POTs: The revolution will not be optimized?

Seda Gürses [1] Rebekah Overdorf [1] Ero Balsa [1]

1. The Optimization Problem

In the 90s, software engineering shifted from packaged software and PCs to services and clouds, enabling distributed architectures that incorporate real-time feedback from users. In the process, digital systems became layers of technologies metricized under the authority of objective functions. These functions drive selection of software features, service integration, cloud usage, user interaction and growth, customer service, and environmental capture, among others. Whereas information systems focused on storage, processing and transport of information, and organizing knowledge —with associated risks of *surveillance*— contemporary systems leverage the knowledge they gather to not only understand the world, but also to *optimize* it, seeking maximum extraction of economic value through the capture and manipulation of people's activities and environments.

The ability of these *optimization systems* to treat the world not as a static place to be known, but as one to sense and co-create, poses social risks and harms such as social sorting, mass manipulation, asymmetrical concentration of resources, majority dominance, and minority erasure. In the vocabulary of optimization, these harms arise due to choosing inadequate *objective functions* that, among other things, 1) aspire for antisocial or negative environmental outcomes (Madrigal, 2018), 2) have adverse side effects (Lopez, 2018), 3) are built to only benefit a subset of users (Lopez, 2018), 4) externalize risks associated with environmental unknowns and exploration to users and their surroundings (Bird et al., 2016)[†], 5) are vulnerable to distributional shift, wherein a system that is built on data from a particular area or domain is deployed in another environment that it is not optimized for (Angwin et al., 2016), 6) spawn systems that exploit states that can lead to fulfillment of the objective function short of fulfilling the intended effect (Harris, 2018), and 7) distribute errors unfairly (Hardt, 2014; Amodei et al., 2016). Common to information and optimization systems is their concentration of both data and processing resources, network effects and ability to scale services that externalize risks to populations and environments. Consequently, a handful of companies are now able to amass enormous power.

To better illustrate the difference between information and optimization systems and the problems the latter pose, in the rest of the paper we focus on location based services (LBS). LBS have moved beyond tracking and profiling individuals to generate spatial intelligence to leveraging this information to manipulate users' behavior and create "ideal" geographies that optimize space and time to customers' or investors' interests (Phillips et al., 2003). Population experiments drive iterative designs that ensure sufficient gain for a percentage of users while minimizing costs and maximizing profits.

For example, LBS like Waze provide optimal driving routes that put users in certain locations at a disadvantage. Waze often redirects users off major highways through suburban neighborhoods that cannot sustain heavy traffic. While useful for drivers, it affects neighborhoods by making streets busy, noisy and less safe. Consequently, towns may need to fix and police roads more often. This example shows that even when users benefit, *non-users* may bear the ill effects of optimization. Users within a system may also be at a disadvantage due to their location. Pokémon Go users living in urban areas see more Pokémon, *Pokéstops*, and *gyms* than users in rural areas. Uber manipulates prices in space and time, constituting geographies around supply and demand that both drivers and riders are unable to control, negatively impacted by price falls and *surges*, respectively. Recent studies report that Uber drivers (who work on commission, sharing a part of the revenue from every ride with the company) make less than minimum wage in many jurisdictions.

Disadvantaged users have worked from within the system to tame optimization in their favor, e.g., by strategically feeding misinformation to the system in order to change its behavior. Neighborhood dwellers negatively affected by Waze's traffic redirection have fought back by reporting road closures and heavy traffic on their streets —to have Waze redirect users out of their neighborhoods. Some Pokémon users in rural areas spoof their locations to urban areas (using stealthy techniques). Other users report

[1] KU Leuven, Belgium. Correspondence to: Seda Gürses <sguerses@esat.kuleuven.be>.

[†] We disagree with this paper's premise that optimization systems will lead to 'optimal' outcomes, with experimentation as its only potential externality.

to OpenStreetMaps —used by Pokémon Go— false footpaths, swimming pools and parks, resulting in higher rates of Pokémon spawn in their vicinity. Uber drivers have colluded to induce *surge* pricing and temporarily increase their revenue by simultaneously turning off their apps, inducing surge, and turning the app back on to take advantage of the increased pricing in the area.

While the effectiveness of these techniques is unclear, they inspire the type of responses that a more formal approach may provide. In fact, these responses essentially constitute adversarial machine learning, seeking to bias system responses in favor of the "adversary". The idea of turning adversarial machine learning around for the benefit of the user is already prevalent in PETs literature (McDonald et al., 2012; Cherubin et al., 2017). It is in fact in the spirit of PETs that we attend to the optimization problem, i.e., we explore ideas for technologies that enable people to recognize and respond to the negative affects of optimization systems.

2. POTs

Optimization systems infer, induce, and shape events in the real world to fulfill objective functions. *Protective optimization technologies* (POTs) reconfigure these events as a response to the effects of optimization on a group of users or local environment. POTs analyze how events (or lack thereof) affect users and environments, then manipulate these events to influence system outcomes, e.g., by altering the optimization constraints and poisoning system inputs.

To design a POT, we first need to understand the optimization system. What are its user and environmental inputs (U, E) and how do they affect the capture of events? Which are the outcomes $O = F(U, E)$ that are undesirable for subpopulations or environments? With a characterization of the system, as given by $F(U, E)$, we identify those who benefit from the system and those placed at a disadvantage. We define a benefit function, $B(X, E) : (x, e, Value) \rightarrow value$ that may include both users and non users $(U \subset X)$ and an environment $E' \supseteq E$. The disadvantaged are those people and environments that reside in local minima of B. We then set an alternative output $B(X, E', Value') : (x, e) \rightarrow value'$ the POT aims to achieve.

