reflektif
SOSYAL BİLİMLER DERGİSİ
JOURNAL OF SOCIAL SCIENCES

2025, Vol. 6(1), 259-268
© The Author(s) 2025
Article reuse guidelines:
https://dergi.bilgi.edu.tr/index.php/reflektif
DOI: 10.47613/reflektif.2025.212
Article type: Commentary Article

Received: 20.01.2024
Accepted: 24.01.2024
Published Online: 25.02.2025

Mahmut Özer*

Can Mathematical Models Be Weapons of Mass Destruction?
Matematiksel Modeller Kitle İmha Silahları Olabilir mi?

Abstract

With the widespread adoption of digitalization, mathematical models have become indispensable in every field. Predictions are made, processes are evaluated and optimized, and future forecasts are developed using mathematical models. The usefulness of these models has accelerated their proliferation. Their application in nearly every aspect of life has triggered a new phase of modeling, where the output of one model can now serve as the input for another, enhancing overall efficiency. Consequently, models are no longer discrete but interconnected, encompassing and influencing human life. At this point, understanding how models operate is critically important for grasping how decisions affecting us are made. Therefore, in this study, mathematical models and algorithms are examined in detail based on Cathy O'Neil's (2016) book '*Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*'. It is emphasized that a model does not encompass everything related to a given field, and therefore prioritizes aspects of the field, assigning weights externally during this prioritization. As a result, every model provides only an approximation for the field, meaning that elements not measurable within the model risk losing value over time. The biases present in the dataset used by a model can lead to biased outputs, thereby reproducing existing inequalities in society. It is particularly emphasized that the fact that models now serve as inputs for one another weakens the possibility of correcting biased outputs and increases the risk of further deepening inequalities. This risk is expected to grow significantly, especially with the widespread adoption of artificial intelligence technologies. Therefore, the study recommends adopting a participatory management approach during the development phase of mathematical models, enabling the involvement not only of domain experts but also of representatives of all stakeholders directly affected by the model. This approach could help prevent the use of biased assumptions and datasets in the models, thereby mitigating the potential negative impact caused by these models.

Öz

Dijitalleşmenin yaygınlaşmasıyla matematiksel modeller her alanın vazgeçilmezleri oldu. Matematiksel modeller ile kestirimler yapılmakta, süreçler değerlendirilerek optimize edilmekte ve geleceğe yönelik kestirimler yapılmaktadır. Modellerin kullanışlılığı yaygınlaşmasını hızlandırdı. Özellikle yaşamın her alanında kullanılması, artık modellerin verimliliğini artırmak için bir modelin çıktısının başka bir modelin girdisi olabildiği yeni bir modelleme fazını tetikledi. Dolayısıyla, modeller artık ayrık değil birbirleri ile bağlantılı çalışmakta ve insan yaşamını kuşatmaktadır. Gelinen noktada, modellerin nasıl çalıştığını anlamak bize yönelik kararların nasıl alındığını anlamak açısından oldukça kritiktir. Bu nedenle bu çalışmada, Cathy O'Neil'in (2016) 'Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy' kitabına dayalı olarak matematiksel modeller ve algoritmalar ayrıntılı olarak değerlendirilmektedir. Bir modelin söz konusu alanla ilgili her şeyi kapsamadığı, dolayısıyla alanla ilgili önceliklendirme yaptığı için her modelin alan için sadece bir yaklaşıklık sağladığı, dolayısıyla modelde ölçülemeyen şeylerin zamanla değer yitirme riski taşıdığı vurgulanmaktadır. Modelin öğrendiği veri setinin yanlılıklar içermesi, çıktıların da yanlı olmasını sağlayarak toplumda var olan eşitsizlikleri yeniden üretebilmektedir. Özellikle, modellerin artık birbirlerine girdi sağlamasının yanlı çıktıların düzeltilme imkânını zayıflattığı ve eşitsizlikleri daha da derinleştirme riskini artırdığı vurgulanmaktadır. Özellikle yapay zekâ teknolojilerinin yaygınlaşması ile bu risk çok daha fazla artmaktadır. Bu nedenle çalışmada, matematiksel modellerin geliştirilme aşamasında sadece alan uzmanlarının değil, ayrıca modelden doğrudan etkilenen tüm paydaş temsilcilerinin katılımına imkân veren katılımcı bir yönetim yaklaşımının benimsenmesi önerilmektedir. Böylece, modeldeki yanlı varsayımların ve yanlı veri setlerinin kullanımının önüne geçilebilmesi mümkün olabilecek ve modellerin yol açabileceği olumsuz etkiler hafifletilebilecektir.

