: beyond the identification of *who* is affected, it also interrogates the relationships between forces, objects, systems, imagination that affect *how* we see “seemingly private, individual errors in system design, datasets, models, thresholds, testing, and deployments”, so that troubles can be made into collective problems (Ananny, 2022, 21). The author calls for participatory configurations to generate "communities of interpretation” that hesitate upon the description of the sociotechnical system that the ML system is entangled with. Concerning the cropping algorithm, PD could help turn the inquiry *downstream*, i.e. into making the trouble not only about revealing the biases of saliency model itself, but also revealing its entanglement with the overall attention economy, with power issues underlying recognisability on social media (Jacobsen 2021) or any other framing that could arise from gathering a diverse community of participants as “interpreters” of the issue. Again this represents a departure from participation 2, as participation 2 did not allow to reframe the problem differently than that of a misbehaving independent and standalone “cropping tool” where user-centred design principles could be called upon to give back control to the user. PD

could help decenter from the cropping itself to emphasise the larger public, cultural or economic issues that the cropping tool contributes to problematize. In that perspective, it does not make much sense to oppose the figures of "the user" and "the algorithm" and calls for different participatory configurations where the industry's interests, and the commitment of design to them, could be challenged.

BIBLIOGRAPHY

Ahmed, Sara (2017). *Living a Feminist Life*. Durham: Duke University Press.

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

Ananny, Mike (2022). "Seeing like an Algorithmic Error: What Are Algorithmic Mistakes, Why Do They Matter, How Might They Be Public Problems?" *Yale JL & Tech.* 24: 342.

Andersen, Lars Bo, Peter Danholt, Kim Halskov, Nicolai Brodersen Hansen, and Peter Lauritsen (2015). "Participation as a Matter of Concern in Participatory Design". *CoDesign* 11(3–4): 250–61.

"A Photo Upload API". *Twitter Developer Platform Blog*. Accessed July 15, 2024. https://blog.x.com/content/blog-twitter/developer/en_us/a/2011/photo-upload-api.

Asaro, Peter M (2000). "Transforming Society by Transforming Technology: The Science and Politics of Participatory Design". *Accounting, Management and Information Technologies* 10 (4): 257–90.

Bakhshi, Saeideh, David A. Shamma, and Eric Gilbert (2014). "Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, New York, NY, USA: Association for Computing Machinery, 965–74.

Baumer, Eric P. S., and Jed R. Brubaker (2017). "Post-Userism". In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA: ACM, 6291–6303.

Binder, Thomas, Giorgio De Michelis, Pelle Ehn, Giulio Jacucci, Per Linde, and Ina Wagner (2011). *Design Things*. Cambridge, MA, USA: MIT Press.

Birhane, Abeba, William Isaac, Vinodkumar Prabhakaran, Mark Díaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed (2022a). ”Power to the People? Opportunities and Challenges for Participatory AI”. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–8. Accessed November 14, 2022. http://arxiv.org/abs/2209.07572.

Birhane, Abeba, Vinay Uday Prabhu, and John Whaley (2022b). “Auditing Saliency Cropping Algorithms”. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA: IEEE, 1515–23. Accessed January 3, 2024. https://ieeexplore.ieee.org/document/9706880/.

Bratteteig, Tone, and Guri Verne (2018). “Does AI Make PD Obsolete?: Exploring Challenges from Artificial Intelligence to Participatory Design’. In *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2*, Hasselt and Genk Belgium: ACM, 1–5.

Bredies, Katharina (2008). “Confuse the User! A Use-Centered Participatory Design Perspective.” In *Designed for Co-Design*, edited by Katja Kattarbee, Andrea Botero, Tuuli Mattelmäki, and Francesca Rizzo. PDC 08 Pre-Conference Workshop.

Chilvers, Jason, and Matthew Kearnes (2020). “Remaking Participation in Science and Democracy”. *Science, Technology, & Human Values* 45(3): 347–80.

Chowdhury, Rumman, and Jutta Williams (2021). “Introducing Twitter’s First Algorithmic Bias Bounty Challenge”. *X Blog*. Accessed July 15, 2024. https://blog.x.com/engineering/en_us/topics/insights/2021/algorithmic-bias-bounty-challenge.

Christensen, Michelle, and Florian Conradi (2019). *Politics of Things: A Critical Approach through Design*. Basel, Switzerland: Birkhäuser.

