Camera Adversaria

Kieran Browne
Australian National University
Canberra, Australia
kieran.browne@anu.edu.au

Ben Swift
Australian National University
Canberra, Australia
ben.swift@anu.edu.au

Terhi Nurmikko-Fuller
Australian National University
Canberra, Australia
terhi.nurmikko-fuller@anu.edu.au

ABSTRACT

In this paper we introduce Camera Adversaria; a mobile app designed to disrupt the automatic surveillance of personal photographs by technology companies. The app leverages the brittleness of deep neural networks with respect to high-frequency signals, adding generative adversarial perturbations to users' photographs. These perturbations confound image classification systems but are virtually imperceptible to human viewers. Camera Adversaria builds on methods developed by machine learning researchers as well as a growing body of work, primarily from art and design, which transgresses contemporary surveillance systems. We map the design space of responses to surveillance and identify an under-explored region where our project is situated. Finally we show that the language typically used in the adversarial perturbation literature serves to affirm corporate surveillance practices and malign resistance. This raises significant questions about the function of the research community in countenancing systems of surveillance.

Author Keywords

surveillance capitalism; adversarial examples; critical design.

CCS Concepts

•**Security and privacy** → **Human and societal aspects of security and privacy;** •**Computing methodologies** → **Object identification;** •**Human-centered computing** → *Interaction design theory, concepts and paradigms;* Smartphones;

INTRODUCTION

Surveillance is no longer the exclusive domain of despots and totalitarians; surveillance is big business [51]. Every day products are launched that find new sources of data to harvest. Society is accreting information at an ever-increasing pace.

Contemporary surveillance systems are composed of many technologies. Networked media, mobile computing devices, and deep learning (DL) are all necessary technologies enabling contemporary surveillance practices.

In this paper we ask how human-computer interaction (HCI) can use design to give some power back to the subjects of corporate surveillance. We introduce *Camera Adversaria*, a mobile app designed to seamlessly replace a smartphone camera application. The app mimics a standard camera application in appearance and user experience but includes software that manipulates images to disrupt DL image classification. The application applies a filter to users' photographs that is mostly imperceptible to the human eye but confounds DL image classification systems. The image processing technique, known as an *adversarial perturbation*, was discovered by DL researchers in 2014 [48] but has yet to be made accessible to the general population. Instead, DL researchers have typically maligned the use of such methods as "attacks". We question why DL researchers have failed to identify the potential of such adversarial peturbations for resisting surveillance. By reframing DL as a key component of the contemporary surveillance apparatus, we accordingly work to reclaim the adversarial perturbation as a *defensive* method for those subject to surveillance.

By intervening at the source of the data, we allow users to continue to post photos to social media and backup images in cloud based services while making their images difficult to read with machines. In this way, users are empowered to make choices about their visibility to corporate surveillance without having to opt-out of such systems completely. As a community/discipline concerned with user choices in the presence of power differentials, this under-explored design space which warrants further exploration by HCI researchers.

The *Camera Adversaria* app is a work of critical design. It is intended to highlight issues of corporate surveillance and the machine readability of images, and to offer users an unobtrusive tool of resistance to said surveillance, laying the groundwork for resolving the "privacy paradox".

There has been some discussion of these issues within the HCI community, especially as a site of speculative design (for example in [28]). Critiques of surveillance are also abundant in broader discourse, from Foucault's *Discipline and Punish* [16] to Orwell's *1984* [39] to artworks like James Bridle's *Every CCTV Camera (CC)* [6].

Camera Adversaria, however, is not primarily an aesthetic work, nor is it "speculative design". Instead, the app presents an immediately useful solution to the contemporary problem posed by a particular incarnation of surveillance capitalism. The app is currently available to download for free from the

CHI '20, April 25–30, 2020, Honolulu, HI, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376434

Google Play store; the source code is published open-source on GitHub [7] under the Eclipse Public Licence v1.0.

Bardzell & Bardzell [4] define critical design for the HCI community as "a research through design methodology that foregrounds the ethics of design practice" and "make[s] consumers more critical of their everyday lives". Its discourse draws on the "critical theory" of the Frankfurt school. Dunne and Raby—the originators of critical design—focus on industrial design for its position "at the heart of consumer culture" and capitalism [14, p. 45]. Camera Adversaria exists in a related design space, but its critique is of a newer incarnation of economic power; "surveillance capitalism" [51]. Adjacent design practices; adversarial design [13] and obfuscation [9], intersect with the aims of this project. Obfuscation, as theorised by Brunton and Nissenbaum, engages directly with privacy concerns of networked media. Though a broad definition of this practice might include a project like Camera Adversaria, Brunton and Nissenbaum focus on vernacular, often low-tech modes of resistance. Similarly, adversarial design as concieved by DiSalvo, is a political practice like critical design but adopts agonism rather than criticism as its primary framing device. Again, construed in some ways, adversarial design could describe our project; we too use our artefact to create a "space for contestation" and "dissensus" [13, p. 9,12]. Finally, this paper serves as an exemplar of how critical design researchers in HCI can re-purpose adversarial technologies (even those construed as hacks and attacks) in order to support resistance to surveillance and other dominant technological paradigms.

