AI Ethics: Data Feminism in a World of Power

Alexandra Radford
Honors Thesis Publication
Advised by Dr. Alexis Elder

AI has contributed to the degradation of workers rights. How technology AI is built and used exploits impoverished minority communities, creating a limited perspective and power imbalance that is built into commercial AI. Looking at how power structures play a part in societal conceptions from dominant groups, we can take a closer peek at *how* AI systems are being created and used. This includes algorithms, feature selection, training, and tracking. An important question this paper goes over is how AI technology is made to further oppress minority communities, and what institutions and corporations have structured in order to retain their own power over those communities. Not only do centralized power structures push minority communities down, but it takes part and results in biases in AI training data itself as well as the lack of relevant data, since missing data and data voids are a result of structural powers dictating what data is collected. What happens then, if that power structure is used in the wrong way? What does it look like to be at the bottom of the bottom? I say 'bottom' because minority communities like ghost workers, are people denied regular worker rights. Amazon Mechanical Turk (AMT) only allows employers the power to rate ghost workers, not the other way around. Therefore, employers are the ones given power from AMT, and exercise that power over ghost workers, who are collecting and cleaning data. To support ghost workers employed through AMT, some activists have developed an accessible tool called Turkopticon. Turkopticon challenges AMT by using data feminism to decentralize and pluralize technology power structures . Data feminism works to challenge the powerful, while Feminist Epistemology acts to dismantle power structures. The aim of Turkopticon is to empower people–to empower minority communities like ghost workers. Work labor rights can then be rebuilt using feminist epistemology as a tool to unite communities and their intersectional perspectives. Turkopticon, as relative to AI technology, takes the first step into empowering work class communities, while also resulting in more accurate AI. This paper suggests the next steps in empowering ghost workers by adding to Turkopticon's features.

What is Data Feminism?

D'Ignazio and Klein (2020), in their book titled *Data Feminism,* define data feminism in the following way:

...feminism begins with a belief in the "political, social, and economic equality of the sexes," as the Merriam-Webster Dictionary defines the term— as does, for the record, Beyoncé.3 And any definition of feminism also necessarily includes the activist work that is required to turn that belief into reality. In *Data Feminism*, we bring these two aspects of feminism together, demonstrating a way of thinking about data, their analysis, and their display, that is informed by this tradition of feminist activism as well as the legacy of feminist critical thought. (p. 3)

As they note, data feminism is derived from feminism and revolves around unjust morality against women from men. But feminism does not solely focus on gender relations, but also social inequities. Gender relations are heavily affected by race and class–and because we aren't focusing on individual parts but a whole issue, feminism holds a holistic approach toward race and class inequities as a whole (Ferguson, 2017). Although feminism isn't a new concept, data feminism is. How can data be used in a feminist way? Collecting, analyzing, and understanding data about or for women is a must–but data feminism directs our attention to where structural inequities arise in that data, and how to fix them, utilizing tools from feminist theory. Data feminism like so helps us understand issues of data involving minoritized groups. Structural injustice limits the inclusion of minoritized communities' perspectives, and data feminism helps bring about justice. Ghost workers are a part of the minority workers community and are major sources of data for AI–but their perspectives are not represented, acknowledged or respected. Irani & Silberman (2013) created Turkopticon as a means to enable rights for ghost workers, who now have a voice and agency against AMT's oppressive system. To take it further, D'ignazio and Klein (2020) introduce data feminism and feminist epistemology as a methodology for improving artificial intelligence. Specifically, more just methods to how AI data is being collected, crowdsourced, coded, feature selection, training, etc. The data that fuels AI needs to be reevaluated, and what better way than by those who are affected most by the companies who build artificial intelligences' degradation of rights? Data science has for a long time been a man's world (Botella et al, 2019), and oversimplifying issues of diversity and gender in the sciences have worsened inequalities in STEM (Alegria & Branch, 2015). But, I do see a

future for pluralistic methodologies that don't oversimplify issues of diversity and gender in technology. Introducing pluralistic approaches has strong implications for cognitive science. Being interdisciplinary, cognitive science includes linguistics, physiology, philosophy, neuroscience, anthropology and computer science–which includes data, which data feminism is worried about. Data feminism explains data through research generation and action–by locating and analyzing socioeconomic factors and forcing us to revisit assumptions about how power structures affect the data collected, used, and interpreted. Data feminism integrates inequities as a whole, while cognitive science is interdisciplinary between humanities and data sciences. This becomes clear when it comes to how cognitive science engages with artificial intelligence. To understand how AI should be built, we use cognitive science, but cognitive science needs data feminism. This means we need to look at what has been unethical in terms of ghost workers rights. The basis of technology, like data and informational frameworks to build AI, are so deeply incorporated into human life that the dominant power group is present within and we must grapple with that.

