
Agree or Cancel? Research and Terms of Service Compliance

Kristen Vaccaro

University of Illinois,
Urbana-Champaign
201 North Goodwin Avenue
Urbana, IL 61801
kvaccaro@illinois.edu

Karrie Karahalios

Center for People and
Infrastructures
University of Illinois,
Urbana-Champaign
kkarahal@illinois.edu

Christian Sandvig

University of Michigan
Room 5385 North Quad
105 South State Street
Ann Arbor, MI 48109
csandvig@umich.edu

Kevin Hamilton

Center for People and
Infrastructures
University of Illinois,
Urbana-Champaign
kham@illinois.edu

Cedric Langbort

Center for People and
Infrastructures
University of Illinois,
Urbana-Champaign
langbort@illinois.edu

Abstract

We address the ACM Code of Ethics and discuss the stipulation that researchers follow terms of service. While the reasons for following terms of service are clear, we argue that there are hidden costs. Using the example of research into algorithm awareness and algorithm transparency, we argue that for some research problems the benefits to society outweigh the harm of violating terms of service. We draw attention to current strategies that researchers use to adapt to the ACM prohibition on research violating terms of service and some results of those approaches. We conclude with recommendations for determining researchers' current understanding of the ACM Code of Ethics and terms of service restrictions and updating the existing guidelines.

Author Keywords

Ethics, algorithm awareness, research guidelines

ACM Classification Keywords

K.4.1 [Computers and Society]: Ethics.

Introduction

Despite the field's historical association with "hacking," the majority of the computing sciences have a strong history of protecting intellectual property and industry interests [17]. And the Association for Computing

Machinery respects licenses and policies both in its social norms and officially. The “General Moral Imperatives” of the ACM Code of Ethics state that “*Violation of...trade secrets and the terms of license agreements is prohibited by law in most circumstances. Even when software is not so protected, such violations are contrary to professional behavior*” [1]. The clear official expectation is that ACM researchers will follow terms of service, terms of use, acceptable use policies and end user license agreements (henceforth combined as “terms of service”).

There are many reasons to follow terms of service. Researchers generally respect the law, including trade secrets and intellectual property. They may also hold the utilitarian view that research practices following corporate terms will be good for industry/academia relations and are likely to result in better long-term research outcomes.

However, there are well-documented problems with terms of service. Terms of service are unilaterally set by providers, can be arbitrary, inconsistent, and can change at any time (and frequently do) [28]. Users have difficulty finding, reading and understanding terms of service agreements [8, 14]. Finally, terms of service may be used for strategic overreach [22, 24], producing periodic popular outcry [27, 30]. Strategic overreach issues are particularly endemic, because terms of service are generally written by corporate counsels to protect companies from liability. The counsel’s strategy is to assign the company the maximum possible benefit and the user the least. Indeed, an incentive exists to serve employers by assigning companies illegal extravagant benefits on the chance they might be enforceable later.

For example, Microsoft and Facebook have employed terms of service that grant them the permanent right to modify and re-sell any user communications, including

patentable material. Acceptable use policies written by Verizon and AT&T have prohibited customers from criticizing them. While the most controversial terms of service are often withdrawn after public protest, many others are found to be unenforceable when challenged – for example, terms available only via a small link at the bottom of a homepage have been found unenforceable.

There are emerging guidelines for what features of terms of service are necessary for enforcement, beyond standard contract grounds [15]. These include providing conspicuous notice of all terms, requiring active assent, prohibiting use prior to such assent and periodic confirmation of agreement. But because of the variety of online terms of service, such guides are still in progress.

These flaws might be reason enough to reconsider the ACM mandate that researchers follow all terms of service. However, we believe that there is another important reason. We argue that some important research problems cannot be investigated while following terms of service. The benefits of this research for society may outweigh the harms to the companies involved. The ACM Code of Ethics already suggests a model for such cases.