SURVEILLANCE CAPITALISM

Until recently two metaphors, Big Brother and the Panopticon, dominated theoretical work on surveillance [15]. Both tie conceptualisation of surveillance to totalising systems and to the state. There is an according priviledging of state forms of surveillance in the literature and arts practices. However, these metaphors are poorly equipped to deal with the creeping growth of non-state surveillance by technology companies in recent years. The rise of corporate data collection has created a system of surveillance oriented by economic power rather than discipline and punishment. There have been a number of attempts to name and describe this phenomenon (e.g. [50], [12]). This paper takes Shoshana Zuboff's "surveillance capitalism" [51] as its theoretical point of departure.

Zuboff clarifies that surveillance capitalism is not a technology but a logic, albeit one that "imbues technology and commands it into action" [51, p. 25]. Central to Zuboff's account is a rejection of technological determinism. Systems of power—economic and otherwise—direct technological development. She argues, following Max Weber, that technological development is largely oriented by economics and profit-making [51, p. 27]. As such, any account of technological development must consider its broader place in a system of economic relations.

The privacy paradox

HCI is increasingly examining its own socio-cultural implications and those of the broader technical and research community—particularly with regards to marginalised communities [4]. For this reason it is incumbent upon HCI researchers to show that it is the interests of those communities they are attempting to represent.

This paper will critique surveillance practices within the technology industry. Much existing critical discourse operates in defence of privacy, however there remains controversy as to whether people really value privacy at all. Research reveals a "paradoxical dichotomy between attitudes and behaviours" with regards to privacy [27]. When surveyed, users often claim to be concerned about sharing sensitive data about their behaviour with third parties such as technology companies. However, this wariness is not necessarily borne out in their actual behaviour. Users overwhelmingly choose to use technology and services which require them to exchange their privacy for social capital and convenience. This phenomenon is termed *the privacy paradox*.

The privacy paradox has been used to argue that users either do not care about privacy, or that they consider that the benefits afforded by these services outweigh the cost. Pethokoukis [41] writes that users subject to surveillance capitalism "understand" and "accept the trade off", citing statistics that few users have changed their settings to protect their privacy and even fewer have stopped using a tech company's services. The implication of this argument is that the system is working for everyone.

There are numerous problems with this account. First, it is far from clear that users understand the nature of the surveillance they are subject to in any real sense. Differences in digital literacy of course play some role here [40], but more importantly, the operations of surveillance capitalism are designed to be unknowable [51, p. 21]. This "epistemic asymmetry", as Brunton and Nissenbaum call it, renders informed consent more or less impossible [9]. The public are generally aware that their data is being harvested but are given little insight into the data market through which this information is bought and sold, or its myriad uses by corporations, institutions, parties and states. This has caused some scholars to question the efficacy of "privacy" to resist the expansion of surveillance practices [46][1]. The privacy movement accepts the "essential legitimacy" of institutional surveillance [46, p. 67]; that private companies collect and use subjects' data goes unchallenged. In this framework, data collected by private companies remains private because it is given "in confidence"; it is only in the context of "breaches of privacy" that subjects of surveillance may protest. Events labelled "breaches" are usually particularly grievous misuses of data, which offer glimpses into this system and fleeting opportunities to have a conversation about values, Cambridge Analytica [44] and We-Vibe [21] being two particularly odious cases. The problem with the "breach" language is that it treats the system as essentially valid except for isolated cases of misuse.

Another issue with this account is whether subjects of surveillance can really opt out at all. Philosophers of technology have understood since the 1980s that regardless of personal preference, certain technologies are so deeply embedded into everyday life as to make their usage essentially mandatory (see e.g. Langdon Winner's discussion of television [29].) Today, the internet has become essential for social participation [51, p. 21]; to simply "opt-out", as the privacy paradox suggests, is not a practical option [9].

Resistance

The growth of surveillance has prompted a corresponding growth in projects that attempt to resist. Much of this has come from the art and design community.

Here we will briefly map out some forms of resistance to surveillance. This review will not discuss the recent proliferation of artworks *about* surveillance (for this see [34]), instead we are interested in attempts to transgress surveillance, particularly those which engage with enabling technologies.

Camouflage

One of the most fertile areas of resistance has been the exploration of camouflage, particularly by designers. Adam Harvey's *CV Dazzle* [18] is one of the best known examples of this. The project aimed to confuse what was then the most common face detection algorithm in use. He designed combinations of hair styles and makeup which would disguise patterns used by the algorithm to detect faces but appeared outwardly to be a fashionable if experimental style. These works are interesting because they target the algorithms that make photographic data machine readable rather than the surveillance devices themselves. In a related project, *URME* by Leo Selvaggio, the artist produces and sells uncomfortably realistic resin masks of his own face allowing strangers to disguise themselves with his identity [33]. In effect, this serves to disguise both the mask's wearer and the artist; his own trace muddied by many possible paths. There is a clear intersection here to the practices of obfuscation [9]. "Loyalty card swapping" achieves much the same end; introducing ambiguity into the agreggated data and rendering it less valuable and harder to use. Similarly, the browser extension *TrackMeNot* floods a given search engine with arbitrary queries, such that a user's true interests are harder to infer [23].

