

DS 760

# Final Option II: Weapons of Math Destruction

---

Rick W. Lentz, MS, MS

3 December, 2017

---

## Internalizing Reality & Meaningful Contributions

Dr. O'Neil was a quant that observed system financial weaknesses while on deck at DeShaw during the financial crisis. Compartmentalization of human intellectual capital, stressful overnight support, and a realization of the ever profitable 'dumb money' trading counterparty, further tarnished the glamor of the \$400k/year quant job. The political decisions after the financial crisis seemed to be the final straw that caused Dr. O'Neil to quit Shaw and move to RiskMetrics in 2009.

After two years of 'rubber stamping' at RiskMetrics, Dr. O'Neil took a Data Scientist job at Intent Media. That work was focused on getting a few pennies value out of website surfers by predicting if they were there to price shop or buy now. The target value was to flip the price shoppers over to competitors and close those truly buying.

During these positions, Dr. O'Neil observed patterns of unsustainable algorithmic power concentration and exploitation. The association and conclusion was that these systems were contributing to long-term, detrimental societal effects. Now Dr. O'Neil is admirably [blogging](#), networking, researching, and reporting on these issues as a self-made algorithmist.


## 'Weapons of Math Destruction'

The book presented many situations that could be characterized by misused mathematical models, misinterpreted models, ignorance, greed, deceit, and flat out prejudice. I've selected a few of these; PredPol, job application screening, US News college rankings, \$50M in University of Phoenix's Google Adword spending, and value-added teacher rankings. Over the next sections, I've been asked to incorporate 12 distinct course topics (per the rubric), naturally and appropriately, while integrating and conveying my position and view on the book's situations listed above. It may seem like a whirlwind of details and ideas, but the effort to produce this product affected meaningful review and appreciation of the course's topics.

## Regulating Networks of Networks

Before jumping into ethical arguments, it is important to reflect on the changes 'going digital' has had on the human rights and freedoms many have sacrificed to protect. The internet is a technology that allows individuals to freely communicate and be both producers and distributors of information and content. This change bypassed traditional, concentrated mass media to distribute non-offensive, non-hateful information and provides decentralized access to a diverse network of uncurated sources of information. It facilitates free communication and formation of pure interest groups, largely independent of traditional geographic presence requirements.

This freedom is challenged from within, by a side pressing to create paywalls for the right to freely communicate on the internet. This side has deep pockets, is powerful, and crafts arguments that resonate strongly with many that don't truly understand what is in their best



interest. For example, one argument is that the government's should not be the one to regulate networks since, by design, it is not efficient and is slow, stifling to innovation, and not a financially efficient organization in comparison to privately managed organizations. Another argument is that people should want more freedom and opportunity by simply letting the free market 'self-regulate'. Further government discrediting leverages NSA and OPM leaks as examples why the government should have the power to regulate the internet.

I support Net Neutrality for the arguments of consumer choice, transparency, competition, and investment and innovation. I hope that US Courts continue support demonstrated in prior rulings, underscoring the importance of the internet as an essential communications and information platform for consumers. Recently, courts have recognized risks of broadband providers degrading the quality of online service data streams to extract tolls and fees from providing companies. Despite being a victim of the OPM breach, I fully support any law, program, or presidential directive that gives certain government agencies the needed authority, tools, and full-scope access (meaning both breadth and depth) needed to keep America safe.<sup>1</sup>

## When $n = \text{All}$

Going digital also means that many of our tools generate data that can persist for a long time. For example, most communications streams are digital voice, text, or video. In the book *Big Data*, Mayer-Schonberger and Cukier's introduced new challenges associated with the persistent electronic form of data, including the demise of the expert.

With so many new sources of data, simple correlations has yielded tremendous efficiency gains. This is because, when used on real-world, population-level data, correlation can help accurately predict what is next, but not why. Further, correlation identifies useful proxies, such as traffic data as a useful proxy for economic recovery. The what's next and proxy finding represents part of the loss of in the expert's value. I'll argue that there is still a need for the expert, since causality remains challenging (even after broadening the information scope to include the both the context and environment).