* National Education, Culture, Youth and Sports Commission, Turkish Grand National Assembly, mahmutozer2002@yahoo.com, ORCID: 0000-0001-8722-8670.

Keywords
Mathematical model, algorithm, artificial intelligence, bias, equality

Anahtar Kelimeler
Matematiksel model, algoritma, yapay zekâ, yanlılık, eşitlik

Introduction

Digitalization has not only led to the generation of big data in every field but has also facilitated the creation of metrics related to various domains and processes using this data. These metrics have accelerated the assessment of the current state, optimization, and future projections in every area (Erdi, 2020). Today, evaluations are conducted based on numerical values across fields ranging from education to healthcare, from the economy to politics. People now make their choices based on rankings created within this context. From the books they read to the items they consume, from the restaurants they dine at to the universities they attend, these numerical values play a significant role in shaping individuals' life choices. For businesses, rankings and metrics serve as guides in determining who to hire, shaping current business processes, and planning future investments. Banks use these metrics when granting loans. University rankings, financial risk factors of individuals, companies, and even countries are all produced through such approaches.

Of course, mathematical models are developed to generate meaningful results from numerical data. The backbone of these models is algorithms. Algorithms determine proxy features related to the field based on their objectives and identify the weights of these features in the intended outcomes (Erdi, 2020). New numerical values related to the field are generated based on assumptions and the proxy features used, and decisions are made based on these values.

It is helpful to first highlight two fundamental characteristics of algorithms. The first is that they provide approximation to understand the context being evaluated. This is because it is not possible to measure everything related to the context for which numerical values are desired. Instead, an attempt is made to approximate the context using proxy features assumed to be significant. Consequently, prioritization is carried out based on assumptions during modeling. Every model operates according to this prioritization. Therefore, every mathematical model or algorithm only provides an approximation to the context.

Naturally, this comes at a cost. On the one hand, the context is often not fully captured; on the other hand, aspects of the context that cannot be measured increasingly become trivialized and ultimately devalued. Alternatively, the risk of manipulation of the measured indicators rises. In short, mathematical models create long-term deformations in the areas they aim to measure.

Models guide the real world through their priorities and assumptions. Consequently, they reshape the real world based on the results they produce. The reflections of these results become more visible in the real world. In fact, this is facilitated by two auxiliary factors: scale and the interdependence of models. As the audience or system influenced by a model grows, the scale of the model expands. Additionally, models are no longer isolated from each other. Outputs generated by a model in one context can serve as inputs for a model developed in another context. This characteristic accelerates the growth of scale. Therefore, in this article, mathematical models and algorithms are comprehensively evaluated based on Cathy O'Neil's book *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (2016).

Every Model Provides an Approximation

Models are inherently designed to achieve a simplified representation of the real world. Therefore, they have limitations. These simplifications are necessary to optimize the model's capacity to perform a specific task, but they also lead to scenarios where the model diverges from reality, potentially producing incorrect results. For this reason, it is essential to be aware of these blind spots and to understand the limitations of a model when evaluating and using it (p. 20).

> To create a model, then, we make choices about what's important enough to include, simplifying the world into a toy version that can be easily understood and from which we can infer important facts and actions. We expect it to handle only one job and accept that it will occasionally act like a clueless machine, one with enormous blind spots.