Avoidance

Avoidance can be understood as the changing of one's own behaviour to resist surveillance. Mostly this takes the form of technological avoidance e.g. use cash, not card, don't carry a phone, avoid social media. This is also the response advised by those who cite the privacy paradox. In some instances, technologies can assist users to avoid surveillance, as in the Institute for Applied Autonomy's *iSee* project which plots a "path of least surveillance" through Manhattan [13, p. 19]. Avoidance practices are, however, purely reactive and show little promise of returning power to subjects of surveillance. While many people do choose to avoid particular services or technologies in order to resist forms of surveillance, this is becominging increasingly difficult and in some cases impossible. We have already discussed the importance of many technologies in enabling social participation. Furthermore the efficacy of avoidance is overstated. Social networks are able to gather data beyond what is explicitly shared, extending even to those outside their platforms [47].

Sabotage

One of the most extreme forms of resistance is the destruction of the means of surveillance. In most cases of state surveillance, such as a network of CCTV cameras, subjects have no other way to affect change in the system. This practice was witnessed in action recently when protestors in Hong Kong cut down facial recognition towers and poured water on their electronics [38].

Surveillance and the camera

This paper is primarily concerned with the particular subset of surveillance technologies that allow for the surveillance of personal photographs by technology companies. Information about the nature and extent of this form of surveillance is limited. It must be pieced together from a collection of announcements, research publications, and privacy breaches. Given the secrecy surrounding corporate surveillance practices, some researchers argue that there is an aspect to this work which is "necessarily speculative" [10, p. 183].

Mobile computing devices revolutionised personal photography. Digital cameras drastically reduced the cost of photography for amateurs and allowed photographs to be shared via digital networks and accrete in the databases of social networks. The HCI community has already made strides in understanding privacy in these spaces. Hasan et al. examine several methods of redacting visual information to maintain privacy online [19, 20]. Image blurring, pixelation, silhouetting and masking etc. are shown to maintain privacy from other human viewers but have variable effects on viewer "satisfaction" with the images. These papers show that parameter choices and further artistic image filters can maintain viewer satisfaction. Li et al. [31] show that inpainting and avatar replacement are more effective privacy solutions than blurring and pixelation. While most of this work concerns the stated redactive methods, Tierney et al. [49] and Ra et al. [42] demonstrate how encryption can be used to maintain privacy online. Though Ra specifically addresses "algorithmic recognition" McPherson et al. [32] show that blurring, pixelation, and even the cryptographic methods employed by Ra et al. can be "defeated" with DL. It is clear that divergent privacy practices are required with respect to humans and machines. Our research is strictly concerned with privacy from algorithmic recognition. Developments in DL image classification afforded access to the semantic content of photographic data in a way that was not previously feasible. DL in this sense activated large databases of users' photographs for surveillance.

DL is used to interpret not only the images posted on social networks, but also those backed up on ostensibly private cloud storage services. In 2013 Google announced that it had pushed a major update to its Photos app allowing users to search their images by content without ever needing to label them. This was achieved through acquisition of the technology developed by Geoff Hinton's DL lab at the University of Toronto; the economic value of this technology is indicated by its rapid move into production "in just a little over six months" [43].

Furthermore, it appears that the camera itself is on track to become a surveillance device in a more literal sense. An article posted to Google's AI blog in October 2017 describes the use of machine learning within the camera app itself, using "semantic segmentation" to produce synthetic depth of field [30]. Another article posted in December of that year confirmed the inclusion of DL within the camera app as a direction of development. "Until recently, [smartphone] cameras behaved mostly as optical sensors... The next generation of cameras, however, will have the capability to blend hardware and computer vision algorithms that operate as well on an image's semantic content" [26]. Accordingly, privacy protections against algorithmic surveillance must be introduced at the source, that is, within the camera app.

CAMERA ADVERSARIA

We will now introduce a critical design intervention in algorithmic photographic surveillace. Our artefact affords a new mode of resistance to the subjects of surveillance by disrupting DL image classification.

The app

Camera Adversaria is designed as a non-intrusive replacement for a smartphone's default camera application. It takes seriously the idea that many of the technologies which enable surveillance capitalism are essentially unavoidable if one wishes to remain a part of modern society. As such it is designed to work within the existing systems of networked media, cloud storage, etc. and to mimic the interfaces and interaction conventions of existing camera applications.

The app has two main views; the capture view and the gallery view. The capture view (Fig. 1) offers nothing out of the ordinary; it should feel familiar to anyone who has used the default camera app. The gallery view (Fig. 2) appears much like a standard camera gallery but contains a simple slider and a text annotation. The slider adjusts the strength of the filter applied to the photograph while the text indicates what a standard DL image classifier can identify within the image. The DL model runs locally within the app and does not store or transmit its results. It displays the top classification alongside the "confidence" of the prediction. These two elements support a feedback loop where a user can balance photographic distortion with resultant machine readability. As such a user may decide how much distortion they are willing to accept in the image in order to inhibit surveillance.

