In: Emre Bayamlıoğlu, Irina Baraluic, Liisa Janssens and Mireille Hildebrandt (eds), BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen, 2018: Amsterdam University Press, 24-28. DOI 10.5117/9789463722124/CH3.

Stirring the POTs: Protective Optimization Technologies[1]

Seda Gürses, Rebekah Overdorf**, Ero Balsa****

Abstract

With the introduction of machine learning, information systems and services are increasingly produced under the logic of optimization. Optimization systems infer, induce, and shape events in the real world to fulfil objective functions. 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. Protective optimization technologies (POTs) are a response to these harms and intend to reconfigure the capture of these systems in order to mitigate their effects on a group of users or local environment. POTs analyse 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. Most importantly, POTs aim to explore and provide technical avenues for people to intervene from outside of optimization systems.

Keywords: Protective optimization technologies, optimization systems, location based services (LBS).

Introduction

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 (Gürses and van Hoboken 2018). In the process, digital systems became layers of technologies metricized under the authority of optimization functions. These functions drive the 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 Optimization Problem

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 mathematical vocabulary, optimization is about finding the best values for an `objective function'. The externalities of optimization occur due to the way that these objective functions are specified (Amodei et al. 2016). These externalities include:

1) Aspiring for asocial behavior or negative environmental ordering (Madrigal 2018, Cabannes et al. 2018),
2) Having adverse side effects (Lopez 2018),
3) Being built to only benefit a subset of users (Lopez 2018),
4) Pushing risks associated with environmental unknowns and exploration onto users and their surroundings (Bird et al. 2016)[2],
5) Being vulnerable to distributional shift, wherein a system that is built on data from a particular area is deployed in another environment that it is not optimized for (Angwin et al. 2016),
6) Spawning systems that exploit states that can lead to fulfillment of the objective function short of fulfilling the intended effect (Harris 2018),
7) Distributing errors unfairly (Hardt 2014), and
8) Incentivizing mass data collection.

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, today a handful of companies are able to amass enormous power.

In the rest of this provocation we focus on location based services (LBS). LBS have moved beyond tracking and profiling individuals for generating spatial intelligence to leveraging this information to manipulate users' behaviour 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 promote individual gain at the cost of generating more congestion. Waze often redirects users off major highways through suburban neighbourhoods that cannot sustain heavy traffic. While useful for drivers, neighbourhoods are made busier, noisier and less safe, and towns need to fix and police roads more often. Even when users benefit, non-users may bear the ill effects of optimization.

Users within a system may also be at a disadvantage. Pokémon Go users in urban areas see more Pokémon, Pokéstops, and gyms than users in rural areas. Uber manipulates prices, constituting geographies around supply and demand that both drivers and riders are unable to control while being negatively impacted by price falls and surges, respectively. Studies report that Uber drivers (who work on commission) make less than minimum wage in many jurisdictions.

Disadvantaged users have developed techniques to tame optimization in their favour, e.g., by strategically feeding extra information to the system in order to change its behaviour. Neighbourhood 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 neighbourhoods. Some Pokémon users in rural areas spoof their locations to urban areas. Other users report 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 temporarily increase their revenue by simultaneously turning off their apps, inducing a local price surge, and turning the app back on to take advantage of the increased pricing.

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 favour of the “adversary”. The idea of turning adversarial machine learning around for the benefit of the user is already prevalent in Privacy Enhancing Technologies (PETs) literature, e.g., McDonald 2012. It is 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 effects of optimization systems.

Introducing POTs

Protective optimization technologies (POTs) respond to optimization systems' effects on a group of people or local environment by reconfiguring these systems from the outside. POTs analyse how capture of events (or lack thereof) affect users and environments, then manipulate these events to influence system outcomes, e.g., altering optimization models and constraints or poisoning system inputs.

To design a POT, we first need to understand the optimization system. What are its user and environmental inputs (U,E)? How do they affect the capture of events? Which outcomes O = F(U,E) 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 by defining a benefit function, B(X, E'):(x,e', Value) ⇾ value that includes both users and non users ($U \subset X$) and affected environments($E \subseteq E'$). The disadvantaged are those people and environments that reside in local minima of B and are gravely impacted by the system. We then set an alternative output B(X, E', Value'):(x,e) ⇾ value' the POT aims to achieve.

A POT's benefit function may attend to different goals (Figure 1). It may attempt to “correct” imbalances optimization systems create, i.e., by improving systems' outcome for populations put at an --often historically continuous-- 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 strategy to produce an alternative to 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 effect change, leveraging the power asymmetries the POT precisely intends to erode.

In: Emre Bayamlıoğlu, Irina Baraluic, Liisa Janssens and Mireille Hildebrandt (eds), BEING PROFILED:COGITAS ERGO SUM. 10 Years of Profiling the European Citizen, 2018: Amsterdam University Press, 24-28. DOI 10.5117/9789463722124/CH3.

