

Policy Statement on the Regulation of Commercial Search Engines

DATE: December 7, 2018

The Honorable Joseph J. Simons
Chairman
Federal Trade Commission
600 Pennsylvania Ave., N.W.
Washington, D.C. 20580

Dear Chairman Simons:

This memo presents the case for the **regulation of selection and sorting algorithms used in commercial search engines (CSEs)**. The case for the proposed regulation is built on two considerations: *first*, the harm to consumers caused by the operation of algorithms trained on data containing bias and errors; and *second*, the need for an adequate regulatory framework to facilitate the development and deployment of future AI-related applications. These two considerations are developed in more detail in Sections II and III, respectively. In Section IV, I present three policy options for the Chairman's consideration. I conclude in Section V by offering a final recommendation based on a thorough assessment of each of their respective merits.

I. BACKGROUND

Over the past two decades, the internet has become an essential part of our daily lives. There are currently close to 275 million internet users in the US (~84% of the total population), with this number expected to grow to 283.5 million (~87%) by 2022 (Statista, 2018). Meanwhile, the selection and sorting algorithms used by CSEs like Google, Bing, Yahoo, and others continue to exert an every-growing degree of influence over the global spread of information and other communication processes (Ahmed, Schulz, & Nakamoto, 2014).

Of the dozens of CSEs that would potentially be affected by this regulation, Google is the most influential by far, controlling between 87% and 92% of the total search engine market (Statcounter, 2017). Over 3.5 billion searches are made on Google every day on topics ranging from “How to make slime” to “What is DACA” (Google, 2017). A number of recent studies, however, have shown that the algorithms used by CSEs are highly susceptible to bias and error. While algorithmic errors in commercial recreational systems may have only limited impact (e.g., sending someone on a bad date), errant algorithms in foundational infrastructure, such as the internet, pose much greater risks to society (Osoba & Welser IV, 2017). For example, Sunstein (2001) and Pariser (2011) have argued that,

when algorithmic information services can be personalized to [the degree that two users can get completely different results to the same query], the diversity of public knowledge and political dialogue may be undermined. We are led—by algorithms and our own preference for the like-minded—into “filter bubbles” (ibid.), where we find only the news we expect and the political perspectives we already hold dear (Gillespie, 2014).

While the technology underlying CSEs has been with us for some time, the science of formally assessing the nature and magnitude of harm done by algorithmic bias and errors is relatively new. For example, publications in the US related to deep learning, specifically, have increased from fewer than 50 in 2012 to over 250 in 2015 (National Science and Technology Council Networking and Information Technology Research and Development Subcommittee, 2016). The FTC has a responsibility to keep pace with what we are learning about the risks that algorithmic bias and error in search engines pose to consumers and to respond in accordance with its mandated responsibilities. The following section provides a more comprehensive classification of the risks posed by CSE algorithms as they pertain to the jurisdictional authority of the FTC along with a sample of typical use cases to guide the policy discussion that follows.

II. CLASSIFICATION OF CONSUMER HARMS

Tarleton Gillespie's notion of a **public relevance algorithm** provides a useful lens through which to view the regulatory responsibilities of the FTC in relation to CSEs. He introduces the term in the following context—

As we have embraced computational tools as our primary media of expression, and have made not just mathematics but *all* information digital, we are subjecting human discourse and knowledge to [the] procedural logics that undergird all computation. And there are specific implications when we use algorithms to select what is most relevant from a corpus of data composed of traces of our activities, preferences, and expressions. These algorithms, which I'll call *public relevance algorithms*, are—by the very same mathematical procedures—producing and certifying knowledge (Gillespie, 2014).

The classification scheme used in this section correspond to the six dimensions of public relevance algorithms that Gillespie identifies as having political valence. For each dimension, I provide the original description he provides followed by one or more typical use cases.

(1) *Patterns of inclusion*: “the choices behind what makes it into an index in the first place, what is excluded, and how data is made *algorithm ready*.”