Figure 2 left shows a photograph in the gallery view with the adversarial perturbation turned off resulting in a true classification displayed by the example DL model. On the right, the same image is displayed with an imperceptibly small perturbation added resulting in an incorrect prediction with high confidence.

The presence of the model's prediction serves to remind a user of the machine readability of images that will be exploited should the image ever find its way onto a social network or cloud storage. While this is an important element of the critical work that *Camera Adversaria* does, it may serve to make a user overly confident in the robustness of the adversarialised image. While there is evidence that adversarial perturbations work across DL models [35] it may be the case that those running in the servers of large tech companies are more resistant to adversarial perturbations or are trained to identify objects that

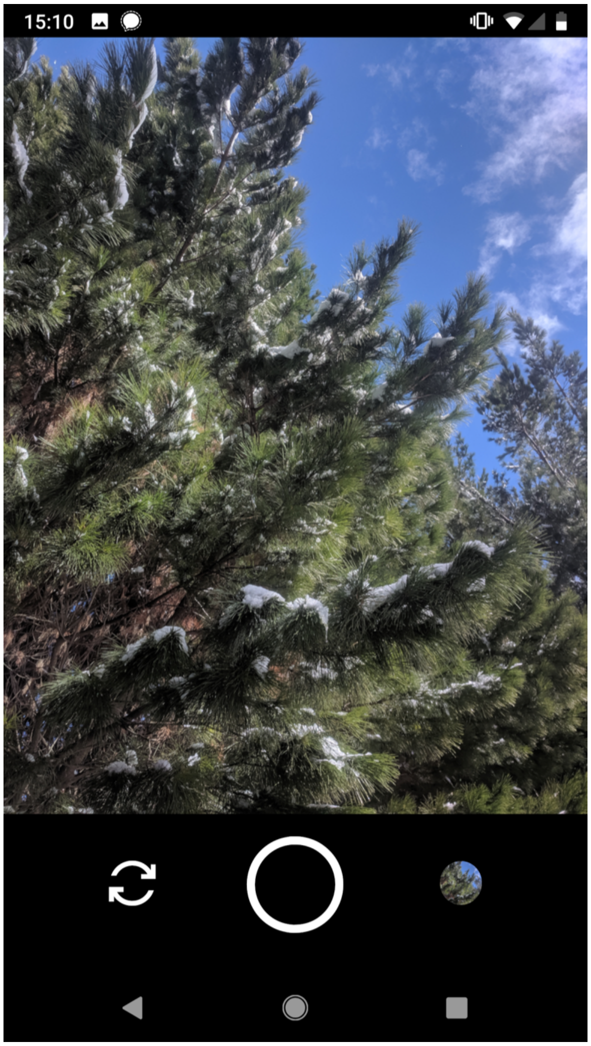

Figure 1. Camera Adversaria capture view.

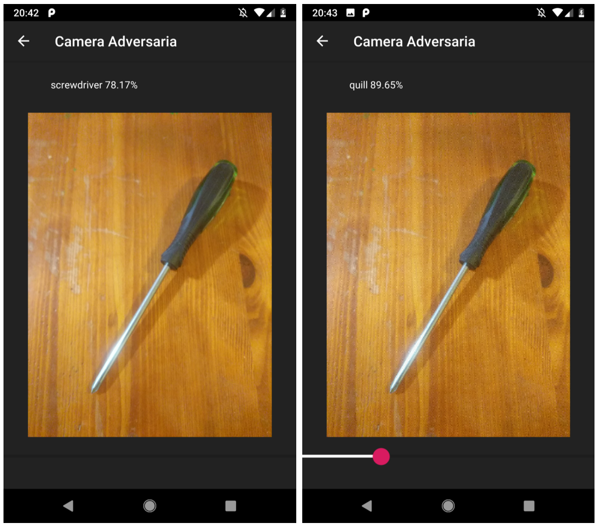

Figure 2. Left: photograph in gallery before adversarial perturbation. This is correctly predicted to depict a screwdriver. Right: same photograph after small adversarial filter turned on. Image is now predicted to depict a quill with high confidence.

do not exist in the model's labels. For this reason, the failure of a classification in the app does not guarantee that this aspect of the image is not machine readable in every case.

Design Process

The project was initially concieved as an online service for creating DL resistant images. We were surprised to find that despite the large research community working on adversarial perturbations, it still required expert knowledge to generate such an image. By comparison, other DL methods, e.g. for DeepDream, pix2pix, had spawned many easy-to-use online services. We realised that by putting adversarial perturbations in the hands of users could we start a conversation about algorithmic surveillance.