In particular, the Code takes a more nuanced approach to accessing computing resources. Although it states “*unauthorized use of a computer or communication system,*” including any “*accounts and/or files...without explicit authorization*” is prohibited, it tempers this by noting that owners only have the “*right to restrict access to their systems so long as they do not violate the discrimination principle.*” Although the discrimination principle clearly states “*these ideals do not justify unauthorized use of computer resources nor do they provide an adequate basis for violation of any other ethical imperatives,*” nevertheless this tempers the mandate.

We make an analogous argument for weakening the terms of service mandate. We draw from a survey of research on algorithm transparency to show that cases that violate terms of service while offering potential benefits to society are a frequent focus of research attention.

Algorithm Transparency and the Problem of Discrimination

A persistent question in computer supported work is how to design systems to encourage the appropriate level of trust [19]. This can be difficult because many users believe that computer systems are trustworthy simply because they exist, leading some to argue that code has social and political power that should be examined [16]. Proprietary algorithms drive many systems making important financial, informational and personal decisions about and on behalf of users. Researchers have begun to suggest that the power and ubiquity of these systems make it essential to probe how they work [9, 23].

FTC Chairwoman Edith Ramirez noted the potential for “discrimination by algorithm,” or “digital redlining,” in predictive scoring of online transactions [21]. Journalists have noted the discriminatory potential of dynamic pricing, noting that even using geography as a pricing feature can “diminish the Internet’s role as an equalizer” by reinforcing biases against rural and poorer areas which tend to have less competition [29]. Consumers have tried to crowdsource information about dynamically priced products [3], while computer science researchers have developed methodologies to test these results [13].

Researchers have also looked at discrimination in ad placement. A detailed comparison between two studies investigating ads displayed by Google suggests the importance of additional research in this area.

In one study, published in the Huffington Post, Nathan Newman investigates potential racial discrimination by manually exploring ads produced by emails with subject lines containing names that are strongly associated with particular races [20]. Little methodological information is provided, including the age of accounts used, the selection process for inputs, or the full frequency of outputs. More fundamentally, personalization models are often developed over time, so an approach based on single emails may not be appropriate. Nevertheless, he claims that this investigation “yielded disturbing results,” providing as output some example ads he argues show “racial profiling.” Google later released a statement calling the article “wildly inaccurate.”

In contrast, Latanya Sweeney investigates the differing frequency of “arrest record” ads by race with care, describing reasonable expectations for how an algorithm might process input data, discussing the automation of the search process, and providing frequency data for her results [26]. While similarly contested by the companies involved, this rigor makes it difficult to reconcile her results with company claims that ads are not associated with these names, and makes a much stronger argument that racial discrimination may be occurring.

The difference in these approaches suggests the value of greater research in this area. However, in both cases, replicating or scaling these studies would likely require terms of service violations as “*Google’s Terms of Service do not allow the sending of automated queries of any sort*” without express permission [10]. Other research has investigated discrimination in computing systems [6, 9], but little of it has been conducted by computer, data or information science researchers or on a large scale.

One of the reasons for the lack of research into the ethics

of algorithms is that most proprietary algorithms function as black boxes, with little information about input features or processing available [5, 11]. Users see only their individual outputs and have little sense of how they compare to others. Because many problematic algorithms can be identified only at scale, discrimination may be completely opaque to individual users.

Several research approaches are under development [5, 23]. Because of the difficulty of obtaining and analyzing code, most approaches focus on reverse engineering some understanding of the algorithm from paired input and output sets [18, 31]. However, this kind of reverse engineering, which often involves scraping data or sending many queries to online systems, frequently violates terms of service. Indeed, not only do many researchers investigating harmful discrimination violate terms of service, in many cases they have to violate terms of service *first*, before they can identify any discrimination.

Implications for Computing Research

Work exploring discrimination in computing systems clearly encounters issues related to the ACM restriction on terms of service, as do other domains, particularly security. There appear to be two primary responses by researchers to these restrictions.

