Is an Algorithm the Answer? Wattpad Books's Challenge to Publishing Infrastructure

Chelsea Humphries

Abstract

This article explores Wattpad's algorithmically defined publishing arm, Wattpad Books, and its claims to be a neutral and unbiased alternative to traditional publishing selection processes. Removing editorial bias with an algorithmic selector, however, does not eliminate bias entirely; rather, it shifts bias from editors to the creators of the algorithm. Wattpad Books thus presents a new and insidious bias in publishing, claiming a neutrality in its data-driven proprietary processes that it does not have.

Keywords: Algorithm, Bias, Publishing, Wattpad, Google

The infrastructure of literary publishing industries are founded upon human judgement. Writers create manuscripts that are then passed along a chain of individuals—agents, editors, designers, printers, and finally sales representatives—all of whom judge and manipulate the writing and its presentation in order to create an item that they think is most likely to sell. This is far from being a flawless system, and it incorporates human bias that even today continues to favour male authors over female, white authors over authors of colour (Ramdarshan Bold, 2018). Methods of self-publishing are means by which written content can circumvent this formal publishing infrastructure, and within the last ten years online creative writing platforms, like Wattpad, have showcased millions of open-access, self-published stories. Wattpad describes itself benevolently as “[giving] people from around the world an audience of millions” (“About”) and “put[ting] the control back in the hands of the audience” (Lau, 2016). It presents an alternative to the infrastructure of the publishing industry—and now, with the creation of its own algorithmically defined publishing arm, Wattpad Books, it is also challenging this industry’s structure. Wattpad Books purports to be a neutral and unbiased publisher, using an algorithm instead of editorial judgement to generate its list of titles. Although it advertises itself as neutral, upon investigation it is clear that this digital, computer-driven selection process only shifts the bias—it is no longer the fallible editors, but the creators of the algorithm and the algorithm itself that now represent an insidious, invisible bias in publishing. Wattpad Books is consequently just as biased as its traditional publishing counterparts, if not dangerously more so.

In January of 2019, Wattpad announced that it will be creating its own publishing company, printing books sourced from user-generated content as selected by an algorithm (“Wattpad launches Wattpad Books,” 2019). This algorithm is meant to undo the bias inherent in traditional publishing and bring to light “stories of diverse communities around the world” (“Wattpad Books”). Such a claim, however, assumes a neutrality that is not possible; Wattpad’s idealistic assertion that the algorithmic selection of stories to publish will “correct for the lack of representation in the publishing industry” (De León, 2019) does not recognize the human labour and biases that underlie both the algorithms and the infrastructure of Wattpad itself. As Bozga (2013) writes, algorithms are “not merely algorithms; they are complex systems composed of human operators and technology. Contrary to popular belief, humans do not only take part in developing them, but they also affect the way they work once implemented. Most of the factors that cause human bias in traditional media still play a role [in digital media]” (p. 210). Rather than dismantling publishing’s “monoculture,” reflecting the views of “editors who have similar backgrounds in some of the biggest cities in the world” (De León, 2019), Wattpad’s algorithmic publishing scheme sets the stage for a new monoculture—one on the heels of Google (“Traffic Report,” 2010), shaping the infrastructure in which creative writing is formed

and shared by manipulating data for profit in a way that presents itself as benign, but ultimately limits user agency and freedom of choice for the benefit of the company.

The “Story DNA Machine Learning Technology” algorithm created by Wattpad is described as being able to “scan hundreds of millions of stories [...] to find themes or elements that might determine a story’s commercial success” (De León, 2019). The unspoken assumption is that these “themes and elements” will be considered blindly—the authors’ sex, age, or nationality will not be a factor—in order to make selections free from bias. Nowhere, however, is this confirmed; the criteria that this technology will use in order to find these “themes or elements” are not outlined on any of Wattpad’s webpages. The algorithm that will be used is proprietary, and it is not transparent. While problematic in this way, Wattpad promotes its new publishing arm by using the rhetoric of neutral technology. As Safiya Noble (2013) writes, “At the heart of the public’s general understanding and trust in commercial search engines like Google, is a belief in the neutrality of technology—a technologically deterministic blind spot to the embedded social values in technology design itself—which only obscures our ability to understand the potency of misrepresentation.” This perceived neutrality effaces the reality that the algorithms created by companies like Google and Wattpad are in fact, human creations. Noble (2013) continues on to note that, at Google, “the ideas for search are a product not of the algorithm on its own (as it suggests in its explanation for why we get objectionable content on some searches) but as a product of its engineers.” Wattpad’s claims that it can use its Story DNA Machine Learning Technology to mitigate the human bias of traditional publishing are just not true; there is no such thing as technology free from human bias.