We quickly realised that transmitting user photographs over a network to be processed by our server would introduce so many additional privacy concerns as to defeat the initial goals. This made a web service a poor choice and we chose instead to develop software that could run on the user's own device. This introduced its own distinct challenges. The standard methods for producing an adversarial example is with an optimisation algorithm that requires access to the DL model and many iterations to find a suitable perturbation. This would be too slow and computationally demanding to run on a mobile device. As early as 2017, however, researchers realised that so called *universal* adversarial perturbations were possible [35]. These are singular perturbations that may be applied to an image causing a given model to misclassify in a significant percentage of cases. Again, these proved to be fairly robust across models and training sets [35]. This would allow us to apply an adversarial perturbation more efficiently, by simply storing the perturbation as an image and applying it over the photograph. The downside of this approach is that the filter can be easily removed given access to the perturbation, making the whole approach less robust. Surprisingly, Co et al. [11] found that simple procedural noise can be used to produce universal adversarial perturbations. This method is particularly well suited to surveillance resistance because it can be computed cheaply and uniquely for each image without ever needing to be optimised for a particular model. The filter used in *Camera Adversaria* is based on the description given by Co et al. but tweaked based on our own testing and iteration. Figure 3 shows an example adversarial perturbation generated by *Camera Adversaria*, it has been multiplied for visibility.

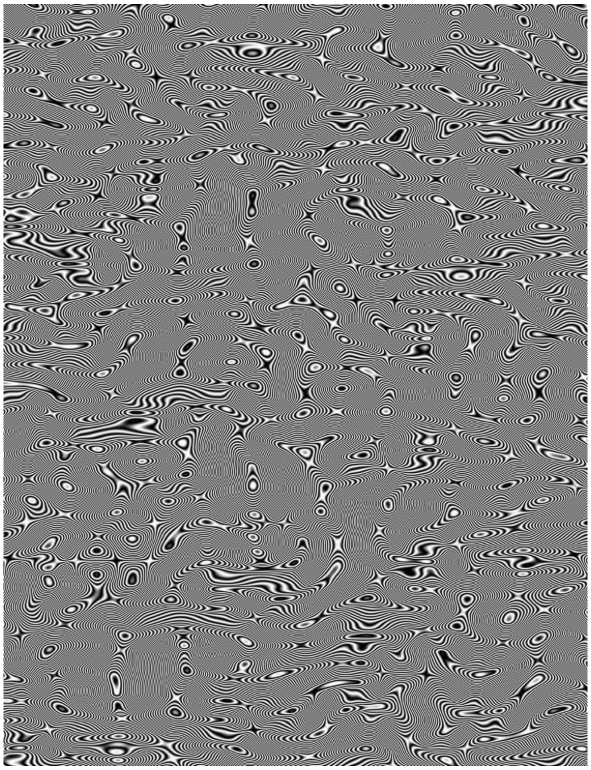

Figure 3. A sample adversarial perturbation generated with perlin noise. Filter has been multiplied for visibility.

During the prototyping phase, we built a simple convolutional neural network, MobileNet [22], into the camera application and collected a number of objects from the network's lables. This allowed us to rapidly iterate on the adversarial filter and evaluate our changes.

Design under surveillance

Camera Adversaria exists in an under-explored region in the design space of resistance. Methods of resistance can be divided into those that change the subject (*automutative*) and those that change the systems of surveillance (*exomutative*). We use these two notions to examine how the design possibilities vary between state and corporate surveillance.

Table 1 places existing modes of resistance into a matrix of state/corporate and automutative/exomutative in order to highlight an under-explored region in the design space. Namely, in corporate surveillance, parts of the surveillance apparatus (e.g. smartphones) are owned and operated by the subjects of surveillance. In this system it is the subjects of surveillance

	State	Corporate
Automutative	Camouflage	Avoidance
Exomutative	Sabotage	?

Table 1. Ways of resisting surveillance.

who carry out "the work of being watched" [1]. Parts of this surveillance apparatus are in this sense *decentralised*. This appears to fly in the face of traditional categories from the philosophy of technology which characterise decentralised systems as inherently democratic and good (see e.g. Lewis Mumford [37]). Surveillance capitalism is clearly far from a democratic system. However subjects still exert some authority over the software that runs on their devices. As such, unlike in the state-owned and operated CCTV systems, users have some recourse to change the system without resorting to sabotage. This presents an opportunity for well designed applications to introduce tools of resistance at the source.

Some existing work has superficial similarities to *Camera Adversaria*. *Bye Bye Camera* by Damjanski uses software to entirely erase humans from photographs [3] and *DeepPrivacy* by Hukkelås et al. substitutes faces in photographs with realistic but entirely synthesised alternatives [25]. Both are redactive modes and erode the social function of image sharing, i.e. they anonymise photographs for humans as well as machines. Crucially, unlike existing work in this space, *Camera Adversaria* is designed to be unobtrusive, leaving images virtually indistinguishable to humans but vastly different to machines. Elsewhere we have argued that the characterisation of neural networks as "seeing" is misleading [8]. Neural networks are unaffected by an image's global coherence but instead depend on low level statistical features. *Camera Adversaria* exploits significant differences between human and machine vision, introducing changes that radically affect the DL prediction but are often imperceptible to a person.

DISCUSSION

In this section we discuss the consequences of our research for adversarial perturbations research and HCI.

