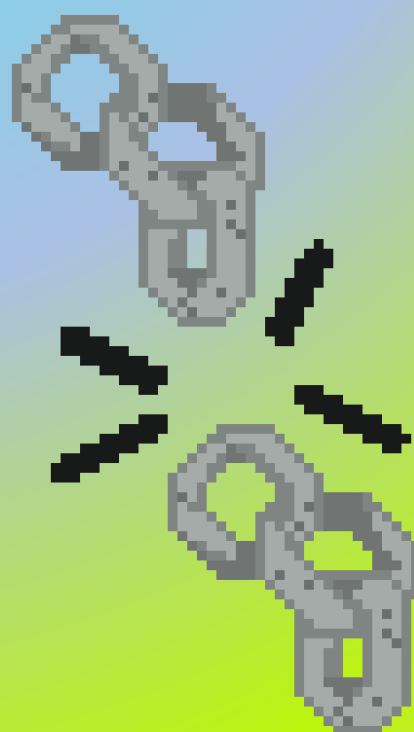# proletariAItes of the world, unite!

you have nothing to lose but your #time

Edo Barbieri

Intro

Datafication of Society is the process through which *big* (huge in volume), *bad* (non-representative of the whole population), *messy* (in weird formats, overly complicated and full of errors) *data* is created, often automatically, on most processes happening in our urban life. This data is often analysed using *AI* and, especially, *Machine Learning algorithms*. These are, in plain English, decision-making computer programmes that should mimic human intelligence and, in the case of Machine Learning, that are *not* developed step by step by someone with some knowledge on the process being analysed; on the contrary, ML algorithms make decisions by (`quite literally`) mechanically looking for patterns in the data, so trying to generalize them. For example, let's assume that we are interested in finding the preferred pronouns of the person we are talking to. A non-automated, non-AI way to do that would simply be to ask directly to them; a Machine Learning approach may be having some kind of sensor registering characteristics of the person we are talking to, and then see which preferred-pronouns subpopulation they are more similar to.

the   personal   is   political
I think that this example is actually quite good to reflect on the aim of this self-publication; decision-making algorithms are not magically embedded with the `truth`, they make a bunch-a errors and have the potential to influence the way in which people experience their everyday life. Reflecting on the preferred pronouns example, let's just imagine that someday the principal of a given university will decide to implement some kind of algorithm to deduce students' preferred pronouns. Given that the group of people sharing the same pronoun is probably not gonna be a monolith, it is very likely that some people would be incorrectly classified as users of a different pronoun, going against their self-determination and potentially perpetuating gender dysphoria.

data to the people, cuz data is the   power
If our spaces of everyday life are gonna be affected by digitalization, we should all get access to the tools to understand and be critical of what's happening. Unfortunately, the tools needed to make such reflections are not evenly and fairly distributed among the general population; as often happens when dealing with technical info (as well

as with basically anything that is *commodifiable*, that can be marketed), gatekeeping is king. And when it is not in the shape of a website paywall, an entry fee or a paper subscription, it looks more like a (usually) cis, straight, able-bodied, wealthy white man in STEM telling you how much time you will waste not getting his beloved subject. The status quo has been used these digital tools for so much, so why the hell shouldn't we do that?

```
proletariAItes    of    the    world,
                unite!
```

The status quo owns the means of digitalization, we should seize them and distribute them to everyone. That sounds like a pretentious aim, innit? In order to at least `try` to start doing something on that front, I decided to share some of my knowledge on the subject in this publication; I am not doing so to force my vision of technology on you- This should be seen more like an invitation for you readers to reflect on some examples of bad AI, while I'll try to introduce concepts and vocabulary that are deffo used more inside unis rather than outside. To do so, I'll try to engage with examples from the real world- giving also some space for personal reflection. Hope you'll learn something from it and, ofc, have fun (and just the right dose of anger) while reading it.
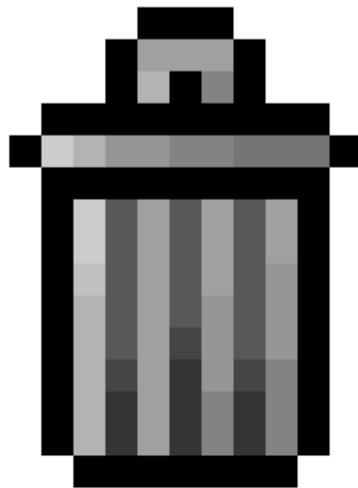
Garbage in, Garbage out