Introduction to a Pluralistic application in algorithms

A prime example of using a pluralistic view involving feminist epistemology to use data more equitably, would be **divergent shuffle**. A divergent shuffle search algorithm takes different domains' search results and shuffles them–providing users with results that represent a variety of perspectives on the world. To better look at how this is used, seeing how perspective sharing can be applied to geographical data is helpful, but it can be used on other criteria as well. Geography is an excellent example, because what happens in more well off geopolitical regions is that they tend to produce search results that overrepresent perspective from the dominant regions. Take North America for example, when you search something like our method 'feminist epistemology', a list of top searches would most likely come from North America, Europe, and and pass over other perspectives from the world. But, instead if we implement results of other continents, countries, and areas, like Africa, South America, Asia and Australia, the best results from these areas would be "shuffled" into a list of top results. This shares differing perspectives and experiences, and can be used to break up dominant powers, and their assumptions, about what the world appears as. It's an interesting idea to implement this connectedness

from where we are getting our information–as it will push us towards a **non-weird** kind of knowledge. Non-weird, meaning viewing others' knowledge to not be outside the norm. How can this be applied to those who are behind filtering, managing, and problem solving, for more equitable AI?

The Workers Behind AI

Those who are behind crowdsourcing (and more) are the ones who, like I've mentioned before, filter, manage, and organize data, in order for AI to function, not directly employed by the company but hired on a per-task basis through online platforms. This leaves them without normal workplace protections. Mary Gray and Siddarth Suri write, " No employment laws capture the on-demand gig economy's odd mix of independence from any single employer and dependency on a web based platform. As the taskmasters of the gig economy, on-demand platforms make their money by matching those buying and selling human labor online, generating a two-sided market of myriad businesses and anonymous crowds of workers." (p. xvii, Gray and Suri, 2019) What categorizes these people is their **ghost work**, as these workers have been made invisible behind the online platform and treated with unfair working conditions and agreements, and not enough organized legal protection. And because ghost workers can't see each other, they aren't able to organize together and share perspectives. Ghost workers are those who create and clean up the corporation's AI training data. They are constantly cleaning another's mess. I mention 'mess' but it should rather be referred to as smaller forms of unclean areas, like dust or a lego left behind in the corner of a room. They aren't cleaning up the big things that you can easily see, but the small legos left under couches and buried in your backpack. The people behind ghost work are diverse in that some are dependent on it being their main income and others are not. Ghost workers are not all in it for the same thing, as some have another full time job. Then there are individuals who are experimenting with ghost work, and found it to not be to their taste–'experimentalists'. 'Experimentalists' usually try a couple projects before finding out that this line of work either isn't for them, or, recognize the unethical ways in which they are being treated for doing work.

Missing Data and Data Voids

D'Ignazio and Klein (2020) introduce the issue of missing data as something that reveals social biases and indifferences. Knowledge of gender, and minoritized communities can influence constructs like justification, knowing, and perspective (Stanford Encyclopedia of Philosophy, 2020). Filling out missing data with perspectives from oppressed people is an example of challenging power with data feminism and feminist epistemology, while also being about to take a step back and recognize what is going on in a community from the inside.

Ghost Workers play a part in the process of filling in missing information by completing surveys and labeling data, but they lack power. Looking at it from the perspective of those who were supposed to be filling these voids and instead have been pushed into oppression, AMT limits their ability to exercise and represent their perspectives in that highly centralized system. Not only are ghost workers helping the corporation grow, but they aren't being empowered as data authors themselves. Filling data voids isn't just about giving the corporations what they want, but starts from the bottom up so the data is actually representative. First we use feminist epistemology as a method to dismantle dominant power structures, specifically corporations holding ghost workers down. Then, when rebuilding and empowering communities like ghost workers, they will have the resources to fill in voids as empowered authorities on their own perspectives. The ideal goal is that there should not be a hierarchical system, and this is what feminist epistemology strives to accomplish. I don't assume all conceptions of power just suddenly disappear everywhere, but practicing methodologies that resist oppressive hierarchies can empower communities to rebuild labor rights and result in better data collection methods and outcomes. The past data voids will still exist as there are too many to count, but then ghost workers would be able to utilize their empowerment for filling those voids.