- ❖ Use case(s): CSE adopts a policy to collect and index all “street view” data, thereby exposing vulnerable populations to risk (collection problem); CSE improperly categorizes gay-friendly content as “adult” (categorization problems); CSE excludes or demotes results containing profanity from its “trending” reports (exclusion and demotion problems).

(2) *Cycles of anticipation*: “the implications of algorithm providers’ attempts to thoroughly know and predict their users, and how the conclusions they draw can matter.”

- ❖ Use case(s): A student who conducts research on the Alt-Right movement begins receiving pro-Alt-Right content in her searches (misrepresentation problem).

(3) *The evaluation of relevance*: “the criteria by which algorithms determine what is relevant, how those criteria are obscured from us, and how they enact political choices about appropriate and legitimate knowledge.”

- ❖ Use case(s): A user chooses which doctor to visit based on the “relevance” score assigned by the CSE (undefined criteria problem); CSE promotes its own products over those of its competitors, or the products of its highest paying advertisers’ websites (commercial aims problem); CSE adopts a “shareholder democracy” approach to page ranking, thereby propagating current political inequalities and power imbalances (epistemological premises problem).

(4) *The promise of algorithmic objectivity*: “the way the technical character of the algorithm is positioned as an assurance of impartiality, and how that claim is maintained in the face of controversy.”

- ❖ Use case(s): CSE fails to disclose that it has outsourced content moderation to human beings (promise of algorithmic objectivity problem).

(5) *Entanglement with practice*: “how users reshape their practices to suit the algorithms they depend on, and how they can turn algorithms into terrains for political contest, sometimes even to interrogate the politics of the algorithm itself.”

- ❖ Use case(s): Media companies begin writing more sensationalist headlines to attract more clicks on the CSE (algorithmic recognizability problem); CSE allows advertisers to bid on preferred placement (backstage access problem); teen users feel invisible and insignificant when a search for themselves on the CSE returns no results (domestication problem); users become less critical of the principles of

capitalism due to frequent use of CSEs that cast them in the role of consumers (knowledge logics problem).

(6) *The production of calculated publics*: “how the algorithmic presentation of publics back to themselves shape a public’s sense of itself, and who is best positioned to benefit from that knowledge.”

- ❖ Use case(s): Political polarization in the US increases after CSEs calculate that representing conservatives and liberals as more radical than they are leads to increased revenues from politically-themed ads (calculated publics problem).

III. OPPORTUNITIES

In this section, I present the case for regulation based not on consumer harms, but rather on the opportunities it affords in terms of facilitating the development and deployment of AI-related applications. In 2016, the Obama Administration published two reports identifying the development of AI technology as a foremost economic priority and calling for significant increases in federal funding to support research and development (National Science and Technology Council Networking and Information Technology Research and Development Subcommittee, 2016). To facilitate the smooth integration of these technologies into American society and to capitalize on the significant economic and societal potential they afford, an adequate regulatory framework urgently needs to be developed. According to Agrawal et al. (2016), “The bottleneck to deployment of AI-enabled, society-enhancing products and services in a number of domains is shifting from technology to regulation. ... Well-designed regulations will influence the rate and direction of innovation by creating incentives for the private sector to invest along trajectories that will most benefit society” (Agrawal, Gans, & and Goldfarb, 2016).

Regulation is also another area in which the US can differentiate itself and gain a competitive edge over other nations. As Agrawal et al. explain, “jurisdictions that are first movers in regulations that are friendly towards self-driving cars may attract significant investment, not only from car companies, but also from complementary industries (e.g., intelligent road infrastructure, delivery services, autonomous fueling services, disabled person services, etc.), even though that region is not a leader in AI research” (ibid.). By investing the effort now to develop policies adequate for the regulation of CSEs, the FTC will be positioning itself to more effectively confront the regulatory challenges posed by AI-enabled technologies in industries as diverse as medicine, finance, transportation, safety, and justice (ibid.).

IV. POLICY OPTIONS

Policy options are presented as follows. First, I provide a general case for adopting the proposed policy. Second, I specify the use cases the policy most directly addresses. Third, I offer a selection of specific regulatory mechanisms the policy might incorporate and rank these by good, better, and best. Finally, I highlight the challenges associated with the regulatory mechanisms proposed.

