Course unit: SCS 3108 - Data Structures and Algorithms

Module name/Topic: Ethical issues in data structures algorithms

Contributors: Evans Miriti, Christopher Chepken, Andrew Kahonge

Duration: 1hr

Reviewed by: Evans Miriti

Module Description

In this module, you will learn about ethical issues that you should be aware of and put into consideration as you design and select algorithms for use in computer systems. The ethical issues that you will learn about are: fairness, bias and discrimination; transparency; privacy; attribution of responsibility (accountability); and environmental impact. Being aware of these issues will enable you to intentionally make decisions that will ensure that use of computer systems to mediate decision making does not result in unethical outcomes.

Module learning outcomes:

By the end of this module, you should be able to:

1. State ethical issues that arise in algorithm design
2. Describe some unethical outcomes of algorithms
3. Discuss techniques of ensuring ethical outcomes in use of algorithms
4. Evaluate systems for compliance with ethical practices

Topic/Module content:

1. Introduction

Ethical Algorithm design is concerned with ensuring that the results of algorithmic computation do not exacerbate existing social problems. It is also concerned with ensuring that the use of algorithms results in the greater good for individuals, groups and the society. In addition, it is concerned with the respect of social values when algorithms are used to aid decision making or when algorithms make autonomous decisions.

In the following sections, you will learn about some of the ethical issues that emerge with the use of algorithms and how they can be resolved or mitigated.

2. Unfairness, bias and discrimination

According to Cambridge Dictionary fairness is “the quality of treating people equally or in a way that is right or reasonable”. Another definition by Oxford Languages is “impartial and just treatment or behavior without favoritism or discrimination”. Algorithms can produce outputs that result in unfair outcomes because they are developed by people who may already be biased e.g. against a certain group of people. It could also be because the data that algorithms learn from (in machine learning) possess these biases. Discrimination occurs where people from different groups are treated differently usually with one group receiving favorable treatment, while the other group(s) receive unfavorable treatment. This occurs in the course of decision making where decisions are made that tend to be favorable to one group and unfavorable to the other. With the growth in automation, we find that many of such decisions are algorithm mediated. The following examples illustrate cases where use of algorithms can or has led to unfair outcomes.

Recidivism Assessment Software

Angwin et al. (2016) reported on use of an application to perform risk assessment in relation to crime in some of the US states. The software was developed by a company known as Northpointe. The software was used to assess recidivism (reoffending) among defendants. The score was used by judges to make major decisions including if the offender was to be released, the bond amounts, the length of sentences and the defendant's rehabilitation needs.   The authors' evaluation of the software found that 1) The application wrongly labeled black defendants as future offenders at twice the rate at which it mislabeled whites 2) The software more frequently mislabeled white people as low risk while in fact they were high risk. The frequency of making the same mistake with blacks was lower.

Table 1: Performance of Risk assessment Software Angwin et al. (2016)

	White	**Black**
Classified high-risk yet did not reoffend	23.5%	44.9%
Classified low risk yet reoffended	44.7%	28%

The authors point out that the validity of these risk assessment tools are hardly ever assessed. Consequently, if there is any bias in the software's recommendations, it is likely to go unnoticed for a long time. The article quotes the former US attorney general Eric Holder who stated that " … risk assessment scores may be injecting bias in the courts". He further adds that "they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society."

This article shows the need for intentional assessment of algorithm based tools to determine if they exacerbate existing biases especially against any group of people. Such biases can be based on race, gender, age, disability and other factors.

STEM Career Advert Reach

A study by Lambrecht A. and Tucker C. E. (2019) found that a STEM career advert was more likely to be shown to men than to women. The advert was run on Facebook. The advert was meant to target both men and women equally. However, it was found that in an effort to minimize costs the advert was shown more to men. The authors stated that advertisers bid for the target audience. However, clicks by women are considered more valuable due their control of the household possessions budget. Consequently, higher paying advertisers crowd out lower paying advertisers in relation to the adverts that the women see. The consequence was that this particular advert was not shown to as many women as to men because it was cheaper to advertise to men than to women.

The policy in many organizations is to try to recruit more women in STEM careers. The article reported that with equal qualifications, a woman is more likely to be recruited than a man. However, the number of applications by women is low. This article shows how an algorithm can unintentionally, have a negative impact on policies that are meant to correct social imbalances.

Other Cases of Unfairness, Bias and Discrimination

A number of cases on algorithms generating results that lead to unfair outcomes have been recorded. The following are some examples:

i. Ads for background check services are more likely to appear after a search for a name associated with African Americans (Sweeney , 2013)
ii. Women less likely to see an advert for an executive coaching service (Datta et al., 2015)
iii. CV rating software rating CVs that contained the term woman lower e.g. Kiriri Women's university
iv. Setting of gender to female results in fewer instances of an advert related to a high paying job being shown compared to when the gender is set to male (Datta et al., 2015)

Other Areas where presence of Unfairness, Bias and Discrimination should be evaluated

i. Loan creditworthiness apps
ii. Tender evaluation software
iii. Learning institution placement software
iv. Any other examples?