First, in some cases, a tacit understanding develops among fellow researchers that terms of service cannot be followed when investigating some questions. Some researchers personally contact companies to see whether they will forgive terms of service violations for academic research. Other researchers simply ignore the ACM restrictions. There are many published papers where researchers violated terms of service to get their work done.

However, this leaves researchers in a tenuous position,

uncertain of the legality of their work. So the second common result is avoiding certain research areas. This may have important consequences for public understanding of computing and users' digital lives. Research in algorithm awareness has found many users are unaware of the existence of algorithms, for example, that Facebook's News Feed is filtered [7]. As a result, users sometimes infer social meaning from algorithm effects, e.g., thinking a friend intentionally blocked posts to them when instead an algorithm filtered out those stories. Such important topics should be studied by diverse researchers, but it is made difficult under terms of service restrictions.

Journalists work in this area, because while their own code of ethics includes mandates to minimize harm, it simultaneously urges them to be "*courageous about holding those with power accountable*" and "*recognize a special obligation to serve as watchdogs*" [25]. While these obligations prioritize important work by journalists, they sometimes lead to an adversarial approach [12]. Perhaps in response to this kind of work, we have found companies are surprisingly aggressive in attempts to censor any such findings. Even for well-conducted research, companies will use the fact that they *no longer* engage in a practice as grounds to not publish that it ever occurred.

While protecting intellectual property and maintaining a good relationship with industry is important, we believe that building systems that enhance the quality of users' lives and providing a rigorous, technical forum to discuss problematic industrial or research practices is essential. For that reason, we make two recommendations.

First, we recommend changing the ACM guidelines on terms of service. Researchers are already comfortable with methods for getting neutral feedback on the importance of their research goals relative to potential costs, for

example, evaluating benefits and harms in the standard Institutional Review Board process. A similar framework to evaluate whether the importance of research goals may outweigh a standard presumption of protecting corporate interests in terms of service is one possibility.

However, even changes to the ACM code will not resolve the ambiguity of the situation. Researchers may be hesitant to run afoul of the US Computer Fraud and Abuse Act (CFAA), a law that has been applied to include terms of service in some cases. Researchers have argued that this legislation is over-broad [23] and that uncertainty about its enforcement inhibits research [4]. To avoid these issues, the ACM should leverage its power to change industry practices, so that some research interests might be explicitly excluded from standard terms of service.

Second, we recommend that the community develop an alternative code of ethics that represent our ethical standards for current computing research. We note that the ACM faces added constraints, including funding issues and the need to develop codes that cover many distinct types of computing research. For these reasons, it may be difficult to change terms of service practices. However, if the community of researchers has already tacitly accepted that violating terms of service may be ethical, if illegal, perhaps these community norms should be made explicit.

A number of models exist that ACM researchers could use to develop guidelines for this work. These include the journalistic Code of Ethics, with its prioritization of the truth, or the Association of Internet Researchers ethical guidelines [2], which develop a series of questions for researchers to consider when designing their work.

Before developing a new framework, however, we argue there is a necessary preparatory step. We must first assess

whether and to what extent researchers are actually aware of the ACM guidelines on terms of service and typical terms of service restrictions. We have found many researchers appear unaware of the ACM Code of Ethics and the topics it covers. Others are unaware that the ACM does not publish research that violates terms of service – some learn this only when their work is rejected on those grounds. Even among those aware of the ACM stance, how aware are they of the terms of service for the platforms they work with and whether their work conforms? And how frequently do researchers knowingly ignore terms of service?

Surveying the community of computing researchers on this topic will be challenging. In studying algorithm awareness among the public, we have found that developing surveys that probe existing awareness levels is extremely challenging. Nevertheless, at present the ACM Code of Ethics serves as a form of terms of service for researchers publishing in ACM venues — determining the current state of understanding becomes an important first step.

Conclusion

One of the missions of the ACM is to ensure the public is well-educated about computing technology. Even concepts fundamental to computing researchers, such as the existence of algorithms, may be completely unknown to the public. But research in this area (and others) is challenged by the ACM prohibition on work violating terms of service. Some researchers may simply ignore the ACM guidelines, but others may avoid work in important areas. So while following terms of service has clear benefits, it may also have hidden costs.