Option 1: Strengthen individuals’ control over their personal data

One way to mitigate the risks described above is to empower consumers to prosecute perceived misuses of personal data by CSEs. Arming consumers with the appropriate legal protections will put pressure on CSEs to act more responsibly with user data. In 2015, for example, the European Union successfully leveraged their General Data Protection Regulation (GDPR) in a lawsuit against Google, providing a model for how such legislation might be applied to the regulation of commercial engines more generally.

Presently, however, consumers in the US don't have any comprehensive data privacy law to draw on to formally contest or prosecute harms suffered on account of having their personal data misused (Greenleaf, 2014). This lack of legal protections over personal data makes US consumers especially vulnerable to the risks posed by CSEs and puts the US at a disadvantage when it comes to the development and deployment of future AI-related applications. As a first step to remedying this situation, the FTC should work to establish a provisional framework for these protections by experimenting with the application of existing laws (through strategic, targeted lawsuits, etc.) with the aim of establishing precedent that can be more effectively leveraged in these contexts. The FTC should also draw on the resources provided by International Human Rights law, specifically as it pertains to nondiscrimination, equality, political participation, privacy, and freedom of expression (Latonero, 2018).

The FTC should not stop here, however, but should work with Congress to develop comprehensive data protection laws that encompass both the private and public sectors. 120 countries and independent territories have already adopted comprehensive data privacy laws, with trends showing such laws spreading globally at an accelerating rate. In every case, these laws meet at least minimum formal standards based on international agreements (Greenleaf, 2017), and the framework that the FTC ultimately develops should do the same.

Use Cases Addressed

This policy will be most useful for addressing the use cases associated with dimensions (2) *cycles of anticipation* and (6) *the production of calculated publics*.

Regulatory Mechanisms

- a) **Good:** Grant individuals the right to know what personal data is being collected and how it is being used.

- b) **Better:** In addition to the above, increase allocation of agency resources and shorten window for investigating and arbitrating data abuse claims.
- c) **Best:** In addition to the above, grant users the right to curate their own “algorithmic identity” and, if desired, remove their data from the platform (“right to be forgotten”).

Challenges

- i. Allowing users to manipulate their algorithmic identity opens the door to deliberate deception and misrepresentation by users or may otherwise diminish the algorithms’ accuracy or effectiveness.
- ii. Increasing individual protections may not translate to equivalent levels of protection for larger populations (as required to address the calculated public problem).

Option 2: Require stricter monitoring and evaluation protocol

Another way to mitigate the harms caused by bias and error in CSE algorithms is to require the operators of these platforms to adopt stricter monitoring and evaluation protocols, which will serve as a legally binding agreement between the company and consumer. At minimum, this document should clearly specify *(a)* the CSE’s plan for identifying and addressing bugs in the system before they are allowed to cause too severe of damage; *(b)* a protocol for suspending activity in the event that risks become unmanageable; and *(c)* standards of fair practice that can be used to determine cases of malpractice and arbitrate allegations of malpractice.

Use Case(s) Addressed

This policy will be most useful for addressing the use cases associated with dimensions *(1) patterns of inclusion, (3) the evaluation of relevance, and (4) promise of algorithmic objectivity.*

Regulatory Mechanisms

- a) **Good:** Require CSEs to disclose whenever they (whether directly or via contracted workers) have engaged in active editing or moderation of search results, as well as the nature of and justification for those interventions.
- b) **Better:** In addition to the above, require CSEs to specify their product's scope of prescribed application (i.e., affirm what their product can and cannot do based on lab and field testing) and disclose any applications that are known to present risks (as well as the nature of those risks). Hold CSEs accountable for product failures within its prescribed scope through the assessment of fines and legal arbitration on behalf of consumers (Kroll, 2018).
- c) **Best:** In addition to the above, require CSEs to allow users the option of self-moderating according to their own preferences using a "menu" of filters and filter packages. Allow CSEs the option of continuing their current practices under a "default" option.