3. Transparency

Lack of transparency in an algorithm can be as a result of several factors such as:

i. Being an inherent characteristic of the model e.g. deep learning models,
ii. Deliberate anonymization of data,
iii. Voluminous amounts of code that are difficult for humans to comprehend,
iv. Large amounts of data that are hard for humans to visualize,
v. Poorly structured and documented code,
vi. Ongoing updates to the system which makes make it difficult to keep track of the current status of the model,

vii. Claim by solution providers of the necessity of protecting their intellectual property.

Lack of transparency can lead to lack of trust of the algorithm based system. It also means that auditability of the system by independent parties is reduced.

However, it has been argued that too much transparency and too much focus on transparency can lead to unfair outcomes. This is because when too much information on the workings of algorithms is provided, users may feel overwhelmed by the information and develop feelings of inadequacy. It may also lead to a situation where some users try to take advantage of the system (also called gaming the system) which may also introduce a form of social inequality. Focus on transparency may also lead to a situation where resources are diverted from important considerations such as safety and accuracy.

A number of ways have been proposed for improving transparency or explainability of algorithms. One of the proposed methods is creation of appropriate documentation when the system is being created, and documenting any updates that are made on the system. Documentation for any new module or component should include the changes that have been made, the functions of the component, test results and how the component should be used. Another approach to improving transparency is an audit test that determines if the algorithms are exhibiting any negative tendencies. These can then be discovered early and remedial measures taken. Transparency factors have also been proposed that solution providers can endeavor to document including: accuracy, data timeliness (is it recent enough), completeness of the data (does it have some missing values or groups), sampling methods used, sources of the data and volume of the data.

In order to improve explainability, companies are creating tools such as Explainable AI by Google that offer users the ability to visualize and understand model predictions. According to Google, Explainable AI helps to build “interpretable and inclusive AI systems from the ground up” and “detect and resolve bias, drift, and other gaps in data and models”. AI Explainability 360 is another tool by the Linux Foundation that serves a similar purpose.

4. Privacy

We will discuss privacy issues in the context of recommender systems although they can occur in other application domains. Privacy issues that can occur in recommender systems include:

i) Collection and sharing of user data without the user's explicit consent,
ii) Stored user data leaking to external agents,
iii) Stored or shared user data being subjected to deanonymization attacks,
iv) Models inferring additional sensitive user data based on user actions or the groups in which the user has been placed.

Breach of privacy puts the user at risk for instance, the risk of being a target of malicious agents e.g. online crime agents. There are risks of personal embarrassment and injury to reputation. Systems may also discriminate against a user based on inferred private information e.g. the credit worthiness of the user or their political inclinations. Inferred information may be beneficial to the user in some domains (for instance in medical practice) but may be detrimental in other areas (for instance in job recruitment).

Methods for solving privacy issues include opt in opt out features for users. However, it has been suggested that users' selections of opt in/out features may also be used to draw conclusions about some personal characteristics of the user. Other methods of dealing with privacy issues include encryption and differential privacy. Other methods include legislation that limit and regulate the collection, use and sharing of individuals' private data. In Kenya, the legislation that governs this is Data Protection Act (2019).

5. Accountability

Accountability is related to taking responsibility for decisions made by systems using algorithms or with the assistance of such systems. Rubel et al. (2019) use the term agency laundering for situations where an agency's decisions are made in an automated fashion by a system. The agency then refuses to take responsibility for bad outcomes (if any) of the system's decision and lays the blame on the algorithm/system.

One example cited in Rubel et al. (2019) is of a report on how Facebook ad targeting mechanism was used by a user to create an ad targeting category called Jew haters. The mechanism then suggested other anti-Semitic phrases that would further improve the ad. Facebook avoided responsibility by stating that they had not expected the system to be used in this way. Another example cited in the same article is that of the Uber ride hailing app predicting higher prices for drivers. This prompts the drivers to make themselves available to offer the riding services. But sometimes, the predicted higher prices fail to materialize.

There are other situations where morally responsible agents may fail to take responsibility and blame it on algorithms. An example is the medical sector where doctors fail to question suggestions of clinical decision support systems and defend themselves by saying the system said so.

Some suggestions on how to avoid shirking of responsibility include development of frameworks for attribution of responsibility for algorithms decisions. Another suggestion is to attribute responsibility to all the actors in the decision making process with the intention of focusing on how mistakes can be corrected.

6. Autonomy

Autonomy is related to humans self-determination and ability to make their own choices and possibly reject those offered by others including algorithmic systems. When we use algorithmic systems, we delegate some of our decision making rights to the algorithmic systems. There is a need to strike a balance between the self-determination powers delegated to algorithmic systems and those retained by users. An example of a proper balance in autonomy is letting a human driver cede control to a car driving agent and be able to take back control when they desire to. According to Tsamados et al. (2021) Lack of user autonomy can be attributed to several factors including:

a) Prevalence of algorithm based systems,
b) Users inability to comprehend how algorithms work,
c) Lack of avenues to appeal algorithmic decisions.