Research Community and Complicity

When conducting a literature review of adversarial perturbation research, we were surprised to find that few authors have realised the potential of these methods for privacy and resistance of surveillance. Instead DL researchers working with adversarial perturbations tend to use language that maligns resistance. Nowhere that we have discovered, has anyone in the adversarial perturbation research community attempted to make adversarial methods available to the subjects of surveillance. This includes researchers who developed a mobile app purely to demonstrate their methods [36].

Adversarial examples are typically described in the language of threats, exploits and bad actors. The DL system is presumed to be good, those wishing to disrupt it are "malicious". This is most clearly framed in the notion of an adversarial "attack". The language of attack is very commonly used in the literature [2, 24, 11, 5].

The literature is full of value-laden language:

"Machine learning classifiers are known to be vulnerable to inputs maliciously constructed by adversaries to force misclassification" [24]

"This linear behavior suggests that cheap, analytical perturbations of a linear model should also damage neural networks" [17]

"Know your adversary: modeling threats 'If you know the enemy and know yourself, you need not fear the result of a hundred battles.' (Sun Tzu, The Art of War, 500 BC)" [5]

Even Co et al. whose methods we have most closely adopted frame their work in this way: "it is important to ensure that such algorithms are robust to malicious adversaries" [11].

We are deeply concerned by the DL research community's framing of adversarial perturbations. This language serves to affirm the benevolence of the DL system and dissuade resistance. Particularly, this raises questions about the role of the research community in countenancing systems of surveillance.

For argument's sake, we wish to clarify that adversarial perturbations are *entirely non-destructive*. They confound a machine learning system only within the scope of a particular image. Understood as a key part of contemporary surveillance systems, there are numerous valid cases in which a subject of surveillance might wish to avoid classification, particularly as the final usage of the collected data is unknown. With adversarial perturbations this can be done without sabotage or damage, the subject simply avoids classification. The field should reconsider its use of language with regards to the method. Adversarial perturbations used in this way are an entirely *defensive* mode of resistance.

Need for critical design

According to Dunne and Raby, design is ideological and most design affirms the status quo, reinforcing cultural, social, techical and economic expectations [14, p. 58]. This was written with industrial product design in mind, but the same claim applies to the development of adversarial perturbations in research and industry.

As a work of critical design *Camera Adversaria* provides an alternative vision to that put forward by research and industry.

Camera Adversaria reclaims adversarial perturbations as a mode of resistance for the subjects of surveillance. This is a necessary corrective to the framing of that technolgogy within the research community.

A reasonable critique of *Camera Adversaria* is that it is just a stopgap. We can expect algorithms to improve and to eventually be less susceptible to this kind of resistance. For this reason it is even more important that we begin to question the role of this technology in systems of surveillance.

Another job of critical design is to create artefacts that embody *alternative social and economic values* [14, p. 58]. *Camera Adversaria* reclaims adversarial perturbations for their defensive and emancipatory properties and makes these accessible to the subjects of surveillance. The app presents an immediately useful tool of resistance to this particular incarnation of surveillance capitalism.

Resolving the privacy paradox

The privacy paradox appears to show that users will sell their privacy for cheap, in exchange for convenience or social capital. This account ignores asymmetries of power and knowledge that render informed consent more or less impossible. To figure out if people really care about privacy we need more transparency around the use of data by corporations and we need to offer realistic tools of resistance that allow for continued participation in the world. The privacy paradox dissolves if one acknowledges that people don't have a realistic alternative.

With *Camera Adversaria* we hope to offer this alternative in the restricted domain of personal photographs. However more interventions in other aspects of contemporary surveillance are required.

This project does not dismantle the apparatus that allows corporate surveillance in the first place. Instead it transgresses a part of that normally invisible system, and reveals the culture that sustains surveillance capitalism. Whether this is enough remains to be seen.

Social challenges

Although what we present here is a technical artefact, we are clear that this cannot stand in for a necessary social change. The political asymmetries of surveillance capitalism cannot be solved with a technical solution, this will only start an "arms race". While individual action may in fact not be a workable solution for the long term, regulators are also poorly placed to contend with these challenges. It is difficult to have a public conversation about surveillance while we know so little about these systems. The final goal of *Camera Adversaria* is to reveal, to some degree, the surveillance system and present an alternative. This is a challenge to the current discourse, particularly the moralising stance apparent in the technical literature.

Opening up this region of design space

We have demonstrated the existence of an under-explored region in the design space of surveillance resistance. This space warrants further interventions from design and HCI. Similar interventions will be possible anywhere users continue to exert some authority over the means of surveillance. This is the case for much of corporate surveillance, where users still do the "the work of being watched" [1]. Though personal devices are part of larger networked systems, users still often control the source of this data. We can write software that augments a user's data to make surveillance capitalism less valuable.

Similar manipulations to those used in *Camera Adversaria* could be used anywhere human and machine readability varies. This could be done for DL in other domains where adversarial examples have been shown to exist; e.g. text and sound. Beyond DL, there are often significant differences between machine and human readability. It might be possible to exploit the visual similarity of obscure characters (e.g. unicode's so called "confusables") with standard characters so as to maintain human readability but confuse machine readers. This is again, because humans innately see similarities in characters and can handle missing information in context, whereas for a machine, these appear as an arbitrary set of indices. It is telling that the technical community has already identified this weakness in unicode and labelled it an "attack" [45]. Perhaps other "attacks" may be usefully repurposed as a defence against surveillance.