Challenges

- i. Requiring CSEs to disclose and justify their editorial decisions may be perceived as coercing them into the role of journalistic editors, thereby increasing their exposure to public backlash and possible lawsuits.
- ii. Given the complexity of the search algorithms, the option of "unbundling" filter features may not be feasible or may impose too high a cost on CSEs.
- iii. One or more of the mechanisms above may involve unknown practicalities that conflict with State and/or Federal trade secrets law and therefore require revision or redefinition.
- iv. One or more of the mechanisms above may increase the CSEs vulnerability to outside manipulation and gaming the system.

Option 3: Support self-regulation through Corporate Social Responsibility initiatives

The benefits derived from Corporate Social Responsibility (CSR) programs in terms of increased long-term profits and shareholder trust have proven a powerful force in incentivizing self-regulation at little or no cost to government. Such initiatives, however, are notoriously susceptible to abuse as companies accept all the gains while making very little of the alleged sacrifices. These risks are especially high in industries where there are no clearly-defined standards of best practices, as is the case in the present context. To discourage such corporate whitewashing, the FTC should take the lead in developing and publicizing best practices and CSR templates for companies in this industry. These recommendations should incorporate elements of the two previous policy options to maximize coverage of the use cases elaborated above.

Use Case(s) Addressed

This policy provides the most flexibility (and, therefore, leniency) over how CSEs choose to address the use cases described above. With the appropriate guidance and incentives, however, it is perhaps the only policy option with the potential to bring about comprehensive reform in the broader CSE industry with respect to the design and functionality of search-related algorithms. As such, it is the most ideally suited for application to use cases associated with dimensions (5) *entanglement with practice*, as well as the subtler examples from the other categories.

Regulatory Mechanisms

- a) **Good:** Without stipulating what terms the CSR document must include, require the CSE to conduct regular performance assessments to be submitted biannually to the FTC for review.
- b) **Better:** In addition to the above, make all CSE performance assessments public record.

- c) **Best:** In addition to the above, conduct periodic audits of CSEs and penalize any that submits fraudulent reports.

Challenges

- i. Without further action by the FTC, none of the terms of the CSR are legally binding or enforceable.
- ii. Audits can be costly and time consuming and penalties for fraudulent reporting may be inadequate to compel compliance.

V. CONCLUSION

In light of the sheer scale at which CSEs operate, the depth of dependence consumers have on the services they provide, and the severity of the risks associated with search algorithm malfunction, the FTC should pursue the most vigorous regulatory measures that, after further evaluation, it deems feasible. In accordance with **Policy Option 1(c) (“best”)**, the FTC should immediately begin working with Congress to develop comprehensive data privacy laws to bring the US into conformity with the broader global trend. These laws should draw on existing legislation and legal precedent but adjusted to accommodate the strictest regulatory mechanisms proposed above.

Furthermore, in accordance with **Policy Option 2**, the FTC should enter negotiations with CSEs to determine which of the proposed regulatory mechanisms provides the maximum protection for consumers (in terms of guaranteeing neutrality and objectivity) at the minimal cost to companies (in terms of increased editorial responsibility, public and legal liability, and exposure of sensitive trade secrets). Pending these negotiations yield no new information to challenge my initial weighing of these costs and benefits, I recommend **Option 2(c) (“best”)**:

distributing the editorial burdens (and accompanying liability) to users by allowing them the option to self-moderate using a menu of filters and filter packages, and otherwise maintaining the status quo under a “default” option. Unless one of these two policy options is deemed unfeasible, I do not recommend Policy Option 3 be pursued at this time.

I hope this letter will help provide the Chairman and the public with a greater sense of certainty concerning how the FTC will exercise its jurisdiction regarding the regulation of commercial search engines. Please do not hesitate to call if I can be of any further assistance.