To dig deeper, it is necessary to look at dense data clusters of similar perspectives, focusing on data collected from Turkopticon's users, ghost workers from across the globe. Turkopticon collects data from these users about their experiences working for AMT, the main purpose is to rebuild worker labor rights to empower perspective sharing upon users.

One important void is ghost workers' inability to share reviews of contractors. A form of these filling data voids, as done by Turkopticon, is having AMT workers fill out what they think is most important for workers' rights and then implementing that into the form of reviews of employers. To help ghost workers become data authors themselves is to fill in the data void presented by AMT with the help of Turkopticon. This empowers workers to take control of their own work conditions. Amazon's Mechanical Turk is made for employers who are looking for contracted or individualized workers to provide essentially just that: their work. The way AMT works is that employers can create an account (with verification) and put up projects that they need work done for. Employers offer pay per project, usually at a rate made up from themselves. The other side of the coin involves workers who select work, depending on their rating. A worker earns their ratings the more accepted work they do, and doesn't personally contribute to their own rating, as everything is left to the employer. Essentially, AMT are the employers to these workers: ghost workers. Sometimes the projects given are ones specifically arranged around a certain rating so that only experienced workers can have access to the project. Essentially limiting the possibilities of people wanting to work. Employers can also refuse to pay ghost workers while keeping their work.

Let's say an employer rejects the hard work you submit, not only will your rating go down, but they can still take your work without paying you. Now it'll be harder for you to get more jobs (because of your rating), you lost money, the hard work put in was stolen, *and* the employer didn't offer a reason for rejection. Although it seems like Amazon was trying to make it easier for two communities to get into contact with each other, AMT has raised a risk of toxicity involving the ethics of its workers. There seems to be an unbalanced power between the workers and employers, with one being categorized as having no power besides submitting work and the other having the power to control over the other group. The workers who experience unfair oppression have gone through those interpersonal experiences with employers, and have related to Amazon comments and concerns, with no avail. The perspective sharing of a dominant group isn't necessarily a good thing, as it ignores the ethics behind crowdwork, and focuses on contract-like workers - who are treated in a separate class than regular workers. These workers are

viewed as producers who aren't a part of the employers corporation, so they don't need to provide minimum wage, 401k, or health insurance. The basic rights an employer usually would give their employee isn't related to at all. The problem with AMT is that it was created to enable employers in power, given by those who have dominant power, and advertising AMT as something helpful for both sides.

What makes data voids even more dangerous is when people may use certain search terms to harm certain communities and individuals–therefore manipulating the search engine into something different. Corporations and institutions, like AMT, can also become involved in this type of coercion–contributing to data voids being altered into other potentially coerced data collection. Making an account and logging in to AMT, you are greeted with employers in charge of every aspect of its workers' user field. Upon logging in, AMT gives power to its employers to regulate work conditions in the form of reviews, money, accepting the work, and reporting. A data void in this instance may lie within the lack of data from the reviews of workers. The employers have more power, so they are able to manipulate review results to the detriment of workers. All aspects of information related to data, especially those ratings, start with a blank slate of no data on the employer. When rated potentially in a negative light, then this void is molded into something as exploitation of its workers.

