

why are black women so



why are black women so angry
why are black women so loud
why are black women so mean
why are black women so attractive
why are black women so lazy
why are black women so annoying
why are black women so confident
why are black women so sassy
why are black women so insecure

ALGORITHMS OF OPPRESSION

HOW SEARCH ENGINES
REINFORCE RACISM

SAFIYA UMOJA NOBLE

Algorithms of Oppression

How Search Engines Reinforce Racism

Safiya Umoja Noble



NEW YORK UNIVERSITY PRESS

New York

NEW YORK UNIVERSITY PRESS

New York

www.nyupress.org

© 2018 by New York University

All rights reserved

References to Internet websites (URLs) were accurate at the time of writing. Neither the author nor New York University Press is responsible for URLs that may have expired or changed since the manuscript was prepared.

Library of Congress Cataloging-in-Publication Data

Names: Noble, Safiya Umoja, author.

Title: Algorithms of oppression : how search engines reinforce racism / Safiya Umoja Noble.

Description: New York : New York University Press, [2018] |

Includes bibliographical references and index.

Identifiers: LCCN 2017014187 | ISBN 9781479849949 (cl : alk. paper) |

ISBN 9781479837243 (pb : alk. paper)

Subjects: LCSH: Search engines—Sociological aspects. | Discrimination. | Google.

Classification: LCC ZA4230 .N63 2018 | DDC 025.04252—dc23

LC record available at <https://lccn.loc.gov/2017014187>

New York University Press books are printed on acid-free paper, and their binding materials are chosen for strength and durability. We strive to use environmentally responsible suppliers and materials to the greatest extent possible in publishing our books.