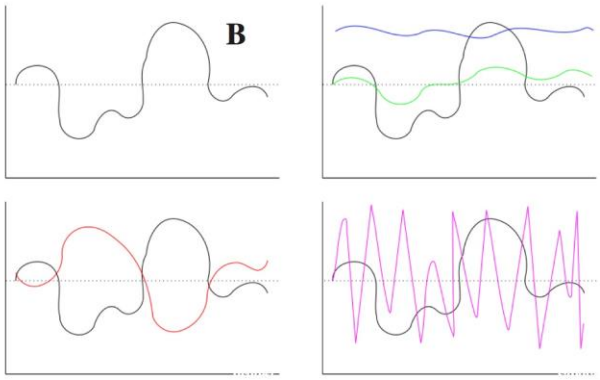

Figure 1. Benefit function (top left). POT strategies include redistribution (top right), protest (bottom left), sabotage (bottom right).

Once we select a strategy, 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 and tweak it as necessary.

We note that POTs may elicit a counter response from the optimization systems they target, with service providers either neutralizing their effect or expelling POT users. Anticipating these responses may require POT designers to aim for stealth or undetectability, e.g., by identifying minimum alterations to inputs or optimizing constraints to prevent detection.

Discussion

POTs come with moral dilemmas. Some of these compare to concerns raised by obfuscation-based PETs, although these focus on protecting privacy and not protecting populations and environments from optimization. In their work on obfuscation, Brunton and Nissenbaum (2015) highlight four ethical issues: dishonesty, polluted databases, wasted resources and free riding.

Since optimization systems are not about knowledge, we may 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 service providers and give rise to negative externalities that impact different subpopulations and environments. In fact, all of the harmful effects of optimization systems may be replicated: POTs may have asocial objective functions, negative side effects, etc. One may argue that if optimization is the problem, then more optimization may even come to exacerbate it. Moreover, POTs users may be seen as free riders. These are serious concerns, especially since whichever benefit function B we choose, there will be users who do not agree with or are harmed by the POT. Yet, this problem is inherent to optimization systems' externalities, especially when users are free-riding on non-users or on existing infrastructure.

Banging on POTs: a digital cacerolazo

Optimization history is also one of counter-optimization as evident in the case of search engine optimization or spammers. As optimization systems spread, POTs ensure that counter-optimization is not only available to a privileged few. One could insist that we should work within the system to design better optimization systems. Given service providers' track record in not responding to or recognizing their externalities, POTs aim to explore and provide technical avenues for people to intervene from outside these systems. In fact, POTs may often be the only way users and non-users can protect themselves and secure better outcomes. While short of a revolution, POTs bring people into the negotiations of how their environments are organized. They also help to provoke a popular response to optimization systems and their many impacts on society.

Notes

* Seda Gürses is an FWO post-doctoral fellow at COSIC in the Electrical Engineering department of KU Leuven where she is a member of the privacy group
** Rebekah Overdorf is a post-doctoral fellow at COSIC in the Electrical Engineering department at KU Leuven.

*** Ero Balsa is a PhD candidate at COSIC in the Electrical Engineering department of KU Leuven, where he is a member of the privacy group
[1] We are indebted to Martha Poon for her original framing of the optimization problem and to Jillian Stone for her empirical insights into Pokémon Go. This work was supported in part by the Research Council KU Leuven: C16/15/058; 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.
[2] We disagree with this paper's premise that optimization systems will lead to 'optimal' outcomes, with experimentation as its only potential externality – we appreciate their highlight of the latter.

References

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

Angwin, Julia, Larson, Jeff, Mattu, Surya, and Kirchner, Lauren. 2016 "Machine bias", *ProPublica*, May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Bird, Sarah, Barocas, Solon, Crawford, Kate, Diaz, Fernando, and Wallach, Hanna. 2016. "Exploring or exploiting? Social and ethical implications of autonomous experimentation in AI", *SSRN Paper* ID 2846909.

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

Cabannes, Théophile, Marco Antonio Sangiovanni Vincentelli, Alexander Sundt, Hippolyte Signargout, Emily Porter, Vincent Fighiera, Juliette Ugirumurera, and Alexandre M. Bayen. 2018. "The impact of GPS-enabled shortest path routing on mobility: a game theoretic approach". Presented at Transportation Research Board 97th Annual Meeting. Washington DC, USA. 7–11 January 2018.

Gürses, Seda and van Hoboken, Joris. Privacy after the Agile Turn. 2018. In Selinger et al. (eds), Cambridge Handbook of Consumer Privacy, Cambridge University Press.

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, 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. 2018. "The perfect selfishness of mapping apps." *The Atlantic*, March 2018. 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.2012. Use fewer in- stances of the lettr "i": Toward writing style anonymization. In PETS Symposium, pp. 299–318. Springer.

Phillips, David, and Michael Curry,. 2002. Privacy and the phenetic urge. In D. Lyon (ed.), Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination. NY: Routledge.