Object and the Knower

In terms of collecting data, before we can assess the actual collected data from users like ghost workers, there's still disconnection between the employers and workers behind systems like AMT. AMT acts as the **knower**, as they are the ones with the power over workers. The ghost workers are the ones presumed to provide information and data for the employers, which the employers exercise that expectation from them already **being** an employer. The data collected is what the employers put together, or piece, in order to create knowledge. But what makes this off-putting is that it ignores workers' perspective totally–they are only viewed as resources that provide data to a greater power. That greater power is exercised by employers to create disconnected data (knowledge), and therefore a distorted **object**. Just as a reminder, the **object** is the study of interest, basically what the data collected

(knowledge) makes up together. Before we can take a look at data clusters, there has to be a dismantlement of the disconnection between the object and knower. Workers need to be seen as not just sources of objects to be spewed out, but as those who can share their perspective knowledge. We want them to be knowers themselves. Anderson (2014) identifies the problem with aperspectivity–and recommends an integrated form of knowing:

> "Feminists question the intelligibility of a "view from nowhere," and a presuppositionless, bias-free science, for both postmodernist (Haraway 1991) and pragmatist (Antony 1993) reasons. Knowers are situated. The underdetermination of theories by evidence implies that biases are needed to get theorizing off the ground. Rather than undertake a futile attempt to inquire without biases, we should empirically study which biases are fruitful and which mislead, and reform scientific practice accordingly (Antony 1993)."

Taking it a step further, the **View from Nowhere** is assumed by those who are employers. They presume and ignore (don't use) workers perspectives, including putting a cap on what exactly *can* be sourced in the first place. The control is taken by employers in totality, meaning there aren't any different point of views from the data source. As ghost workers are at the bottom of the power structure, they are therefore used as resources to put collected data together. Only the point of view is stripped away, specifically producing what the employers want. Reestablishing the worker as knower will create a healthier relationship between the two, and will empower ghost workers to spread perspective through communities and AI .

Turkopticon

Turkopticon was created as something that can be used as a tool to empower workers on AMT. First it is important to understand what Turkopticon means, where the name stemmed from. Silberman & Irani (2013) have compared Turkopticon to Foucault's central prison design–panopticon–a prison surveillance system where the guard is at the center of the prison and can monitor everyone in sight. Although the guard can see everyone else, the prisoners do not know if, or when, they are being watched, so they act as if they are being watched. This reflects on the purpose of Turkopticon as it keeps AMT

from acting out of order. In a smaller sense, its browser extension and web-based technology exaggerates the case between the worker and the employer, a mutual aid for workers. This leads to the question, "how does Turkopticon work?" and eventually "how does Turkopticon implement feminist epistemology as a way to empower the worker class?" To address the first question, in order to regulate the wrongs of AMT, the creators behind Turkopticon created a 'bill of rights' by giving surveys to AMT worker users.

Using the survey results of users, they came up with a classification system for rating contractors in four categories: 1) Communicativity 2) Generosity 3) Fairness 4) Promptness. Each one of these is rated out of a five-point scale decided on by the worker in the form of a review–which rates a contractor. When a worker decides on the project they are interested in, they will be shown what the employer's rating is before doing the work, as a way of giving them the choice whether to proceed. Another thing Turkopticon allows is for workers to not specify their real names or emails, which help maintain their privacy from employers. Different from AMT, Turkopticon uses moderators to review issues and incidents. You can flag something and the moderator will look at the reported issue, and only the moderator can get rid of that review once it's reported. A strength of the moderators Turkopticon uses is that they have themselves experienced being ghost workers. They are selected this way because a moderator needs to have multiple perspectives and experiences in order to appropriately understand what is going on when a review or issue is flagged.

Turkopticon does use a feminist epistemology approach in building this web-extension, to help workers to have regular or more rights in their profession. Feminist epistemology, as mentioned before, centers the "particular perspectives of the knower" (Stanford Encyclopedia of Philosophy, 2020). Turkopticon lets workers share perspectives to rebuild power. Rebuilding that power forms a stronger relationship together instead of pushing workers farther apart, as the AMT structure encourages. Something specific that Turkopticon takes from panopticon is its control through surveillance. Basically it will tell you if a previous worker has reported the terms and conditions of a previous contract were violated, as it will show up in the review when prospective workers first look at an employer's reviews.

One complicated issue is that most employers do not offer a set minimum wage. Because they do not have to offer a certain amount, people who use the web-browser to find work projects are all over the world–meaning what minimum wage is somewhere, it will be different geographically. The issue with this is that an employer can, if they want to, possibly target certain workers from a specific part of the world. Let's say that someone from Australia is an employer, and they offer three cents per project, but are aiming exceptionally low to capture those from impoverished countries. They can do this in order to not pay the minimum wage it costs for people to live in Australia. This brings up the question, should there be a universal minimum wage set for all Turkopticon employees, which employers have to start with and can only go up from there? This is made difficult because of varying global costs of living. Rather than measure all worker interests the same way, the four categories to help workers to share and research based on their own perspectives, including but not limited to employers respecting those who want more money for all the work they are doing.