Algorithms attempt to capture the reality they aim to represent through proxy features. However, since it is impossible to fully capture reality through these features, there is always a margin of error. While this margin of error may appear mathematically reasonable, each error can have a devastating impact on the individuals it affects (pp. 17–18). The biggest problem lies in how accurately the assumptions used in algorithms represent the reality being sought. As O'Neil points out, in decision-making algorithms, scoring replaces the reality it is supposed to represent: *"Instead of seeking the truth, scoring begins to represent the truth"* (p. 7). As a result, when the representation is flawed and the outcome is contested, you must provide far more evidence than the algorithm ever had to prove your case. In this context, O'Neil frequently refers to a model that measures teachers' performance based on their students' academic achievements. Although this model disregards many contributions—some of which are not measurable—when assessing a teacher's actual performance, it is often preferred by education administrators because it provides a support model that is difficult to question (p. 21).

Thus, this type of model is open to criticism because it does not evaluate all aspects and contributions of teachers, nor does it fully reflect the complexity of educational processes. However, it is preferred because it offers a quick and practical solution to meet the administrative needs of the education system. This shows that models are shaped according to the goals and priorities of their creators and that, in achieving these goals, they may sometimes overlook important elements. In fact, since approximations are made during modeling, things that cannot be measured are excluded from the model. As a result, what cannot be measured is devalued through models. Why should importance be given to things that the model does not value by failing to measure? On the other hand, while creating models, priorities are taken into account, and the features considered are determined according to these priorities. Therefore, interventions in models and algorithms are made based on values (p.21):

> Models are opinions embedded in mathematics. Whether or not a model works is also a matter of opinion. After all, a key component of every model, whether formal or informal, is its definition of success. This is an important point that we'll return to as we explore the dark world of WMDs. In each case, we must ask not only who designed the model but also what that person or company is trying to accomplish.

In short, when evaluating a model, it is important to understand the values and priorities underlying its design. Elements such as which data are included or excluded, which variables are considered more important, and which outcomes are targeted directly shape the structure and results of the model. Although the mathematical nature of models gives the impression that they are objective and precise, it should not be overlooked that models are shaped by human choices and values. More critically, the priorities of a given field begin to shift based on what is valued in rankings. Institutions, striving to improve their scores on indicators measured in rankings, gradually devalue unmeasured aspects, which eventually fall into obsolescence. Furthermore, when those conducting the rankings modify the indicators or adjust their weightings, institutions quickly attempt to adapt themselves to the new circumstances.

These metrics also open the door to manipulations. For instance, since publication and citation performance are considered in university rankings, financially strong institutions may strike high-paying agreements with highly productive researchers in these areas. Consequently, these researchers are not required to spend more than a short period each year at the contracting institutions but list those institutions in their publications' affiliations. Additionally, when everyone places the same emphasis on identical indicators, it harms diversity, causing the ecosystem to drift away from variety and toward standardization. Therefore, such rankings deform ecosystems, such as the higher education ecosystem. As the scale of using such models expands, the scale of deformation—and ultimately the extent of the damage—also increases (p.54):

Self-Fulfilling Prophecy

These models, due to their biases based on race, gender, religion, culture, and socioeconomic status, place the advantaged in an even more favorable position while further disadvantaging the underprivileged—a phenomenon known as the Matthew Effect (Merton, 1968; Ozer, 2023a; 2023b; 2024a; Perc, 2014). Despite the assumption that such systems operate impartially, given the nature of mathematics, they in fact continually increase the advantages of socioeconomically privileged groups while making it nearly impossible for the disadvantaged to break this cycle. Moreover, these models not only systematize these biases but also render them invisible. In other words, algorithms reinforce injustice by repeating the conscious or unconscious biases reflected in past human decisions. This leads to the embedding of past discrimination and biases into algorithms, perpetuating such discrimination on a broader and more systematic scale. For example, it has long been known that models used in the justice system exhibit racial discrimination (p.24).