A POT benefit function B may attend to different goals. A POT may attempt to *"correct" imbalances* optimization systems create, i.e., by improving a systems' outcomes for populations put at a disadvantage. Conversely, it may also strategically attempt to reverse system outcomes as a form of *protest*, highlighting the inequalities these systems engender. This further hints at the subversive potential of POTs. POT designers may concoct a B to *contest the authority* of optimization systems, challenging the underlying objective functions these systems optimize to and their very *raison d'être*. To do that, a POT may attempt to sabotage or boycott the system, either for everyone or for an impactful minority that are more likely to affect change, leveraging the power asymmetries the POT precisely intends to erode.

Once we select B, we must choose the techniques that implement it. These techniques involve changes to the inputs that users have control over and alterations to constraints over the objective function to reconfigure event capture (i.e., the system's mechanism of detection, prediction, and response to events). Lastly, we deploy and assess the impact of the POT both in terms of local and global effects on users and environments as intended by B and tweak it as necessary.

We note that POTs may elicit a counter response from the optimization systems they target. The latter may either attempt to neutralize POTs or expel those deploying them from the system. Anticipating these responses may require POT designers to aim for *undetectability*, e.g., by identifying minimum alterations to inputs and constraints, or optimizing constraints to prevent detection.

3. Discussion

As with PETs, POTs come with moral dilemmas. Some compare to concerns raised by obfuscation-based PETs, although the latter focus on privacy and not the protection of populations and environments from optimization. In their work on obfuscation, (Brunton & Nissenbaum, 2015) highlight three ethical issues: dishonesty, wasted resources and polluted databases. Since optimization systems are not about knowledge, one could argue using POTs cannot be judged as dishonesty but as introducing feedback into the cybernetic loop to get optimization systems to recognize and respond to their externalities. POTs are likely to come at a greater cost to the service providers, and may give rise to negative externalities that simply impact different subpopulations and environments. In fact, all of the issues that we discussed as harmful effects of optimization systems can be replicated in POTs: they may have an antisocial objective function, have serious side effects, benefit a few and so on. Seen that way, it is possible to argue that if optimization is the problem, then more optimization may even come to exacerbate it.

This paper intends to provoke questions around optimization systems and their many impacts on society. Optimization history is also one of counter-optimization as evident in the case of search engine optimization or spammers. POTs ensure that counter-optimization is not only available to a privileged few. One could argue we should just design better optimization systems. That, however, requires working within the system; POTs explore technical avenues for intervening from outside of the system. While short of a revolution, they bring inhabitants into the negotiations of how their environments are organized.

Acknowledgements

We are indebted to Martha Poon for her original framing of the optimization problem and to Jillian Stone for her empirical insights into Pokemon Go.

This work was supported in part by the Research Council KU Leuven: C16/15/058. In addition, this work was supported by the European Commission through KU Leuven BOF OT/13/070 and H2020-DS-2014-653497 PANORAMIX and generously supported by a Research Foundation - Flanders (FWO) Fellowship.

References

Amodei, Dario, Olah, Chris, Steinhardt, Jacob, Christiano, Paul, Schulman, John, and Mané, Dan. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*, 2016.

Angwin, Julia, Larson, Jeff, Mattu, Surya, and Kirchner, Lauren. Machine bias: There's software used across the country to predict future criminals. and it's biased against blacks. ProPublica, May 2016. URL https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Bird, Sarah, Barocas, Solon, Crawford, Kate, Diaz, Fernando, and Wallach, Hanna. Exploring or exploiting? social and ethical implications of autonomous experimentation in ai. ssrn scholarly paper id 2846909. *Social Science Research Network, Rochester, NY.*, 2846909, 2016.

Brunton, Finn and Nissenbaum, Helen. *Obfuscation: A user's guide for privacy and protest*. Mit Press, 2015.

Cherubin, Giovanni, Hayes, Jamie, and Juarez, Marc. Website fingerprinting defenses at the application layer. *Proceedings on Privacy Enhancing Technologies*, 2017(2): 186–203, 2017.

Hardt, Moritz. How big data is unfair? Medium, September 2014. URL https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de.

Harris, Malcolm. Glitch capitalism: How cheating ais explain our glitchy society. New York Magazine, April 2018. URL https://nymag.com/selectall/2018/04/malcolm-harris-on-glitch-capitalism-and-ai-logic.html.

Lopez, Steve. On one of l.a.'s steepest streets, an app-driven frenzy of spinouts, confusion and crashes. Los Angeles Times, April 2018. URL https://www.latimes.com/local/california/la-me-lopez-echo-park-traffic-20180404-story.html.

Madrigal, Alexis C. The perfect selfishness of mapping apps. The Atlantic, March 2018. URL https://www.theatlantic.com/technology/archive/2018/03/mapping-apps-and-the-price-of-anarchy/555551/.

McDonald, Andrew WE, Afroz, Sadia, Caliskan, Aylin, Stolerman, Ariel, and Greenstadt, Rachel. Use fewer instances of the letter "i": Toward writing style anonymization. In *International Symposium on Privacy Enhancing Technologies Symposium*, pp. 299–318. Springer, 2012.

Phillips, David, Curry, Michael, and Lyon, D. Privacy and the phenetic urge. *Surveillance as social sorting: Privacy, risk and digital discrimination*, pp. 137–152, 2003.