We recognize that with current tools all data can be taken into account and that it is no longer efficient to make decisions about what variables to examine by relying on hypotheses alone. Today's expert needs to have the skills of the statistician, software programmer, infographics designer, and storyteller. In a world where  $n$  equals all, value shifts to ideas and ownership of the data itself.

In the next sections, we address privacy and data ownership. These concepts are critical for understanding who captures value from information. Since data intermediaries will likely continue aggregating data from multiple sources, they can innovate directly or sell access to those that will gain the most.<sup>2</sup>

---

<sup>1</sup> Discuss the two sides to net neutrality

<sup>2</sup> What does Mayer-Schönberger and Cukier mean when they suggest that data science has brought about "the demise of the expert?" Why do they not think this is a bad thing?

## Privacy and Property Rights

In Floridi's, *Right to be Forgotten: A Diary of the Google Advisory Council Tour*, we observed key differences in the right to freedom of digital opinion and expression in Europe and the United States. The European Union has a history of efforts aimed at providing protections for citizens data, including laws on data gathering, processing, and retention. For example, a directive issued in 1998 limited the transfer of data to parties without adequate protections or standards, reducing the risks of theft and injury. While explicit in the EU, our system protects some categories of information but favors the self-regulating approach. I'll advocate for support of national level programs in addition to the self-regulating' approach due to simple fact that US has had an embarrassing run of private and government breaches. These demonstrate the true level of difficulty in protecting information and people from harm.<sup>3</sup>

Most sites and services have both a Terms of Use and Privacy Policy governing the bulk of individual data. For example, if you use Facebook for sharing images, videos or text, read someone else's stream, search, click a like button, etc, you are granting ownership of that use and data including its derivative and future value to Facebook.

The system created before digital provides rights protection through patents, copyrights, trademarks, and trade secrets. A patent is a limited property right relating to an invention in exchange for public disclosure of the invention. A copyright grants the creator of original work, such as writings, music, and art, an exclusive right to its use and distribution. A trade secret is a formula, practice, process design, instrument, pattern or compilation of information that is not generally known or reasonably ascertainable by which a business that can obtain an economic advantage over its competitors. A trade secrets protection is formed through non-disclosure, confidentiality, and, noncompete agreements.<sup>4</sup>

## Software Rights


Not to confound the true context of the upcoming ethical cases, but it is important to understand software rights and the arguments supporting both sides of the public and private ownership debate. Richard Stallman was an early advocate of software rights. He stated that one needs the ability to run a program, for any purpose, study how the program works and be able to adapt if for your needs. This included the right to redistribute copies and improve through released code.

Stallman applies this view from a sociotechnical perspective, showing that paid software causes psychosocial harm to the spirit of self-reliance. It prevents others from learning new skills and harms the spirit of scientific cooperation. He illustrates the harm through a make-work (work that counts for the GNP but is waste) example. A programmer at a bank was making something that

---

<sup>3</sup> The right to privacy and the right to freedom of opinion and expression seem to be weighed differently in mainland Europe and in the United states. Explain. Refer to the Floridi reading on this topic and explain.

<sup>4</sup> Describe the difference between copyrights, patents and trade secrets. How do these three apply to data science?



already existed but was proprietary. If the software was 'free', the make-work wouldn't be needed. Stallman argues that competition and 'lack of helping' damages social cohesion and human relationships. If a developer demands a hefty sum to perform the job, others will as well. His view is that people should help their neighbors but through ownership, (he calls this hoarding), we can end up sending the opposite message. He gives examples of reward someone for 'obstruction' or admiring them for the wealth they have. His point is that this pattern of behavior causes society to disregard a natural spirit of welfare for society in exchange for personal gain. "The antisocial spirit feeds on itself, because the more we see that other people will not help us, the more it seems futile to help them. Thus society decays into a jungle."

Stallman mentions that many people loved programming enough to break up their marriage. From an increase in business productivity perspective, he said that if it were free, we have broader adoption due to lower barriers to entry, more code examples, make-work reduction, and be free to make customizations and changes as the business grows.