But there is a problem because different workers have different interests. Subgroups of workers may fall into dense data clusters. Dense data clusters are categorized as 'dense' because they hold similar perspectives, characteristics, interests, and views from each other, so users are able to roughly connect to those with these similar aspects. Turkopticon, though, connects and categorizes these pieces of data into clusters, specifically clusters that may relate to each other based on commonalities of either employers or users. Data collected from a specific contractor might have a dense cluster of perspectives between the Chicago and Detroit locations, and therefore will have a stronger link between them. When moving to somewhere outside Turkopticon's initial scope (North America), however, then the link between other workers' aspects can weaken. It isn't about geographic location, but the lack of data connecting the two to begin with–meaning data voids hinder communication within communities. If we can identify not just where the missing data is coming from, but why the link network isn't as strong for something like 'labor rights in the army' and 'labor rights in Whole Foods'. The high-density data clusters are the ones who are actively able to converse with each other, but in a not-so-dense space of the network there might also be data that hasn't been built upon yet. By building on those weaker links between dense and less dense

spaces, we can start to pick up the pieces of missing data to start sharing more perspectives and conversations within communities.

<p style="text-align:center">**Turkopticon vs. Amazon Mechanical Turk**</p>

Turkopticon functions to help ghost workers who have not received rights as workers. As influential as this web-extension has been, they still have their own faults, but are willing to change and revise for the better of the working class–for the better of ghost workers behind the scenes of corporations. Turkopticon pushes forward with the enhancement of AMT and presents a representation of feminist epistemology–a decentralized and pluralistic way to break up power of dominant groups. One aspect that makes Turkopticon so pluralistic is the use of moderators compared with AMT. Turkopticon, as mentioned before, works so well because moderators are the 'guards', while AMT contractors are the 'prisoners' (Foucault, 1975). Being able to write reviews is what gives workers the power to regulate employers' unfair treatment. Using a review scale in the form of four categories–communicativity, generosity, fairness, and promptness–these are used to share perspective with ghost workers within the AMT/Turkopticon realm. Turkopticon also values their worker users' privacy from employers.

Turkopticon is an independent platform, not directly integrated into AMT, but it lets workers see each other and share perspectives. Turkopticon highlights workers' rights by basing its system on a review classification system to share experiences with employees, while giving users control through privacy settings, and promoting orderly discussion through moderators to help reach shared perspective across the user network. AMT, by hiding workers' perspectives, presents itself as the 'view from nowhere', therefore severing ties between the 'knower' and 'object' of study. AMT establishes who does and doesn't get power by making requestors/employers those who are able to rate ghost workers but not the reverse. This favors the employers, not the workers. Pay isn't great because AMT doesn't have a minimum wage because it is task-based, not time based. Remember that ghost workers are contracted, but not full employees. The exploitation of workers is what makes ghost workers at the bottom of the bottom. AMT works to connect ghost workers with those who have the power to suppress them. This happens with reviews from requestors, low pay, or taking your work without paying without a reason. Ghost

worker rights are in the hands of Amazons' employers, not the other way around.Turkopticon helps ghost workers gain agency by sharing perspectives.

Empowerment vs. Dominating Power

Turkopticon is a living example of something trying to dismantle power, but in what ways can we break down how that system came to be? We will be looking at this in the view of Turkopticon and AMT, starting with those who currently have the most power: Jeff Bezos. These examples are intended to reiterate how dominant power has a hold on minority communities or those outside the scope of power. Jeff Bezos is, right now, near the core of power, as he holds the most power over his employees and even people outside Amazon's workforce (the customers). Then we have those who work for him under regular employee contracts. They are the first round of employees who have a 401k, set wages, covered insurances, and are distributed power from higher up. The employees themselves then are able to hold power, as they have what the CEO has given, and can now choose to use that power over others. Most may not think that having a regular office job comes with this kind of ability, but that is because businesses like Amazon are run as their own hierarchy system inside another hierarchy system: capitalistic society. Inside our society lives Amazon and other big corporations that act within each other, and ultimately bring forth a more constricted centralized institutional structure. At the top we have the CEO, then their workers below, and then the ghost workers. Ghost workers have no connections to the regular workers, let alone the CEO–and are hidden so as to not be seen in the first place. Purposefully this can be used by dominant groups to keep ghost workers oppressed and. The problem is that there is this kind of structure to begin with that not only holds so much power, but is able to distribute amongst those it deems worthy. To look closer at what that is, is to reveal centralized power itself. The issue of power isn't that ghost workers aren't getting any, but where it's coming from: the CEO. We want to dismantle the distributor/institution and rebuild a system which is rich to empower specifically amongst ghost workers.