We argue that computing researchers should not simply abandon this important work. We believe that the ACM

should reconsider the prohibition against violating terms of service and encourage research exemptions. Alternatively, the researcher community can develop its own ethical guidelines, but an essential first step is to better understand researchers' current awareness of the ACM Code of Ethics, its terms of service mandate, and how their work accords with terms of service.

These changes may be hard. But when community standards have been established for conducting work that, while violating terms of service offers real benefits to society, computing research will benefit.

References

- [1] Acm code of ethics. <http://www.acm.org/about/code-of-ethics>.
- [2] Ethical Decision-Making and Internet Research. <http://aoir.org/reports/ethics2.pdf>.
- [3] Bidding For Travel. <http://biddingfortravel.com>.
- [4] Burstein, A. J. Conducting cybersecurity research legally and ethically. *LEET* (2008).
- [5] Diakopoulos, N. Algorithmic accountability reporting. *Tow/Knight Brief* (2013).
- [6] Diakopoulos, N. Sex, violence, and autocomplete algorithms. *Slate* (2013).
- [7] Eslami, M., Rickman, A., Vaccaro, K., Sandvig, C., Hamilton, K., and Karahalios, K. "i always assumed that i wasnt really that close to [her]". *In submission*.
- [8] Fiesler, C., and Bruckman, A. Copyright terms in online creative communities. *CHI* (2014).
- [9] Gillespie, T. Can an algorithm be wrong? *Limn* (2012).
- [10] Google Terms of Service. <http://www.google.com/policies/terms/>.
- [11] Hamilton, K., Karahalios, K., Sandvig, C., and Eslami, M. A path to understanding the effects of algorithm awareness. *CHI* (2014).
- [12] Hanitzsch, T. Deconstructing journalism culture. *Communication Theory* (2007).
- [13] Hannak, A., Soeller, G., Lazer, D., Mislove, A., and Wilson, C. Measuring price discrimination and steering on e-commerce web sites. *ICM* (2014).
- [14] Jensen, C., and Potts, C. Privacy policies as decision-making tools. *CHI* (2004).
- [15] Kahn, S., and Kiferbaum, D. Browsewrap arbitration? *NY Dispute Resolution Lawyer* (2012).
- [16] Lessig, L. *Code and Other Laws of Cyberspace*. Basic Books, 1999.
- [17] Levy, S. *Hackers*. Dell, 1985.
- [18] ProPublica Message Machine. <http://projects.propublica.org/emails/>.
- [19] Muir, B. Trust between humans and machines. *IJMMS* (1987).
- [20] Newman, N. Racial and economic profiling in google ads. *Huffington Post* (2011).
- [21] Ramirez, E. Ftc chairwoman opening remarks. In *Big Data: A Tool for Inclusion or Exclusion* (2014).
- [22] Russo, J. How To Read "Terms of Use" Agreements. <http://web.stanford.edu/class/ee380/Abstracts/010411.html>.
- [23] Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. Auditing algorithms. *ICA* (2014).
- [24] MS Passport. <http://yro.slashdot.org/story/01/04/03/1535244/>.
- [25] Society of professional journalists code of ethics. <http://www.spj.org/ethicscode.asp>.
- [26] Sweeney, L. Discrimination in online ad delivery. *CACM* (2013).
- [27] By Agreeing To Terms Of Service, You've 'Consented' To Search. <http://www.techdirt.com/articles/20141104/18041429043>.
- [28] TOSBack: Terms of service tracker. <http://tosback.org/>.
- [29] Valentino-Devries, J., Singer-Vine, J., and Soltani, A. Websites vary prices, deals based on users' information. *WSJ* (2012).
- [30] Wortham, J., and Bilton, N. What instagram's new terms of service mean for you. *NY Times* (2012).
- [31] XRay. <http://xray.cs.columbia.edu/>.