Therefore, these models exhibit the characteristic of a self-fulfilling prophecy. In disadvantaged cases, whether it is a job application, hiring process, or sentencing based on the likelihood of reoffending, individuals are evaluated not by what they have done but by where they belong. As a result, their socioeconomic status becomes their destiny. Based on their past and the neighborhood they live in, the model may determine a higher likelihood of reoffending, resulting in a longer sentence. After imprisonment, the likelihood of finding a job decreases, financial difficulties trigger other familial and social problems, and ultimately, without the chance to start afresh, the individual becomes involved in another crime, receiving a much longer sentence than they originally deserved. In the end, the model has seemingly validated itself—not because it was accurate, but because it unjustly amplified the disadvantages of the underprivileged! (p.48). Such models also embed an additional unjust assumption into their algorithms: the presumption that sentencing individuals with a higher likelihood of reoffending to longer prison terms is more beneficial for society (p.97-98).

A similar situation is observed in ranking models as well. Those who rank lower in one evaluation are doomed to remain at the bottom in subsequent rankings. As the scale of these rankings grows, the harm suffered by the disadvantaged continuously increases, and disadvantage gradually turns into a predetermined destiny through these rankings. O'Neil cites the ranking initiatives of higher education institutions as an example of this phenomenon (p.53):

> U.S. News's first data-driven ranking came out in 1988, and the results seemed sensible. However, as the ranking grew into a national standard, a vicious feedback loop materialized. The trouble was that the rankings were self-reinforcing. If a college fared badly in U.S. News, its reputation would suffer, and conditions would deteriorate. Top students would avoid it, as would top professors. Alumni would howl and cut back on contributions. The ranking would tumble further. The ranking, in short, was destiny.

Another situation is observed in models used to determine patrol areas for efficiently allocating police resources, which are widely implemented in the United States to identify high-crime areas (p.86). These models, relying on recorded crimes as training data, discriminate based on race and socioeconomic status. They increase the number of patrols in areas densely populated by Black and disadvantaged communities. This leads to a vicious cycle where crime recording rates in these areas continually rise, creating a self-fulfilling prophecy. However, when white and socioeconomically advantaged communities face similar issues, crimes are often resolved without being officially recorded. Consequently, the number of patrols in these areas remains lower, and the likelihood of apprehending offenders is significantly reduced. Similarly, for example, it has been observed that 85% of individuals stopped by the New York Police Department due to suspicion are young African American or Latino men (p.92). In other words, those who are initially advantaged in crime records maintain this advantage after the model is implemented, while the disadvantaged are trapped in a vicious cycle, being pushed into an even more disadvantageous position.

Once you are caught in this spiral, your likelihood of reoffending increases, leading to harsher sentences. After serving your sentence, your chances of finding a job diminish. Again, your likelihood of being accused in a random stop rises. Models, now pervasive in all areas of life, continually push you into a disadvantaged position and ultimately determine your fate. In short, the groups harmed by these models are, strangely enough, always the socioeconomically disadvantaged—the poor (p.97). Meanwhile, for example, the financial sector, due to its economic importance and political lobbying power, is often subjected to less scrutiny. This can, in some cases, lead to high-profile crimes going unpunished (p.89-91).

Every day, a new mathematical model is introduced. These models, which initially appear to be implemented with good intentions, increasingly begin to extract data from individuals across various domains. As the diversity of data expands, it becomes evident that a model claimed to focus on one specific area also utilizes data from outside that domain, determining outcomes not only based on that field but also by incorporating data and performance from many other areas. At this point, individuals become completely vulnerable to the opaque, unjust, and harmful results produced by the model. This creates a feedback loop that reinforces injustice and systemic bias. Such biases in the system lead to continuous unfair treatment of certain segments of society. However, justice has no equivalent in the code of these models (p.199-200). Since models are data-hungry, new input are continually added to their datasets, making the cycle even more destructive with each passing day (p.175-176).

Another characteristic of these models is that they do not evaluate individuals independently but based on their surroundings, social networks, and geographic locations. In other words, individuals are assessed within the context of their socioeconomic status, which reinforces the disadvantages of these environments and makes it nearly impossible for individuals to break out of them. Thus, it becomes evident that justice is not merely a legal concept but one deeply intertwined with social and economic factors (p.146-147).