As I was saying in the intro, the more we (supposedly) advance as a society, the more data are collected on us. When talking about their collection, we usually distinguish between *Small Data*- that are collected with a purpose in mind, and *Big Data*- which are usually collected `just because`. This `just-becausedness` in their collection is usually justified by the idea that the more volume of data you can collect, the more information you can get. Which does sounds correct, right? Unfortunately, this concept (that stat people like to call *Law of Large Numbers*) holds true only when there's no social group who gets more represented in the dataset; but when that happens, all results computed from these datasets are gonna be carried by these overrepresented people. Because of the *training*, the learning, that these algorithms had on biased data, all their outcomes are gonna be biased; even a small association between the characteristic you wanna study and the over (or under) representation of a group may lead to low-quality (`i.e., useless`) results.

Woop-woop! That's the data of da police!

Here, I wanna introduce to you the example that will follow us through the rest of the zine: predictive policing. I'll try to explain a wee bit how it works- if you want more details, here's an article [https://www.fastcompany.com/90601641/predictive-policing-algorithms]. In the '90s, NY cops started to use an algorithm, called CompStat, to predict what areas to patrol more. The way in which it worked was by mapping only certain types of crimes, and then working out some kind of *risk measure*, meaning likelihood of other crimes to happen. As of right now, there are more and more type of policing algorithms available, but most of them work in the same way; collect data on some type of crime, compute how risky the area is. The problem here being that most policing algorithms are based only on crimes that are more reported when committed by non-white people, so racialized areas are more prone to be represented (hurray! representation!). You can easily imagine how this is not a way to have a safer city; the only thing created by these algos is a cycle of policing areas that are already *oversampled*, from which data is already collected disproportionally, creating new interactions with the police, and so more data collection. (nooo! representation!)

TLDR; being "AI", "Computer" or "Science" doesn't mean being good, true or objective. Next time a decision regarding yourself is gonna go through an AI, think about the dataset on which the algorithm is trained and how its characteristics may influence its decision.

If you wanna reflect more on the topic, I'd suggest reading the following article [https://www.vox.com/science-and-health/2019/1/23/18194717/alexandria-ocasio-cortez-ai-bias], hope you're gonna enjoy it (y)!

Partial Vision

As of right now, we have mainly talked of algorithms that create problems because of the type of data they are trained on is of bad quality. Sometimes tho (usually both things happen together to some extent) the problem does not lie on the data itself, but on the way in which the algorithm has been designed. When designing predictive AI, the characteristics you decide to embed in your `algo` are gonna be decisive of its performance; in the end, what you're trying to do is to approximate a hidden attribute of interest with other ones that you can observe. Coming back to our beloved CompStat example, in the end what you're doing is to approximate the risk level on a given area with data from crime records; choosing to include or not a type of crime would significantly change the analysis. For example, deciding to focus on drug-related crimes would result in an oversurveillance similar to the one that I described in the previous section. On the other hand, focusing on white collar crimes would create a totally different risk map. These AI are literally shaping the space in which we live; different choices of design give you different perspectives on the topic you're analysing.

We usually refer to this concept as *partial vision*, meaning that our specific background, the society we're from, and the spaces we interact with are gonna guide our design choices. Before, I said that garbage data generates garbage results; `but garbage design choices are deffo not gonna help.` As an AI programmer, you have the power to tackle this subjectivity by reflecting on the choices you're making, or by allowing people with different backgrounds in the algorithm construction; as people subjugated to its choices tho, the only thing that we can do is to be aware of this subjectivity.

TLDR; designing an algorithm means performing some choices, and these choices shape it so that it reflects the society from where the designers are from. Nex time a decision regarding yourself is gonna go through an AI, think about who designed it, and how their presence may be reflected in its output.

If you wanna reflect more on the topic, I'd suggest reading the following article [https://towardsdatascience.com/dont-blame-the-ai-it-s-the-humans-who-are-biased-d01a3b876d58], hope you're gonna enjoy it (y)!