Costanza-Chock, Sasha (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge, Massachusetts: The MIT Press.

Crawford, Kate (2017). “The Trouble with Bias”. 2017 NIPS Keynote. Accessed July 15, 2024. https://www.youtube.com/watch?v=ggzWIipKraM.

Delgado, Fernando, Stephen Yang, Michael Madaio, and Qian Yang (2023). “The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice”. Accessed December 14, 2023. http://arxiv.org/abs/2310.00907.

Dewey, John (2005). *Art as Experience*. New York, New York: Penguin Publishing Group.

DiSalvo, Carl (2022). *Design as Democratic Inquiry: Putting Experimental Civics into Practice*. The MIT Press.

Dixon, Brian (2020a). "From Making Things Public to the Design of Creative Democracy: Dewey's Democratic Vision and Participatory Design". *CoDesign* 16(2): 97–110.

Dixon, Brian S. (2020b). *Dewey and Design: A Pragmatist Perspective for Design Research*. 1st ed. 2020 édition. Cham: Springer Nature Switzerland AG.

Edwards, Paul N. (2018). "We Have Been Assimilated: Some Principles for Thinking About Algorithmic Systems". In *Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology*, IFIP Advances in Information and Communication Technology, eds. Ulrike Schultze et al. Cham: Springer International Publishing, 19–27.

Falco, Gregory (2019). "Participatory AI: Reducing AI Bias and Developing Socially Responsible AI in Smart Cities". In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 154–58. Accessed January 24, 2024. https://ieeexplore.ieee.org/abstract/document/8919542.

Hansson, Karin, Laura Forlano, Jaz Hee-jeong Choi, Carl DiSalvo, Teresa Cerratto Pargman, Shaowen Bardzell, Silvia Lindtner, and Somya Joshi (2018). "Provocation, Conflict, and Appropriation: The Role of the Designer in Making Publics". *Design Issues* 34(4): 3–7.

Haraway, Donna J. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press.

Hern, Alex (2020). "Twitter Apologises for 'racist' Image-Cropping Algorithm". *The Guardian*, September 21, 2020. https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm.

Jacobsen, Benjamin N (2021). "Regimes of Recognition on Algorithmic Media". *New Media & Society*: 14614448211053555.

Ji, Jiaming, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan,

Zhonghao He, Jiayi Zhou, Zhaowei Zhang, Fanzhi Zeng, Kwan Yee Ng, Juntao Dai, Xuehai Pan, Aidan O'Gara, Yingshan Lei, Hua Xu, Brian Tse, Jie Fu, Stephen McAleer, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, Yike Guo, and Wen Gao (2024). "AI Alignment: A Comprehensive Survey". Accessed February 4, 2024. http://arxiv.org/abs/2310.19852.

John-Mathews, Jean-Marie, Robin De Mourat, Donato Ricci, and Maxime Crépel (2023). "Re-Enacting Machine Learning Practices to Enquire into the Moral Issues They Pose". *Convergence* 0(0): 13548565231174584.

Kalluri, Pratyusha (2020). "Don't Ask If Artificial Intelligence Is Good or Fair, Ask How It Shifts Power". *Nature* 583(7815): 169–169.

Kümmerer, Matthias, Thomas S. A. Wallis, and Matthias Bethge (2016). "DeepGaze II: Reading Fixations from Deep Features Trained on Object Recognition". Accessed December 20, 2023. http://arxiv.org/abs/1610.01563.

Latour, Bruno, and Peter Weibel (2005). *Making Things Public*. MIT Press.

Lee, Min Kyung, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, and Ariel D. Procaccia (2019). "WeBuildAI: Participatory Framework for Algorithmic Governance". *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW): 1–35.

Lorusso, Silvio (2021). "The User Condition". Accessed November 1, 2022. https://theusercondition.computer/.

Marres, Noortje (2005). "Issues Spark a Public Into Being: A Key but Often Forgotten Point of the Lippmann-Dewey Debate". In *Making Things Public*, eds. Bruno Latour and Peter Weibel. MIT Press, 208–17.

Marres, Noortje, and David Stark (2020). "Put to the Test: For a New Sociology of Testing". *T he British Journal of Sociology* 71(3): 423–43.

Majewski, Taylor (2024). "It's Time to Retire the Term 'User'". *MIT Technology Review*, April 19, 2024. https://www.technologyreview.com/2024/04/19/1090872/ai-users-people-terms/.