On the other side, Johnson argued that software increases social utility and owners are justified for incentives to create and innovate. Johnson further argued that we have a moral obligation to obey the laws and stealing and using software is harmful. The argument is strongly supported by taking away the capacity to profit and through the rival advantage gained by unpaid use.

Johnson also used Locke's account, stating that developers can argue that the software they create is rightfully theirs because they produced it with their intellectual and physical labor. Appropriating the software without resource attribution is the same as making the developer their slave. Arguments stating that stealing software is illegal as well as protections under the patent and copyright systems very strongly supported Johnson's view for ownership rights.<sup>5</sup>

## Rights in Practice

We have very briefly covered software, individual, privacy, network and data rights. In the OkCupid case, we observe a few rights interacting.

OkCupid is a web service that allows users to share information about themselves (e.g. sexual history and preferences) and search for potential partners based on similarity over many topics. Researchers collected publicly available information and disseminated the aggregated data, creating a public outcry for greater privacy protection.

From a privacy perspective, the Fourth Amendment protection someone using this service may expect, doesn't apply. The Supreme Court ruled in *Smith v. Maryland* that an individual has no "legitimate expectation of privacy in information he voluntarily turns over to third parties." The cumulative effect of the public exposure and third-party doctrine renders data scraping, collection, and mining largely outside the scope of the Fourth Amendment protection.

From a network rights perspective, collecting and aggregating the data was allowed, by both researchers and data aggregating services. OkCupid's service was by designed to maximize

---

<sup>5</sup> Johnson and Stallman take opposite perspectives on whether software should be privately owned. Yet, both offer utilitarian arguments for their respective sides. Outline these two utilitarian arguments.

---

profitability and accurately match seekers. This Individual privacy rights exposure caused the site change the service by adding layers of protections that were not there earlier.

From a software rights perspective, OkCupid (and not the customer) was able to claim copyright protection of the information under the 1998 Digital Millennium Copyright Act. This resulted in a 'DMCA takedown' to parties hosting the aggregated data.

Here we can see that individual privacy rights were lost under the third-party doctrine and that aggregation of publicly searchable profile records was allowed. However, releasing the aggregated information publicly, even under research fair-use was not allowed due under OkCupid's DCMA protection.<sup>6</sup>

## Algorithms Shift Power

PredPol is a system designed to increase effectiveness of policing by programmatically scheduling policing activities to certain areas at certain times. The system relies on prior incident data rather than experts. O'Neil's main criticism is that biased source data creates biased systems. Further criticism points to the shift in power, from police organizations to the private company that writes the algorithm used in assigning resources.

O'Neil also made this argument for job applicant screening systems. These systems are designed to reduce the candidate pool by filtering out certain troublesome personalities. O'Neil showed the harm to candidates that were flagged based on an attribute signature and couldn't get a job since many employers used the same screening system (same algorithm).

The case of value-added teaching evaluations was that of policy based on flawed analysis. The idea with the system was that good teachers deliver good scores. O'Neil showed the harm to teachers that were weeded out by the system unfairly and without a sound basis. I agree with O'Neil's analysis and views. It is a cautionary tail regarding basing public policy on flawed analysis.


## Scale and Rankings

O'Neil criticized University of Phoenix's behavior on several fronts. As a for profit university, the Apollo Group spent \$50M on Google Adwords to promote admissions into certain programs. This amount as a percentage of education costs was much higher than other universities, suggesting that quantity over quality was the primary driver. O'Neil also pointed to the high costs relative to the diploma's respective career salary as predatory bias.

On the other end of the spectrum, O'Neil pointed to US News College Rankings as perpetuating top school's admissions practices. Namely, that successful families are able to afford to send children to their alma mater. The fear is that having high admission standards is creating a permanent underclass in society.

---

<sup>6</sup> Describe the situation that transpired at OKCupid. Who is to blame, in your opinion, for the security breaches? Explain. Do you think additional policies or laws need to be made to prevent this sort of thing in the future?