Manufactured in the United States of America

10 9

Also available as an ebook

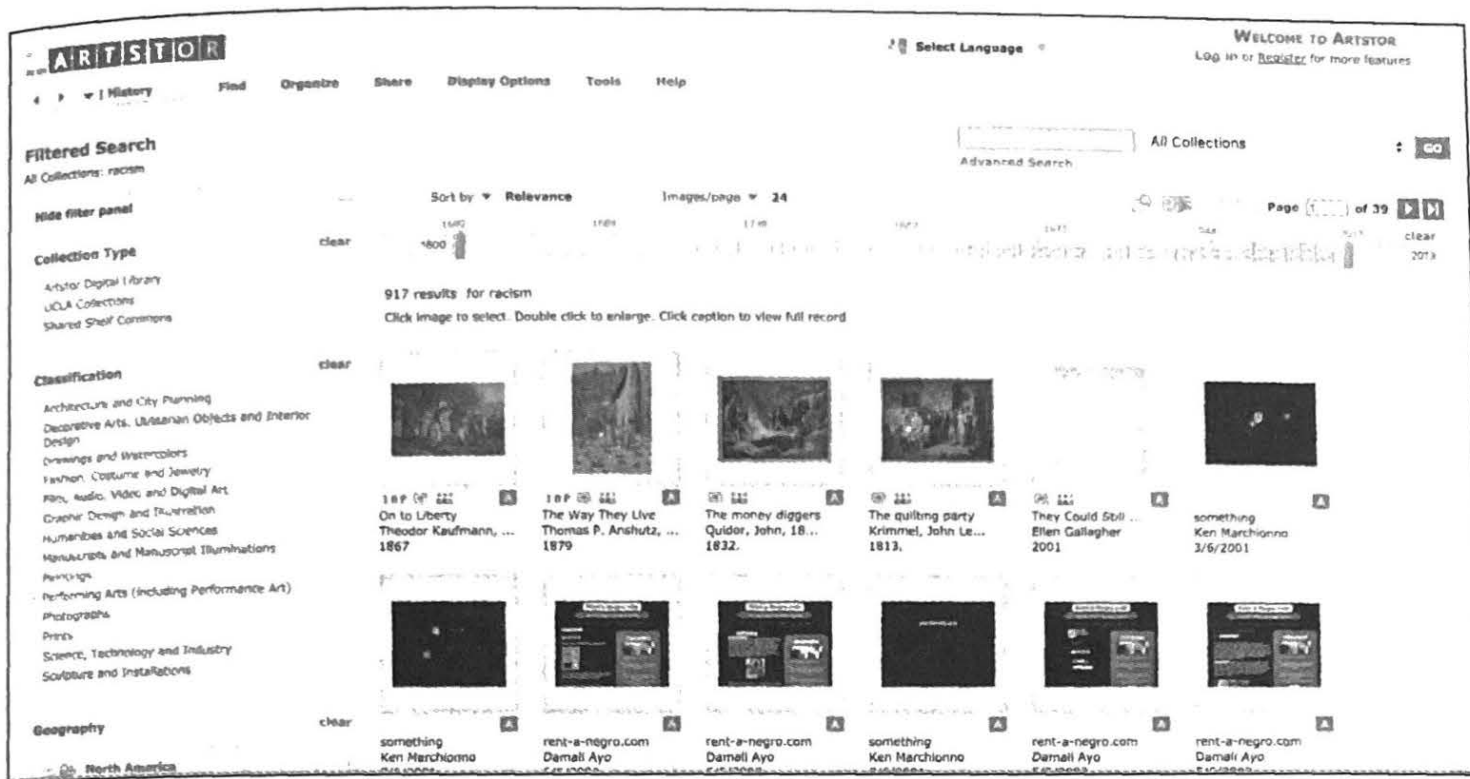


Figure 5.5. A satirical piece by the artist Damali Ayo and her online piece *Rent-A-Negro*, which is a critique of liberal racial ideologies that tokenize African Americans. The work is cataloged as “racism.”

think critically about the long-term consequences of misidentification of people and, in this case, concepts about works of art.

Search as a Source of Reality

Indeed, problematic results in ArtStor are just a small window into a long and troubled history of misrepresentation in library subject cataloging and classification systems, which are faithful reflections of the problematic representations in mainstream U.S. culture. Our ability to recognize these challenges can be enhanced by asking questions about how technological practices are embedded with values, which often obscure the social realities within which representations are formed. The interface of the search engine as a mechanism for accessing the Internet is not immune, nor impartial, to the concerns of embedded value systems. Search is also more than the specific mathematical algorithms and deep-machine learning developed by computer scientists and software engineers to index upward of a trillion pages of information and move some from the universal data pile to the first page of results on a computer screen. The interface on the screen presents an information reality,

while the operations are rendered increasingly invisible.³⁶ The media and communications scholar Alex Galloway destabilizes the idea that digital technologies are transparent, benign windows or doors providing a view or path to somewhere and in themselves insignificant—the digital interface is a material reality structuring a discourse, embedded with historical relations, working often under the auspices of ludic capitalism, where a kind of playful engagement of labor is masked in vital digital media platforms such as Google.³⁷ Search does not merely present pages but structures knowledge, and the results retrieved in a commercial search engine create their own particular material reality. Ranking is itself information that also reflects the political, social, and cultural values of the society that search engine companies operate within, a notion that is often obscured in traditional information science studies.

Further, new digital technologies may constitute containers for old media discourses, and the web interface (such as a plain Google search box) is a transitional format from previous media forms.³⁸ Certainly in the case of digital technology such as commercial search engines, the interface converges with the media itself. Commercial search, in the case of Google, is not simply a harmless portal or gateway; it is in fact a creation or expression of commercial processes that are deeply rooted in social and historical production and organization processes. John Battelle, who has carefully traced the history of Google, describes search as the product of our needs and desires, aggregated by companies:

Link by link, click by click, search is building possibly the most lasting, ponderous, and significant cultural artifact in the history of humankind: the Database of Intentions. The Database of Intentions is simply this: the aggregate results of every search ever entered, every result list ever tendered, and every path taken as a result. . . . This information represents the real-time history of post-Web culture—a massive clickstream database of desires, needs, wants, and preferences that can be discovered, subpoenaed, archived, tracked and exploited for all sorts of ends.³⁹

Undoubtedly, search is also pivotal in the development of artificial intelligence. In many ways, Google Search is an attempt to use computer science as a basis for sorting and making decisions about the relevance and quality of information rather than human sorting and web-indexing

practices—practices that search engine companies such as Yahoo! and those of the past invested in heavily and that were both expensive to implement and limited and less responsive in real time.⁴⁰

Providing Context for Information about People

In a narrow sense, information is a series of signals and messages that can be expressed through mathematics, algorithms, and statistical probabilities. In a broader sense, however, Tefko Saracevic, a professor emeritus of information science at Rutgers, suggests that information is constituted through “cognitive processing and understanding.”⁴¹ There is a pivotal relationship between information and users that is dependent on human understanding. It is this point that I want to emphasize in the context of information retrieval: information provided to a user is deeply contextualized and stands within a frame of reference. For this reason, it is important to study the social context of those who are organizing information and the potential impacts of the judgments inherent in informational organization processes. Information must be treated in a context; “it involves motivation or intentionality, and therefore it is connected to the expansive social context or horizon, such as culture, work, or problem-at-hand,” and this is fundamental to the origins of information science and to information retrieval.⁴² Information retrieval as a practice has become a highly commercialized industry, predicated on federally funded experiments and research initiatives, leading to the formation of profitable ventures such as Yahoo! and Google, and a focus on information relevance continues to be of importance to the field. Information science is essentially deeply entwined with the history of library science and has primarily been concerned with the collection, storage and retrieval, and access to and use of information. Saracevic notes that “the domain of information science is the transmission of the universe of human knowledge in recorded form, centering on manipulation (representation, organization, and retrieval) of information, rather than knowing information.”⁴³ This foregrounds the ways that representations in search engines are decontextualized in one specific type of information-retrieval process, particularly for groups whose images, identities, and social histories are framed through forms of systemic domination. Although there is a long, broad, and historical context for

addressing categorizations, the impact of learning from these traditions has not yet been fully realized.⁴⁴