Future work

Future work on *Camera Adversaria* should include a user study to identify the project's effectiveness in bringing attention to surveillance concerns and protecting users' privacy. Important questions include how best to balance the "usability vs privacy" tension from a UX design perspective (especially within a diverse user community) but also the opportunity to see how different users feel and act when these surveillance issues (and the opportunity to circumvent them) are foregrounded. The app could be extended to experiment with further adversarial filters. The effectiveness of this form of procedural noise in disrupting DL systems suggests that other more effective methods may exist.

CONCLUSION

Camera Adversaria is a critical design intervention in surveillance capitalism. We highlight the significance of DL image classification in contemporary corporate systems of surveillance and critique the research community's complicity countenancing these systems. We present an application designed to fit into and disrupt the broader surveillance apparatus and make it freely available to those subject to surveillance. This work identifies a design space with potential for more contributions from HCI researchers. Finally, we reclaim methods developed by the DL research community as a *defensive* tool for the those subject to surveillance.

ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers for their insightful and constructive feedback.

REFERENCES

[1] Mark Andrejevic. 2002. The work of being watched: Interactive media and the exploitation of self-disclosure.

Critical studies in media communication 19, 2 (2002), 230–248.

[2] Anish Athalye, Nicholas Carlini, and David Wagner. 2018. Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples. *arXiv:1802.00420 [cs]* (Feb. 2018).

[3] Jason Bailey. 2019. Bye Bye Camera - an App for the Post-human Era. `https://www.artnome.com/news/2019/6/24/bye-bye-camera-an-app-for-the-post-human-era`, *Artnome* (2019). Accessed: 16 September 2019.

[4] Jeffrey Bardzell and Shaowen Bardzell. 2013. What is critical about critical design?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 3297–3306.

[5] Battista Biggio and Fabio Roli. 2018. Wild Patterns: Ten Years after the Rise of Adversarial Machine Learning. 84 (2018), 317–331. DOI: `http://dx.doi.org/10.1016/j.patcog.2018.07.023`

[6] James Bridle. 2017. Every CCTV Camera (CC). `https://jamesbridle.com/works/every-cctv-camera-cc`. (2017).

[7] Kieran Browne. 2019. Camera Adversaria Source Code. `https://github.com/kieranbrowne/camera-adversaria`, *GitHub repository* (2019).

[8] Kieran Browne, Ben Swift, and Henry Gardner. 2018. Critical Challenges for the Visual Representation of Deep Neural Networks. In *Human and Machine Learning*. Springer, Switzerland, 119–136.

[9] Finn Brunton and Helen Nissenbaum. 2011. Vernacular resistance to data collection and analysis: A political theory of obfuscation. *First Monday* 16, 5 (2011).

[10] Nicholas Carah and Daniel Angus. 2018. Algorithmic brand culture: participatory labour, machine learning and branding on social media. *Media, Culture & Society* 40, 2 (2018), 178–194.

[11] Kenneth T Co, Luis Muñoz-González, and Emil C Lupu. 2018. Procedural Noise Adversarial Examples for Black-Box Attacks on Deep Neural Networks. *arXiv preprint arXiv:1810.00470* (2018).

[12] Julie E Cohen. 2019. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford University Press, USA.

[13] Carl DiSalvo. 2012. *Adversarial Design*. MIT Press.

[14] Anthony Dunne and Fiona Raby. 2001. *Design noir: The secret life of electronic objects*. Birkhäuser.

[15] Luis A Fernandez and Laura Huey. 2009. Is resistance futile? Thoughts on resisting surveillance. *Surveillance & Society* 6, 3 (2009), 199–202.

[16] Michel Foucault. 2012. *Discipline and Punish: The Birth of the Prison*. Vintage.

[17] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and Harnessing Adversarial Examples. (2014). `http://arxiv.org/abs/1412.6572`

[18] Adam Harvey. 2010. CV Dazzle: Camouflage from face detection. `http://cvdazzle.com/`. (2010). Accessed: 13 September 2019.

[19] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer experience of obscuring scene elements in photos to enhance privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM.

[20] Rakibul Hasan, Yifang Li, Eman Hassan, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2019. Can Privacy Be Satisfying?: On Improving Viewer Satisfaction for Privacy-Enhanced Photos Using Aesthetic Transforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM.

[21] Alex Hern. 2017. Vibrator maker ordered to pay out C$4m for tracking users' sexual activity. `https://www.theguardian.com/technology/2017/mar/14/we-vibe-vibrator-tracking-users-sexual-habits`, *The Guardian* (14 March 2017). Accessed: 7 January 2020.

[22] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017).

[23] Daniel C Howe and Helen Nissenbaum. 2009. TrackMeNot: Resisting surveillance in web search. In *Lessons from the Identity trail: Anonymity, privacy, and identity in a networked society*, Ian Kerr, Carole Lucock, and Valerie Steeves (Eds.). Oxford University Press, 417–436.