Before commenting in detail on these situations, I'll explore proposed tools useful for address these situations: frameworks, external oversight, self oversight

## A Need for Frameworks

We saw in the previous example that the digital world is changing the intersections of rights in sometimes surprising ways. We will also be presented with cases where a particular situation starts off as clear, but once analyzed, we find a need for clarification. This is termed called a conceptual muddle. The text provided the example of a supervisor reading a subordinate's email. Should the email be treated like correspondence on company letterhead stored in company files or more like a private phone conversation? Any time a prior activity category is used to cover a new activity, we can see conceptual muddles. The goals of those working on technology ethics is to formulate frameworks that provide clarity for new activities. The frameworks allow clarity and let organizations move forward to formulate and justify appropriate policies.<sup>7</sup>

## External Oversight and Censorship

We covered the concepts of government data search as well as the idea of an external algorithmist providing assurance of proper algorithmic use. The Harvard Law Review article proposed dog-sniffing criteria as a good model for regulating governmental searches of data. The criteria are: The sniff must only analyze information that is legally obtained; the sniff must only detect illegal activity; humans must not participate in any search until probable cause has been established by the sniff, and the sniff must have a low false-positive rate.<sup>8</sup>

As a software developer, I struggled with the idea of external algorithmists. The idea that people will continue to build algorithms and that algorithms will be in use long enough before improving that people will review them seems odd given the continual improvement cycle build in to current generation deep learning systems via reinforcement learning. I there may be some value for this role, but I will take any bet that the job I currently perform will no longer contribute meaningful value in 10 years.<sup>9</sup>


In addition to government and external algorithmist oversight, there is also the idea that data scientists should be licensed by external bodies. I disagree with this and find the argument that without licensing there is no requirement for heightened care and no concept of professional malpractice is invalid. One does not need a 'requirement' to care about the quality and lasting impact of one's work.

---

<sup>7</sup> Know Moor's notion of a conceptual muddle

<sup>8</sup> The Harvard Law Review article finds dog-sniffing a good model for regulating governmental searches of data. Explain.

<sup>9</sup> What are external algorithmists? What might they do? (See Mayer-Schönberger and Cukier.) Why do Mayer-Schönberger and Cukier see them in our future?



Take the open-source community we covered earlier. Hundreds of thousands of professionals think their work is good enough to warrant public scrutiny. The system promotes and improves good work products through adoption and refinement. The net benefit of these tools is higher than many commercial products. I further offer these objections: 1) Licensing offers little net economic value to society. The burden of the licensing barrier introduces artificial scarcity, which impacts an otherwise natural supply-demand balance. 2) It shifts the utility of the profession from an individual's net contribution to that of a role-based contribution. The role-based view changes the perspective to a semi-static set of activities. Without licensing there can be an implied set of expectations of continued learning and growth that vary based on an individual's capabilities and growth desires. 3) Licensing has dramatically inflated the cost of education while stagnated salaries for doctors, dentists, and pharmacists. Pharmacists, for example, earn the same whether they are a newly minted or have 30 years of stellar performance in the field. Finally, credentials, and Nanodegrees, coupled with a current public skills portfolio meet the net societal benefits argued by the author.<sup>10</sup>

Lastly, we need to explore the thresholds for external censorship. Weckert finds that one needs to meet two thresholds: universality of offensiveness and reasonable avoidability. For offensiveness, to be sufficient to warrant coercion, it should be the reaction that could be expected from almost any person chosen at random from the nation as a whole, regardless of sect, faction, race, age or sex. For reasonable avoidability, "No one has a right to protection from the state against offensive experiences if he can effectively avoid those experiences with no unreasonable effort or inconvenience." Weckert feels Feinberg's universality principle is unacceptable because it is implausible due to the act of singling out an individual from the group. Weckert feels Feinberg's avoidability principle is plausible since tortuous paths on the www are avoidable. He is referring to serious cases (e.g. revenge porn on Facebook) since individuals have little power to change what has already happened, 'where self-respect is at stake.'<sup>11</sup>

## Internal Oversight Framework

After covering several concepts relating to external oversight, we need to examine the arguments for companies naturally taking on this effort. Davis observed several significant cases showing how casual handling of business ethical inquiry resulted in deep financial losses. Davis further identified positive benefits for explicit ethical handling and routine within the company: 1) Greater and faster customer adoption 2) Having explicitly shared values speeds innovation and collaboration 3) Reduced long-term risks 4) Increased goodwill through alignment


Davis's alignment methodology framework is a cycle of inquiry, analysis, articulation, and action. Inquiry is described as the discovery and discussion of core organizational values. Analysis is the review of current data-handling practices and an assessment of how well they align with core

---

<sup>10</sup> Explain the benefits and drawbacks of licensing data scientists.