Using Divergent Shuffle

Understandably, AMT works most prominently in the U.S, but is trying to reach out. In other words, expand beyond just a few communities. That expansion is what puts those outside the central power ring to be even further oppressed and pushed, possibly further down than those who are inside the central power. Turkopticon does present access to everyone, especially those who are outside the main central power ring, with the ability to review and rate - something that can and will change how employers treat and respect people working behind the scenes. There is one little problem here, the disconnection of rights is further than regular ghost workers. Pay-wise, every country has a different economy (way of life), and employers can take advantage of, let's say someone living in India, whose wage differs from Canada, employers could possibly alter their minimum offering as a way of potentially targeting individuals (and communities) outside of their realm of centralized power. I'm not specifically saying that all employers are doing this, but there is potential for AMT to be used in order to broaden its scope of centralized power. To disable and dismantle would mean to focus on how Divergent Shuffle can highlight the voices of those currently outside the main scope of power, hopefully as a tool to **empower**. Turkopticon mainly focuses on giving ghost workers rights, but what about those who may not be located in North America in the first place? They have different rules, kinds of power concepts, and rights all together. Thus how can the Divergent Shuffle aid those who are further outside the outsiders (Ghost Workers in the U.S)?

As we begin to inspect Divergent Shuffle as a tool, the difference between shuffling individual perspectives compared to communal/community based ones, can drive us to a more decentralized empowerment of individuals. What we want is to push past not just empowering those individuals, but to empower communities. If the Divergent Shuffle were used individually it would most likely, as mentioned above, shuffle around more privilege - not dismantle individual (dispositional) power. If you look at it in the terms of office workers, ghost workers, and CEO, then in the world of corporations those who are handed (passed) power down to them, have that individual privilege. So then what about those in the basement? Since they are outside that initial power scope (although still affected by it), they aren't being any less oppressed. Instead it reestablishes individual power, the opposite of dismantling and empowering communities.

Divergent Shuffle, like mentioned above, breaks up power within search engines, and rebuilds a more diversified field of knowledge from multiple persons (countries) perspective. It works to implement more perspective from *everywhere*, not centrally just North America. Divergent Shuffle can also work for Turkopticon, as multiple people are using the system from different locations of the world, to break up dense clusters and enable more connections between them. Divergent Shuffle can help Turkopticon to bring forward rights for those outside the centralized power. Something to highlight, and would be interesting to see, is what Turkopticon would look like if it focused on those who are outside the centralized power *with* Divergent Shuffle as a method to dismantle a centralized power.

To improve, if the Divergent Shuffle approach was incorporated into Turkopticons system, it would need to specifically target networked perspectives. If this happens then there could be a better resource for them to find on their own (specifically those that have had less power), and empowering networked perspectives–networked communities of Ghost Workers across the world. Networked perspectives are those with diverse intents supporting each other in the user field of Turkopticon, hence a 'network' is formed. Sharing that perspective through Turkopticon with reviews isn't enough, as these supporting diverse intentions are hindered from dominant social groups. Inside Turkopticons network lies North America's group cluster, but what happens outside of that network cluster?