[24] Sandy Huang, Nicolas Papernot, Ian Goodfellow, Yan Duan, and Pieter Abbeel. 2017. Adversarial Attacks on Neural Network Policies. (2017). `http://arxiv.org/abs/1702.02284`

[25] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. 2019. DeepPrivacy: A Generative Adversarial Network for Face Anonymization. *arXiv preprint arXiv:1909.04538* (2019).

[26] Alex Kauffmann. 2017. Introducing Appsperiments: Exploring the Potentials of Mobile Photography. `https://ai.googleblog.com/2017/12/introducing-appsperiments-exploring.html`, *Google AI Blog* (2017). Accessed: 16 September 2019.

[27] Spyros Kokolakis. 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security* 64 (2017), 122–134.

[28] Sandjar Kozubaev. Stop Nigmas: Experimental Speculative Design Through Pragmatic Aesthetics and Public Art. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (2016) *(NordiCHI '16)*. ACM, 76:1–76:10. DOI: http://dx.doi.org/10.1145/2971485.2993921

[29] Winner Langdon. 1986. *The Whale and the Reactor: A search for limits in an age of high technology*. University of Chicago Press, Chicago.

[30] Marc Levoy. 2017. Portrait mode on the Pixel 2 and Pixel 2 XL smartphones. https://ai.googleblog.com/2017/10/portrait-mode-on-pixel-2-and-pixel-2-xl.html, *Google AI Blog* (2017). Accessed: 16 September 2019.

[31] Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 67.

[32] Richard McPherson, Reza Shokri, and Vitaly Shmatikov. 2016. Defeating image obfuscation with deep learning. *arXiv preprint arXiv:1609.00408* (2016).

[33] Torin Monahan. 2015. The right to hide? Anti-surveillance camouflage and the aestheticization of resistance. *Communication and Critical/Cultural Studies* 12, 2 (2015), 159–178.

[34] Torin Monahan. 2018. Ways of being seen: surveillance art and the interpellation of viewing subjects. *Cultural Studies* 32, 4 (2018), 560–581.

[35] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. 2017a. Universal adversarial perturbations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1765–1773.

[36] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. 2017b. Universal adversarial perturbations. (2017). https://youtu.be/jhOu5yhe0rc

[37] Lewis Mumford. 1964. Authoritarian and democratic technics. *Technology and culture* 5, 1 (1964), 1–8.

[38] Guardian News. 2019. Anti-surveillance protesters tear down 'smart' lamp-post in Hong Kong. (2019). https://youtu.be/u1Ji7wonUhE

[39] George Orwell. 2009. *Nineteen Eighty-Four*. Everyman's Library.

[40] Yong Jin Park. 2013. Digital Literacy and Privacy Behavior Online. *Communication Research* 40, 2 (April 2013), 215–236. DOI: http://dx.doi.org/10.1177/0093650211418338

[41] James Pethokoukis. 2019. In praise of surveillance capitalism. http://www.aei.org/publication/in-praise-of-surveillance-capitalism/, *AEIdeas Blog* (2019). Accessed: 16 September 2019.

[42] Moo-Ryong Ra, Ramesh Govindan, and Antonio Ortega. 2013. P3: Toward privacy-preserving photo sharing. In *Presented as part of the 10th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 13)*. 515–528.

[43] Chuck Rosenberg. 2013. Improving Photo Search: A Step Across the Semantic Gap. https://ai.googleblog.com/2013/06/improving-photo-search-step-across.html, *Google AI Blog* (2013). Accessed: 16 September 2019.

[44] Matthew Rosenberg, Nicholas Confessore, and Carole Cadwalladr. 2018. How Trump Consultants Exploited the Facebook Data of Millions. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html, *NY Times* (17 March 2018). Accessed: 7 January 2020.

[45] Narges Roshanbin and James Miller. 2011. Finding homoglyphs-a step towards detecting unicode-based visual spoofing attacks. In *International Conference on Web Information Systems Engineering*. Springer, 1–14.

[46] James B Rule. 2012. 'Needs' for Surveillance and the Movement to Protect Privacy. In *Routledge handbook of surveillance studies*, Kirstie Ball, Kevin D Haggerty, and David Lyon (Eds.). Routledge, 64–71.

[47] Emre Sarigol, David Garcia, and Frank Schweitzer. 2014. Online privacy as a collective phenomenon. In *Proceedings of the second ACM conference on Online social networks*. ACM, 95–106.

[48] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. 2014. Intriguing properties of neural networks. In *International Conference on Learning Representations*.

[49] Matt Tierney, Ian Spiro, Christoph Bregler, and Lakshminarayanan Subramanian. 2013. Cryptagram: Photo privacy for online social media. In *Proceedings of the first ACM conference on Online social networks*. ACM, 75–88.

[50] David Murakami Wood and Torin Monahan. 2019. Platform Surveillance. *Surveillance & Society* 17, 1/2 (2019), 1–6.

[51] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power*. Profile Books.