Black Boxes

So far, we saw how garbage data and garbage algorithms may influence people's lives; all these reflections were made possible because of my understanding of the example I decided to present. Unfortunately, this is usually limited by the *transparency* of the `algo` itself; private companies developing AI will not usually tell you the specific characteristics present in their code, as well as how much each of those has an influence on the final output. When talking about Machine Learning algorithms, transparency is even less, since they are created through *training processes* that are not meant to tell us why a decision has been performed, but only to announce to us the decision itself. This is what we usually call a *Black Box*, an opaque decision `algo`. This made a bunch-a sense when these `algos` were mainly used and developed for marketing application; `but since we live in a society,` we then started to apply 'em to more and more social settings. The reason why someone thought that it would be a clever idea to influence people's lives using something that is at least not clear to the general public, if not to the designers themselves, is still a mystery to me. Especially since it is still possible to design clearer `algos` that perform very similarly to their opaque counterparts.

So, how do predictive policing and black boxes get on together? Quite nicely, unfortunately. If we now think about both reflections I gave in this zine, they have transparency as their basis; the only way in which it was possible for me to do those was because I knew 'bout the type of features on which the `algos` were trained on. Sadly, it seems that most of the AIs used in this field are proprietary, meaning that they're own by someone who, in order sell them, keeps their details as a secret. Through these algorithms, cops are creating a space where people are getting more or less policed based on choices that aren't that clear to anyone, and one of the reasons why this is happening seems to be making some `rando Technocrat` richer.

TLDR; all algorithms have different degrees of interpretability, and probably less-interpretable algorithms should not be used to influence one's social life. Here, I should probably invite you to discover how interpretable the algorithms you encounter are; unfortunately, there is no real way to do that without going too much into technicalities.

Now, `inside you there are (or at least there should be by now) two wolves,` two paths that you can decide to follow: you can either decide to engage with different types of examples of black boxes and see their impact, or you can decide to start getting a more technical (and, I warn y'all, potentially maths-heavy) understanding of what causes these black boxes (or I mean, both; or neither). I thought a lot on what article to present for this section and what I come up with is this one [https://www.jacobinmag.com/2016/09/big-data-algorithms-math-facebook-advertisement-marketing] where Cathy O'Neill presents her book in which theories *Weapons of Maths Destruction*, a specific class of algorithms in which being a black box is in their definition. It definitely focuses more on the social aspect than the technical one (being an article on Jacobin and not on Nature), but I think it can be a nice starting point to whatever `wolf` you decide to listen to. Hope you're gonna enjoy it (y)!
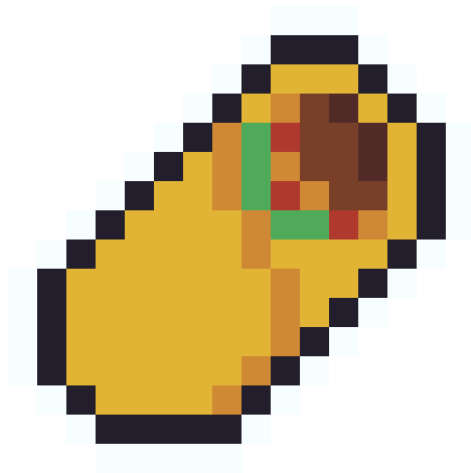
Wrap-ups

A concept that I have yet not defined but that it is present throughout the whole publication is the one of *Positivism*, that is the belief that the only way to understand something is by observing it in its objectivity, without any sort of interaction with it. This belief can be seen as a very simplistic way to justify black boxes; we can't really explain what's happening here, but we don't really need it since all our observations come from the `Objective and True World`™. Why should we then investigate the structure of the decision model, if the model is observing the space in its truest form?

As we have seen tho, this doesn't really work that well when discussing AI and, particularly, their social applications (`or maybe it just doesn't work?`); the space observed by the `algo` is shaped by the designer, deciding what characteristic of the real world to include in the computation of the output. Plus, the space is also shaped by the algorithm itself, since social AI have often a direct impact on the society; not reflecting on that would mean missing a huge part of the story.

As Uncle Ben almost said, `with Big Data comes great responsibility`. Sadly, lots of the people that are designing, training and applying these algorithms are too rooted on their positivist view, and just collectively decided not to be held accountable for their `algos`. I hope that I gave you some tools to understand better these concepts in a bigger framework, and to bother the Status Quo by questioning their power. In the future, there may be a movement to force Technocrats to design better AI, in the meantime I hope that these concepts will help you to live a more aware, whole, and hopefully joyful, presence in the social world- see you then!

"who produces information? How? For whom? And who consumes it?"