Accessibility can reach everyone in concept, and because of that users will naturally cluster more together; like with shared interests, characteristics, geographical locations, etc. And, these clusters represent many users' perspectives - many live differently among others. These populous networks where there are dominant social groups present can become swamped, or in other words: congested. They need a tool to get a broader representation. Because these networks are congested, workers aren't necessarily getting that information from less congested areas. And that information will serve in their best interest for broader representation. Applying this to examples, it's safe to say that the North America cluster lacks that kind of broader representation. Using a dominant social group like AMT, the North America cluster condenses interests and perspectives because of its geographical location and socioeconomic status. The problem with SES and geography is that it's on a global scale, not just in North America. Although AMT

primary is located and deals with workers in the United States region, the social status of workers has always been below the CEO or institutional power–so that makes those working in the NA region dense *and* disconnected from somewhere like India or Australia. Vice -versa this problem is global because information isn't always available outside of North America and the same can go for India or anywhere else. So if India and North America have a highly dense network in their own geographical area and SES (all within Turkopticon), then there really isn't much information being passed through to each other. Although Turkopticon works as one system, and it isn't trying to split these communities up, mixing in randomized perspectives into weak network lines between India and North America is what can further decentralize workers by connecting individual differences to each other.

What does it look like to use DS to broaden Turkopticons reach? Identifying dense clusters, analyzing why they are so dense (lack of info sharing), and then applying DS as a means of breaking them up. Breaking up something can also push it to connect/build. The application of DS is what would be plucking from different perspectives and using them to build relationships globally. Weaker network links can then expand and thicken with a broad perspective, which can't really be done when there are only similar perspectives, likes/interests, and characteristics within one dense network.
Turkopticon gives its users the tools to share perspectives in the form reviews that continue to remain solid over time. Sometimes, and it can be assumed that reviews can become too dense–a part of the dense cluster. So using DS can break up that review cluster and bring other informational perspectives into connection with each other. If a dense cluster has been identified and analyzed as EA1, EA2, EA3, then plucking random (but relevant) reviews from disconnected places like S1, E1, N1, can break up congested reviews about requesters from individuals with similar likes, interests, characteristics, etc. It goes both ways too, EA is getting differing perspective reviews from outside its congestion, while places like S, E, and N are getting their information as well. In a way, using DS creates a parallel processing network between these communities of workers. Now, reviews can be seen across the board, especially a review about a requester that was outside that dense cluster. Dismantling AMT with Turkopticon using DS aids

in two important ways: (1) helps Turkopticon as a constitutive to push back against AMT, and (2) helps people see what's up, and to organize beyond Turkopticon.

Algorithmic Pluralism

Diving deeper into how algorithms are taking information from users and places, one of the first questions to be addressed is to whom this information is desirable? In our case it usually involves the creators of AI algorithms, involving feature selection and its epistemic values. Having the creators of algorithms specifically be all white men would be an issue because it not only represents a single view, but when also taken to be a 'view from nowhere', it blocks other points of views and elevates dominant perspectives. Furthermore, being a group that is dominant in power means they have control over what we see (ads and information) and how much of it we may see. Essentially, the 'view from nowhere' reinforces those who have preexisted with power; however, we don't want the 'view from nowhere', so we need to break this view up into a more pluralistic ideology. Taking into consideration how universally reaching we want our AI to be, something like feature selection, as it picks pieces of predictors and variables as a network, should be decentralized and pluralistic. Feature selection is made up of predictors, so when something goes into it that may not be as broad or is oppressive, that comes from the ghost workers themselves. Hancox-Li & Rumar (2021) have suggested a starting point, that instead of trying to make data true we try to make data less false. That would be to expand perspectives–and specifically ghost workers who are a part of minority communities. Less false doesn't mean 'right' or 'wrong', but more a scale of 'less true' and 'more true'. So when making feature selection predictors less false, we'd need to analyze what kind of learning method to use. And since our aim is to connect perspectives using DS, then the method of predictor should be 'K-Means Clustering', which will work to cluster data points that are similar to each other. To make the 'K-Means Clustering' predictor complement DS, it must associate perspectives based on interests, characteristics, and needs that are similar.

Turkopticon's creators laid the foundation for this with ghost workers of AMT by presenting a survey of rights, so to see what AMT isn't doing correctly and what kind of rights they need that will eventually be written by them in the form of reviews. Employers are rated by factors: communicativity,

fairness, promptness, and generosity. These factors are features that can then be the main selection predictors for K-means clustering. This data learning method works to put datasets into non-overlapping subgroups. It can do this using measurements of inter-clusters and intra-clusters. Inter-clusters measure distance between data points across different clusters, while intra-clusters measure the distance of data points within the subgroup cluster (Sreedhar, Kasiviswanath & Reddy, 2017). This lets systems capture both how inter-cluster data points are similar to each other, but also different at the same time (Kashyap & Singh, 2021). After identifying these highly dense clusters, DS can work to spread results from other clusters, by selecting dissimilar inter-cluster results from differing clusters with more similar intra-cluster data points, and then shuffling them together in displayed results.