Attention to “the universe of human knowledge” is suggestive for contextualizing information-retrieval practices this way, leading to inquiries into the ways current information-retrieval practices on the web, via commercial search engines, make some types of information available and suppress others. The present focus on the types of information presented in identity-based searches shows that they are removed from the social context of the historical representations and struggles over disempowering forms of representation. These critiques have been levied toward other media practices such as television and print culture. Whether human beings believe that the information delivered in search is relevant has consistently been the basis of judgment about information quality,⁴⁵ but what is underdiscussed is that retrieval of information in commercial platforms such as web-based search engines is not unique to the individual searcher. A web-based commercial search engine does not entirely “know” who a user is, and it is not customizing everything to our personal and political tastes, although it is aggregating us to people it thinks are similar to us on the basis of what is known through our digital traces.

Finding Culturally Situated Information on the Web

The field of LIS is significantly engaged in information classification and organization work, which can inform the framework for thinking about developing ICTs that are focused on surfacing prioritized results, such as the search engine. Critical race theory in this process of developing information-organization tools is of great value, particularly when thinking about the phenomenon of excessive recall of documents on the web that are irrelevant or decontextualized. Responses to the kinds of problematic biases in large commercial search engines are part of the growing motivation behind a host of culturally situated search engines that are emerging, particularly Blackbird (www.blackbirdhome.com), a Mozilla Firefox browser designed to help surface content of greater relevance to African Americans. Blackbird has been met with mixed reviews, from support and interest to wholesale rejection.⁴⁶ In any case, organizations and individuals are responding to the limits of

traditional commercial search engines through the development of such search engines. Identity-focused websites, a combination of web-based browsers and web directories, are emerging to prioritize the interests of specific communities on the basis of the human-curated practices of the past and can be seen in search engines such as BlackWebPortal (www.blackwebportal.com); GatewayBlackPortal (www.gatewayblack.com), which is based on international models such as JGrab, a Jewish search engine; BlackFind.com (www.blackfind.com); and Blackbird. Sites such as Jewogle (www.jewogle.com), which serves as an online encyclopedia of the accomplishments of Jewish people; Jewish.net (<http://jewish.net/>), which is used to “search the Jewish Web”; JewGotIt (www.jewgotit.com); and Maven Search (www.maven.co.il), which catalogs over fifteen thousand Jewish websites, have emerged in the hundreds, some tongue in cheek, across religion, culture, and national origin. Much of this is a response of communities that are seeking control over relevant content and representation, as well as access to quality information within racial or group identity.

One of the fundamental challenges for these culturally situated search engines is the way in which they make visible the contradictions and biases in search engines, which André Brock discusses in relationship to Blackbird. He notes that “Blackbird’s efforts to foreground African American content were seen as an imposition on the universal appeal of the internet, highlighting the perception of the browser as a social structure limited by Black representation.”⁴⁷ Brock’s work indicates that though there is a demand for culturally relevant Internet browsing that will help surface content of interest to Black people, its value works against norms on the web, making it less desirable.

Reproducing Social Relations through Information Technologies

Online racial disparities cannot be ignored because they are part of the context within which ICTs proliferate, and the Internet is both reproducing social relations and creating new forms of relations based on our engagement with it. Technologies and their design do not dictate racial ideologies; rather, they reflect the current climate. As users engage with technologies such as search engines, they dynamically co-construct content and the technology itself.⁴⁸ Online information and content

available in search is also structured systemically by the infusion of advertising revenue and the surveillance of user searches, which the subjects of such practices have very little ability to reshape or reformulate. Lack of attention to the current exploitative nature of online keyword searches only further entrenches the problematic identities in the media for women of color, identities that have been contested since the advent of commercial media such as broadcast, print, and radio. Noticeably absent in the discussions of search is the broader social and technical interplay that exists dynamically in the way technology is increasingly mediating public access to information, from libraries to the search engine.

Now, more than ever, a new conception of information access and quality rooted in historical, economic, and social relations could have a transformational effect on the role and consequences of search engines. It is my goal through this research to ensure that traditionally under-represented ideas and perspectives are included in the shaping of the field—to surface counternarratives that would allow for a questioning of the normalization of such practices. Rather than prioritize the dominant narratives, Internet search platforms and technology companies could allow for greater expression and serve as a democratizing tool for the public. This is rendered impossible with the current commercial practices.

What we need are public search engine alternatives, united with public-interest journalism and librarianship, to ensure that the public has access to the highest quality information available.