Meunier, Axel, Jonathan Gray, and Donato Ricci (2021). "A New AI Lexicon: Algorithm Trouble". *A New AI Lexicon*. Accessed November 1, 2022. https://ainowinstitute.org/publication/a-new-ai-lexicon-algorithm-trouble.

Meunier, Axel, Donato Ricci, Dominique Cardon, and Maxime Crépel (2019). "Les glitchs, ces moments où les algorithmes tremblent". *Techniques & Culture. Revue semestrielle d'anthropologie des techniques*. Accessed September 16, 2021. https://journals.openedition.org/tc/12594.

Noble, Safiya Umoja (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

"Picture This: More Visual Tweets". *X Blog*. Accessed July 15, 2024. https://blog.x.com/en_us/a/2013/picture-this-more-visual-tweets.

PhotoShelter (2020). "Vision Slightly Blurred: Twitter's Image-Cropping Algorithm". *Facebook*. Accessed July 15, 2024. https://www.facebook.com/PhotoShelter/videos/vsb-twitters-image-cropping-algorithm/745693006153772/.

Rahwan, Iyad (2018). "Society-in-the-Loop: Programming the Algorithmic Social Contract". *Ethics and Information Technology* 20(1): 5–14.

Redström, Johan (2008). "RE:Definitions of Use". *Design Studies* 29(4): 410–23.

Redström, Johan, and Heather Wiltse (2020). *Changing Things: The Future of Objects in a Digital World*. London, UK New York, NY: Bloomsbury Visual Arts.

Shaffer Shane, Tommy (2023). "AI Incidents and 'Networked Trouble': The Case for a Research Agenda". *Big Data & Society* 10(2): 20539517231215360.

"Share a Photo via Text Message'. *X Blog*. Accessed July 15, 2024. https://blog.x.com/en_us/a/2011/share-a-photo-via-text-message.

Simonsen, Jesper, and Toni Robertson (2012). *Routledge International Handbook of Participatory Design*. Routledge.

Sloane, Mona, Emanuel Moss, Olaitan Awomolo, and Laura Forlano (2020). 'Participation Is Not a Design Fix for Machine Learning'. *arXiv:2007.02423 [cs]*. Accessed November 23, 2020. http://arxiv.org/abs/2007.02423.

Steen, Marc (2013). "Co-Design as a Process of Joint Inquiry and Imagination". *Design Issues* 29(2): 16–28.

Stray, Jonathan, Ivan Vendrov, Jeremy Nixon, Steven Adler, and Dylan Hadfield-Menell

(2021). “What Are You Optimizing for? Aligning Recommender Systems with Human Values”. Accessed January 26, 2024. http://arxiv.org/abs/2107.10939.

Suchman, Lucy (2012). “Configuration”. In *Inventive Methods*, Routledge.

Theis, Lucas, Iryna Korshunova, Alykhan Tejani, and Ferenc Huszár (2018). “Faster Gaze Prediction with Dense Networks and Fisher Pruning”. Accessed December 16, 2023. http://arxiv.org/abs/1801.05787.

Theis, Lucas, and Zehan Wang (2018). “Speedy Neural Networks for Smart Auto-Cropping of Images”. *X Blog*. Accessed July 15, 2024. https://blog.x.com/engineering/en_us/topics/infrastructure/2018/Smart-Auto-Cropping-of-Images.

“Twitter Algorithmic Bias - Bug Bounty Program” (2021). *HackerOne*. Accessed July, 15 2024. https://hackerone.com/twitter-algorithmic-bias.

“Twitter Photos: Put a Filter on It’. *X Blog*. Accessed July 15, 2024. https://blog.x.com/en_us/a/2012/twitter-photos-put-a-filter-on-it.

Yee, Kyra, Uthaipon Tantipongpipat, and Shubhanshu Mishra (2021). “Image Cropping on Twitter: Fairness Metrics, Their Limitations, and the Importance of Representation, Design, and Agency”. *Proceedings of the ACM on Human-Computer Interaction* 5(CSCW2): 1–24.

Young, Meg, Iretiolu Akinrinade, Ania Calderon, Rigoberto Lara, Eryn Loeb, and Tunika Onnekikami (2023). “Shaping AI Systems By Shifting Power”. Accessed 15 July 2024. https://medium.com/datasociety-points/shaping-ai-systems-by-shifting-power-ee95f7c3edf9.