Using knowledge and outlook of lower class communities, like ghost workers, can introduce new ways of creating universal feature selection predictors that integrate differing perspectives. This can even help us interact with our idea of knowledge–keeping one foot in design and the other in critique, instead of looking at it from a distance. Having the object of study interacting and in conversation with the knower (person) can create more than one way of thinking about that object of study, meaning algorithms will have potential to decentralize into a less false predictor framework without settling on a 'view from nowhere'.

Conclusion

In the current world AI technology has contributed to a disbalance of power towards not only minority groups, but specifically working labor classes. The way technology is built and produced plays a crucial role in how power can be used against communities of workers. Essentially workers and minority communities are most affected by technology's own power structure *and* how it's used. And, because of this, the main focus of this project is on how feminist epistemology can inspire the usage of data feminism: pluralistic, critical, and decentralized, will challenge conventional and centralized models of AI; which affect working labor issues and all minority groups/communities. Looking within AI's design; feature selection, algorithms, and training, leads the way towards understanding AI reflects power imbalances in people's lives. It is important to note the potential harm of not empowering data authors,

compared to letting data authors fill out their own data themselves, and will only push communities down further in our capitalistic society. Similarly, work labor classes face issues involving their own work behind AI training, design, data filtering, and managing, and have been overlooked by corporations/industries and legal mechanisms for labor protections. In order to work towards a decentralized, critical, and pluralistic society, we need to rebuild labor and data collection practices to empower others, not exploit them. And to do this we need to take a look at how deeply interrelated power is with data extraction, algorithms, work class structures, and AI technology itself.

References

Anderson, Elizabeth, "Feminist Epistemology and Philosophy of Science", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/feminism-epistemology/>.

Botella, C., Rueda, S., López-Iñesta, E., & Marzal, P. (2019). Gender Diversity in STEM Disciplines: A Multiple Factor Problem. *Entropy*, 21(1), 30. https://doi.org/10.3390/e21010030

Coeckelbergh, M. (2022). *The political philosophy of AI*. Polity Press.

Chen, A. (2019). How Silicon Valley's Successes Are Fueled by an Underclass of 'Ghost Workers.'. *The Verge*, *13*.

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. The MIT Press.

Irani, L., & Silberman, M. (2013). Turkopticon: interrupting worker invisibility in amazon mechanical turk. *Conference on Human Factors in Computing Systems - Proceedings*, 611–620. https://doi.org/10.1145/2470654.2470742

Eveleth, R. (2019, October 1). *The biggest lie tech people tell themselves - and the rest of Us*. Vox. https://www.vox.com/the-highlight/2019/10/1/20887003/tech-technology-evolution-natural-inevitable-ethics

Ferguson, K. E. (2017). Feminist theory today. *Annual review of political science*, 20(1), 269-286.

Golebiewski, M., & Boyd, D. (2019). Data voids: Where missing data can easily be exploited.

Hancox-Li, L., & Kumar, I. E. (2021). Epistemic values in feature importance methods: Lessons from feminist epistemology. *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 817–826. https://doi.org/10.1145/3442188.3445943

Kashyap, S., & Singh, A. (2021). Testing Data Feminism in India. *Scholars Journal of Arts, Humanities and Social Sciences*, *9*(10), 516–530. https://doi.org/10.36347/sjahss.2021.v09i10.006

Keyes, O., & Creel, K. (2022). Artificial Knowing Otherwise. *Feminist Philosophy Quarterly*, *8*(3/4). https://doi.org/10.5206/fpq/2022.3/4.14313

Sreedhar, C., Kasiviswanath, N. & Chenna Reddy, P. Clustering large datasets using K-means modified

inter and intra clustering (KM-I2C) in Hadoop. *J Big Data* 4, 27 (2017).

https://doi.org/10.1186/s40537-017-0087-2