<sup>11</sup> Johnson's Ch. 6 as well as Wechert's "Giving Offense on the Internet" discuss the tension between allowing freedom of speech and limiting the causing of offense. How does Wechert propose to draw the line between what should be allowed and what should be banned? Why does he draw the line where he does?





organizational values. Articulation means the explicit, written expression of alignment and gaps between values and practices. Action is the activity of using tactical plans to close alignment gaps that have been identified and to encourage and educate how to maintain that alignment as conditions change over time. Davis would say that companies could achieve higher and faster customer adoption, speedier innovation and collaboration through commonly shared values, reduced business risks, and increased leadership status relative to competitors.<sup>12</sup>

## Applying the Tools to O'Neil's Cases

University of Phoenix spending \$50M a year on Adwords seems excessive when compared to other private schools. This focus on growth and high pressure car sales like tactics is going to get the company in legal trouble with the US Government. This trouble force the public company to sell to a private investment consortium earlier this year. The sale ousted several board members who pushed for the company's unfortunate excessive focus on growth. The new board would be wise to self regulate, promote an internal ethics framework and policies, and rebuild their reputation.


I found the technology platform and resources superior to many other schools and online programs in 2001. I personally benefited greatly from the knowledge gained through an MS Computer Information Systems. Over the years I have worked with instructors who taught a few courses each term for extra money. Looking ahead, with the senior leadership replaced, I feel O'Neil's consultancy could help ensure that the new senior leadership team address policy and governance in a way that would repair the institution's reputation.

Regarding O'Neil's argument against the US News College Rankings, I don't feel that system needs any current focus and the systemic underclass argument is overblown. Having completed undergrad under scholarship at the #2 public rated school doesn't bias my feelings towards the case. I believe that those that adopt a lifelong passion for learning can energetically find a path into the elite schools for graduate education.

I think that those damaged by the value-added education evaluation policy are winning court cases. State education departments that dismissed teachers based on the indisputably arbitrary and capricious evaluation metrics are being subject to court ordered reparations. Several of the lawsuits are looking to get higher courts to nationally stop the once federally championed, misused statistical process. O'Neil could certainly be a strong champion for these cases. A simple-to-set-up GoFundMe campaign would likely cover her team's time, travel and incidental costs.

---

<sup>12</sup> Kord Davis offers a method of aligning values and actions in a big data business. Describe two "current practices" talked about in Chapter 3 that might change if companies followed his method. What does he see as the benefit of doing this?



O'Neil's views on job application screening seem valid. I didn't read any supporting arguments suggesting criminal intent. Since only one screening company was used by multiple employers, O'Neil's consultancy may be able to offer feedback directly or through social media platforms to the screening company. If not effective, additional media attention following a show the harm strategy may help raise awareness and educate both the public and companies using the screening service.

PredPol doesn't cross any boundaries in my mind. The intent seems to be to help departments use applied prediction models for scheduling resources. There is room for improvement by taking in input from the front line officers. Looking ahead, I see far greater change happening to many professions as we apply deep learning, machine learning and AI based systems to automate tasks. These advancements are under central review by the London based [DeepMind Ethics and Society](#) group. I'd strongly advocate for those serious about working on the most significant ethical issues of our lifetime to follow the work and perhaps join that team.

## Conclusion

I started by reviewing the impacts going digital and datafication has had on society. Correlation is a good predictor of what is next - but not why. I reviewed Net Neutrality, individual privacy, software ownership, geographical data protection differences, and the OkCupid case in detail.

We reviewed several of O'Neil's cases from both a scale and power shift perspective before reviewing the external and internal governance concepts. From this foundation of ideas, I provided my own opinion on each case as well as where I see the greatest challenges ahead.

—