Patrick Campbell
Policy Advisor

BIBLIOGRAPHY

- Agrawal, A., Gans, J., & Goldfarb, A. (2016, December 21). The Obama Administration's Roadmap for AI Policy. *Harvard Business Review*.
- Ahmed, A., Schulz, P. J., & Nakamoto, K. (2014). "The Impact of Search Engine Selection and Sorting Criteria on Vaccination Beliefs and Attitudes: Two Experiments Manipulating Google Output". *J Med Internet Res*, 16(4).
- Gillespie, T. (2014). The Relevance of Algorithms. *MIT Press*, 1- 32.
- Google. (2017, October 25). *Google Trends*. Retrieved from Google Trends: <https://trends.google.com/trends/yis/2017/US/>
- Greenleaf, G. (2014). Global Data Privacy Laws: 89 Countries, and Accelerating. *Social Science and Electronics*, 1-14.
- Greenleaf, G. (2017). Global Data Privacy Laws 2017: 120 National Data Privacy Laws, Including Indonesia and Turkey. *145 Privacy Laws & Business International Report*, 10-13.
- Kroll, J. A. (2018). The fallacy of inscrutability. *Royal Society*, 376(2133).
- Latonero, M. (2018). *Governing Artificial Intelligence : Upholding Human Rights & Dignity*. California: Data & Society.
- National Science and Technology Council Committee on Technology, National Science and Technology Council Committee on Technology. (2016). *Preparing for the Future of Artificial Intelligence*. Washington, D.C: Executive Office of the President.
- National Science and Technology Council Networking and Information Technology Research and Development Subcommittee. (2016). *The National Artificial Intelligence Research and Development Strategic Plan*. Washington, D.C.: Executive Office of the President.
- Osoba, O., & Welser IV, W. (2017). *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. Santa Monica, California: RAND Corporation.
- Statcounter. (2017, December 31). *Statcounter*. Retrieved from Global Stats: <http://gs.statcounter.com/search-engine-market-share>
- Statista. (2018, December 7). *Number of internet users in the United States from 2015 to 2022 (in millions)*. Retrieved from Statista.com: <https://www.statista.com/statistics/325645/usa-number-of-internet-users/>

APPENDIX 1. PLANS FOR FUTURE DEVELOPMENT

In order to convert this memo into a fully developed policy memo, I would need to do more in-depth research into the jurisdictional authority of the FTC to determine the extent to which the three policy options recommended above are feasible. I would be especially interested in reviewing the FTC case and proceeding files as well as the 70+ laws that the Commission is responsible for administering or enforcing in order to identify what specific laws and policies best map onto the present context. Along these same lines, I would also want to try to measure my recommended response to the nature of consumer complaints being filed with the FTC against commercial search engines. It may be difficult to motivate the case that regulation is needed to protect consumers if consumers aren't actually making complaints of this nature.

I also have unresolved questions concerning the potential role of the Federal Communication Commission in the implementation of my three policy options. If the FCC does have jurisdiction in these matters that I haven't properly represented, I would need to consider how to include them. For example, other policy researchers including Gillespie (see Gillespie 2014) have proposed that Section 230 (Safe Harbor) of the Communications Decency Act of 1996 may have application in the context of moderating the use of search engine algorithms, which would be particularly relevant to the fleshing out of Policy Option 2. In the case that these limited investigations fail to yield an appropriate legal basis for the proposed regulations, I would be interested to perform a more extensive search of court decisions in relation to the sorts of harms outlined in section 2 of this memo in the interest of establishing a more appropriate legal and regulatory framework for addressing these cases.

A third thing I would like to do to further develop this memo is to add a section that attempts to more clearly define what a commercial search engine *is*—or more specifically, how it

should be viewed by policy-makers in particular. This may take the form of a more traditional market categorization, as for example Napoli and Caplan attempt to do when they define these entities as “hybrid tech-media companies” (Napoli and Caplan 2017), or may even look outside of the market paradigm altogether, perhaps toward other closely related frameworks such as those applied to the regulation of utilities and public and collective goods.

Finally, as I continue to refine and narrow my problem definition to identify the specific companies that may be implicated so as to move forward with enforcement, I would want to do more thorough investigations into the potential costs associated with regulation, specifically in terms of effects on the job market, GDP, and other important economic indicators. In general, I’m a bit apprehensive about regulating companies like Google, for fear that it may limit or impair my future access to the service they provide. For that reason, given more time, I would want to look harder for regulatory options that provide a win-win for both consumers as well as CSEs.