In fact, the self-fulfilling prophecy arises from the existence of numerous models that now guide life, most of which contain similar biases, unjustly scaling people's destinies. When the scale grows to affect the entire ecosystem, the ranking game that unfolds perpetually accredits the advantaged by keeping them at the top of the list while forcing others into a competition they cannot win. In this game, whose rules they did not set, the disadvantaged are consistently clustered on the losing side. O'Neil describes this situation as the exponential growth of the model and its capacity to scale across all aspects of life (p.29-30).

Meanwhile technological transformations, such as the widespread adoption of automation, are reducing the demand for mid-skilled workers in the labor market, gradually pushing them into the pool of low-skilled workers (Markovitz, 2019). Consequently, on one hand wages for low-skilled workers are continuously declining, on the other hand most businesses prefer part-time workers over full-time employees for low-skilled jobs (Ozer, 2024b). Additionally, with the advancement of models, such workers are constantly monitored for productivity, forcing them into inhumane working conditions (p.128).

Vulnerable Populations Exposed to Errors

When assumptions about purpose are made and a value is generated through a model, even if the model operates efficiently, it can still produce erroneous results (e.g., false positives). While these errors may affect only a small number of individuals, they can have a devastating impact on the lives of those individuals. A model that works well for the majority but produces biased or incorrect results for a minority raises the critical question: Can this cost be sacrificed for efficiency? According to O'Neil, since these algorithms are designed to encompass large populations, the wealthy are often excluded from such evaluations.

The information of the masses is collected and recorded in every possible way. In the marketplace, this information is purchased by companies and added to their existing data collections. Naturally, in an environment with such vast amounts of data, the market's concern is not whether the data is accurate but whether it is usable (p.151).

Most of the time, even though there are errors in this data, correcting these mistakes is only possible if the affected individuals investigate why they were impacted or have the means to uncover it (p.152-153). Others, unaware of how their profiles are created, are doomed. The Matthew Effect becomes even more destructive through such models, and disadvantages become unbearable (p.155). Since models are applied to large-scale populations, the errors or injustices of algorithms are not taken seriously, as they are considered individual cases and do not compromise the model's overall efficiency. Therefore, they are largely ignored, allowing the manipulation of the masses to continue unchecked (p.111). At this point, people have become vulnerable to major tech companies, making them susceptible to all kinds of manipulation (p.181).

Discussion

Mathematical models and algorithms now encompass all areas of life. As highlighted in the study, if left unregulated, these models can, as O'Neil emphasizes, turn into destructive weapons for the masses. The failure to verify the accuracy of data, the inclusion of biased data, and the assumptions and prioritizations made during algorithm development can exacerbate their destructive impact. These models can reinforce biases, systematically exclude certain groups, or create a false sense of security and justice. Most importantly, the models not only negatively affect relatively disadvantaged socioeconomic groups, but they also deeply impact and pull down the middle classes, which have long been losing ground due to automation.

Mathematical models approach the target domain using only measurable proxy features that have numerical values. Therefore, each model offers only an approximation for the domain. The degree of representational authority granted to each proxy feature is determined by algorithms. As a result, algorithms operate based on a prioritization (valuation). As the value attributed to the model increases, unmeasurable aspects of the domain become progressively devalued. Ultimately, this deforms the domain, reconstructing it around the measurable elements, valued attributes, and proxy features. When biases exist in the data used by the model or in the assumptions embedded in the algorithm, the decisions made by the model also become biased.

The exponential advancements recently observed in artificial intelligence (AI) have rapidly paved the way for AI applications across all aspects of life, from education to healthcare, from finance to the defense industry, leading to the swift formation of an AI ecosystem (Ilikhan et al., 2024; Ozer, 2024c; Ozer, 2024d; Perc et al., 2019; Tanberkan et al., 2024). AI has increased automation in labor markets, contributing to rising unemployment (Ozer and Perc, 2024), while also deepening the disruptive effects of mathematical models highlighted by O'Neil. The issues O'Neil points out in mathematical models—lack of transparency, scalability, and destructive impacts—are greatly amplified by the capabilities of AI. For this reason, countries are already striving to implement measures to address the negative effects that AI is expected to have on employment (Ozer et al., 2024a). On the other hand, efforts are also being made to prevent AI from exacerbating social inequalities, particularly through biased outcomes influenced by factors such as socioeconomic status, race, gender, and religion.