This bias continues, even when considering the “democratic” use of user data from Wattpad in informing the results generated by this technology. Wattpad Books aims to use “audience data from global reading trends” to “publish a mix of international Wattpad hits and hidden gems” (“Wattpad launches Wattpad Books,” 2019). This suggests the creation of a variety of books and a resulting large number of diverse choices for readers. The algorithmic selection of stories that are primed for “commercial success,” however, immediately suggests a limitation on the content that is actually chosen and produced as books; the stories selected by Wattpad run the risk of literally becoming formulaic, as determined by unspecified criteria, to the ultimate financial benefit of the company.

Furthermore, while the manipulation of “audience data” appears to be a democratic system of analysis in determining what the most publishable stories on Wattpad might be, putting “control back in the hands of the audience,” according to Wattpad’s CEO (Lau, 2016), this approach to publishing cannot be truly democratic until its methods of manipulation are rendered transparent. The audience data that is collected is unspecified, and how it is weighted algorithmically is unknown. Furthermore, the validity of using such audience data is not

assessed—and it is possible that this data existing within Wattpad's own infrastructure is biased as well.

As Bozgađ (2013) writes about bias existing in algorithmic filtering and personalization on social media platforms, it is important to consider the biases of individual users. It is possible that users may only follow like-minded individuals and share information with which they already agree; on Wattpad, this may result in certain stories being liked and shared more than others as readers explore what is already being shared and up-voted by those whom they know, leaving some stories that would otherwise fit Wattpad Books's publishing criteria (problematic as it might be) undiscovered. If users are not aware of how their data is being used by Wattpad in order to determine the stories that Wattpad Books will be publishing, it will not be possible for them to make informed decisions and interact with the writing that they encounter more critically, such that they are meaningfully involved in the publishing process. Instead, the use of user data comes to superficially represent user involvement and power, while the decision-making power is still wielded by Wattpad.

There is also no assessment of how user data might fairly or unfairly influence the selection of books to be published. For instance, could there be a need for Wattpad Books's algorithm to offer counter-measures that will balance results skewed in favour of English-language stories? As Ramdarshan Bold (2018) notes, "English is still the dominant language" on Wattpad, but "there is a clear demand for works in a variety of languages and works in translation." Similarly, while emerging genres, like fan-fiction are "attracting large readerships," the stories on Wattpad that receive the most followers and votes are by "authors who write within the popular genres" (Ramdarshan Bold, 2018). Without clear indication to the contrary, the Story DNA Machine Learning Technology used by Wattpad may perpetuate these trends, selecting English content in already popular genres without contributing to greater diversity of published content at all, despite Wattpad Books's stated goals. This is already evident when one examines the first six books that have been chosen for publication by Wattpad Books in the fall of 2019; they are all written in English and are all young adult fiction in popular genres like fantasy, romance, and thrillers ("Wattpad Books"). Although the authors are of varied backgrounds (being five female authors and one gay-identifying male author of unspecified ages from the United States, England, and New Zealand), the true diversity of the published material is questionable. It certainly does not reflect the fact that approximately 23% of Wattpad stories are not written in English and that nearly half of the Wattpad stories that have been picked up by traditional publishers in the past have been written in Filipino or a hybrid of Filipino and English (Ramdarshan Bold, 2018).