The current state of artificial intelligence enhances the capabilities of these models and carries the potential to significantly amplify their destructive effects if left unchecked. Therefore, a participatory approach must be adopted in the development of models, ensuring the active involvement of relevant experts, especially those groups likely to be affected by biases, as well as unions, civil society organizations, and stakeholder representatives throughout all processes (Ozer et al., 2024b). This can help mitigate the adverse impacts of these models to some extent. Using a participatory approach during the development phase of models will also enhance the accountability and transparency of mathematical models.

Since the models are no longer isolated, they interact with one another; when the output of one model becomes the input for another, the scale, scope, and disruptive impact of models continuously grow. In this context, the decisions made by models perpetuate the past into the future, making it increasingly difficult for disadvantaged groups to break free from cycles of disadvantage. For this reason, the use of big data and algorithms must be developed in a fair and inclusive manner, tested continuously, and designed to address rather than amplify existing biases. Greater transparency, regulation, and the establishment and enforcement of ethical standards are necessary to ensure individual justice in the design and implementation of these models.

References

Erdi, P. (2020). Ranking: The Unwritten Rules of the Social Game We All Play. *Oxford University Press*.

İlikhan, S., Özer, M., Tanberkan, H., & Bozkurt, V. (2024). How to mitigate the risks of deployment of artificial intelligence in medicine? *Turkish Journal of Medical Science*, 54(3), 483-492.

Markovitz, D. (2019). The Meritocracy Trap: How America's Foundational Myth Feeds Inequality, Dismantles the Middle Class, and Devours the Elite. *Penguin Press*.

Merton, R. K. (1968). The Matthew effect in science. *Science*, 159, 56-63.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. *Penguin Books*.

Ozer, M. (2023a). The Matthew effect in Turkish education system. *Bartın University Journal of Faculty of Education*, 12(4), 704-712.

Ozer, M. (2023b). Matta etkisi. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 19(4), 974-984.

Ozer, M. (2024a). The Matthew effect in the game of success and asymmetrical distribution of reward. *Reflektif Journal of Social Sciences*, 5(1), 187-197.

Ozer, M. (2024b). Dynamics of the meritocracy trap and artificial intelligence. International Journal of Management Economics and Business, 20(3), 845-869.

Ozer, M. (2024c). Potential benefits and risks of artificial intelligence in education. *Bartın University Journal of Faculty of Education*, 13(2), 232-244.

Ozer, M. (2024d). Impact of ChatGPT on scientific writing. *The Journal of Humanity and Society*, 14(3), 210-217.

Ozer, M., Perc, M. (2024). Human complementation must aid automation to mitigate unemployment effects due to AI Technologies in the labor market. *Reflektif Journal of Social Sciences*, 5(2), 503- 514.

Ozer, M., Perc, M., Suna, H. E. (2024a). Artificial intelligence bias and the amplification of inequalities in the labor market. *Journal of Economy, Culture and Society*, 69(1), 159-168.

Ozer, M., Perc, M., & Suna, H. E. (2024b). Participatory management can help AI ethics adhere to the social consensus. İstanbul University Journal of Sociology, 44(1), 221-238.

Perc, M. (2014). The Matthew effect in emprical data. *Journal of Royal Society Interface*, 11(98), 20140378.

Perc, M., Ozer, M., Hojnik, J. (2019). Social and juristic challenges of artificial intelligence. *Palgrave Communications*, 5, 61.

Rigney, D. (2010). The Matthew Effect: How Advantage Begets Further Advantage. *Columbia University Press*.

Tanberkan, H., Ozer, M., Gelbal, S. (2024). Impact of artificial intelligence on assessment and evaluation approaches in education. *International Journal of Educational Studies and Policy*, 5(2), 139-152.