How might this skew the future creation of written content and the ways that readers interact with stories? A worst-case scenario is one in which this new algorithmic publishing

infrastructure devolves into a digital popularity contest. The selection of authors to publish may no longer be obviously biased, as it can be in traditional publishing, but it may rather be biased by authors' abilities to foster votes and likes, encourage reader interactions, and meet a commercial formula—one that is invisible and supposedly “neutral,” but is perhaps insidiously informed by the social biases inherent in human interaction and judgement, subsequently manifested in human technology, and influenced by the financial goals of Wattpad. Wattpad's Story DNA Machine Learning Technology masks these issues with rhetoric that promises to open up the publishing industry and encourage diversity, but this rhetoric is unsupported.

The only way to address systemic biases like those within the publishing industry is to generate discourse around them; from this discourse, solutions can then be offered openly and with the input of users—in this case, readers and writers. Until Wattpad Books's algorithm is made public and transparent, these conversations cannot happen. Like the subjective agents and editors of the traditional publishing world, Wattpad is currently keeping its decision-making processes to itself, and in doing so, it is only replacing one problematic part of the publishing infrastructure with another that is even more challenging to dismantle: an algorithm is harder to point at and blame for biased selection than a human.

The implications of Wattpad's algorithmic approach to publishing may be far reaching. Wattpad's “Master Plan” indicates a desire to use data-driven techniques, like the algorithm behind Wattpad Books, to generate entertaining content “across all formats, on multiple platforms” (Lau, 2016). As CEO Allen Lau (2016) writes, “with our technology, data, and community-driven approach to entertainment, we represent the future of the multi-billion dollar industry.” Wattpad is positioning itself to be to entertainment and storytelling what Google is to information searching—transforming from a “neutral purveyor of choice to an agent which preselects choices” (“Traffic Report,” 2010). Without critical assessment along the way, Wattpad's strategy may result in limiting the diversity of material that users are able to encounter. While Wattpad Books states that it will “introduce diverse, new, and undiscovered voices [...] find new genres, category-bending spins on traditional genres, narratives that tackle the experiences and struggles of marginalized communities, and [...] speak to the diversity of Wattpad's global community” (“Wattpad Books”), evidence in support of this is yet to be seen. Wattpad will profit from Wattpad Books and its Story DNA Machine Learning Technology, but readers and writers must beware the rhetoric of its “neutrality” and supposed democratic benefit until more is disclosed about how its algorithm is created and operates.

References

- About. *Wattpad*. (2019). Retrieved from <https://www.wattpad.com/about/>
- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3), 209-227. doi:10.1007/s10676-013-9321-6
- De León, C. (2019). Wattpad, the storytelling app, will launch a publishing division. *The New York Times*. Retrieved from <https://www.nytimes.com/2019/01/24/books/wattpad-books-publishing-division.html>
- Lau, A. (2016). The master plan. *Wattpad*. Retrieved from <https://company.wattpad.com/blog/2016/11/30/the-master-plan>
- Noble, S. (2013). Google search: Hyper-visibility as a means of rendering black women and girls invisible. *In Visible Culture*, (9). Retrieved from <http://ivc.lib.rochester.edu/google-search-hyper-visibility-as-a-means-of-rendering-black-women-and-girls-invisible/>
- Ramdarshan Bold, M. (2018). The return of the social author. *Convergence: The International Journal of Research into New Media Technologies*, 24(2), 117-136. doi:10.1177/1354856516654459
- Traffic Report: How Google is Squeezing out Competitors and Muscling into New Markets. *Consumer Watchdog*. (2010). Retrieved from <https://www.consumerwatchdog.org/report/traffic-report-how-google-squeezing-out-competitors-and-muscling-new-markets>
- Wattpad Books*. Retrieved from <https://books.wattpad.com/>
- Wattpad launches Wattpad Books, a new publishing division to bring diverse, data-backed stories to book-lovers everywhere. *Wattpad*. (2019). Retrieved from <https://company.wattpad.com/blog/2019/1/23/wattpad-launches-wattpad-books-a-new-publishing-division-to-bring-diverse-data-backed-stories-to-book-lovers-everywhere>