Users' autonomy may also be lost in that they may have no ability to control what information is stored about them. Users may be forced to provide information to systems that they would rather have kept private. This can include gender especially in the age of fluid gender identities. Other user categories can be inferred by algorithm’s internal logic. Consequently, systems may place users in categories that they are uncomfortable with.

Lack of autonomy may also be as a result of algorithms based systems persistence in offering or nudging a user to make certain choices.

Participatory design has been advanced as one of the methods that can be used to put into consideration the values of end-users and increase their autonomy. The goal of this technique is to ensure users knowhow and experiences are factored into the algorithms’ designs. Society-in-the-loop framework is a conceptual framework that seeks to include all stakeholders in the design process so that the algorithms are a means to implement or mediate a social contract that has been agreed by the stakeholders involved. In the context of recommender systems, participatory design can ensure that the categories that are created by the system are more accurate and that no categories that are of interest to users have been excluded.

7. Environmental Impact

One of the areas of concern in regards to environmental impact of algorithms is the power consumption by powerful computers executing computing intensive workloads. Very high power consumption leads to increases in utilization of fossil fuels to generate power for powering the computers and for cooling purposes. Examples of such computing intensive uses are blockchain and some artificial intelligence (AI) applications. Rashid, Ardito and Torchiano(2015) showed that the running time of an algorithm was the main determinant of how much power was consumed.

Some blockchain applications, for instance Bitcoin, rely on a verification mechanism known as Proof of Work (PoW). PoW requires miners to compete in solving a complex cryptographic challenge in order to win the right to add the next block into the chain and in so doing, be rewarded with Bitcoins. The challenge to be solved keeps becoming harder and harder, meaning that miners have to keep increasing the power of their computers in order to stand a chance of winning the challenge. In 2021, it was estimated that the Bitcoin blockchain consumes approximately 500TWh of power. Kenya generates about 12TWh of power per annum. It has also been claimed that blockchain mining contributes to emission of 65 megatons of CO_2 per annum.

In AI, modern learning models such as deep learning models require powerful computing resources for training purposes. The problem is exacerbated by the fact that training models such as large language models require huge amounts of data. Consequently, a lot of power will be required both for the storage and training purposes.

In blockchain applications, new validation mechanisms that are less power intensive are being introduced in the industry. As an example, the Ethereum and Algorand blockchain technology use the less power intensive proof of stake consensus mechanism. Other methods to reduce power consumption include shifting workloads to cooler places that will require less use of air conditioning for cooling. Other methods to deal with the issue include use of green energy sources and carbon credit trading.

AI can be used to offset its carbon footprint by being applied in fields such as precision agriculture. Using AI to control the systems can result in less power, fertilizers and electricity being consumed.

Delivery Methods

{ Face to face lecture /Online lecture/Pre Recorded lecture/(Lecture notes)

Supplementary reading materials

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The ethics of algorithms: Key problems and solutions. *AI & Society*.

Cases:

Club X plans to implement a notification system that sends members their invoices and payment details via SMS. The SMS message will also contain the members unpaid balance (or overpayment) and their credit limit. Suggest some ethical concerns that the club should consider in developing the systems

Privacy – The messages will be sent using service providers. Club X should ensure that service providers respect privacy. The club must also respect the privacy of members in terms of mobile phone details and balances.

Autonomy – Some members may opt not to receive the message, or receive the messages only at certain times, or receive the messages in other formats

Transparency – How much it costs to provide the service. Are members charged extra for the SMS?

Kenya Bank is planning to offer online access to its services. For a client to access the online services portal, they will be required to login using a username and a password. In addition, the client will be required to enter a one time password (OTP) that is sent as an SMS to their phone.

The OTP will expire if it’s not submitted by the client within one minute. Discuss the ethical issues that can arise in the provision of this service and how they can be addressed.

- Stakeholders – Clients, bank, SMS service providers,
- Clients with some disabilities or elderly may be unable to enter the OTP on time
- SMS service provide will have access to telephone details of the clients

  Some solutions
- Alternatives – mobile app, email to send the OTP
- Client to set the time limit for expiry of heir OTP (autonomy)
- Agreements with SMS service providers on data protection to ensure respect of privacy

Assignments /Quizzes:

i. Discuss how the use of algorithms can lead to injustice
ii. List some techniques that can be used to make algorithm based systems more transparent
iii. Discuss why it is important to protect people's privacy
iv. Discuss some methods that can be used to minimize the environmental impact of algorithms
v. Participatory design can help address which ethical issues?

vi. Which of the following can be a source of Bias in algorithms

a. Programmers (Answer)

b. Data (Answer)

c. Users (Answer)

d. Computer hardware

vii. When system users refuse to take responsibility for decisions of their systems, this is an example of

a. Lack of transparency

b. Lack of autonomy

c. Lack of accountability (Answer)

viii. In Kenya, the law that is concerned with the protection of data is called?

Data Protection Act (2019)

ix. Which blockchain consensus mechanism is more power intensive

a. PoS

b. PoW

x. Use of Northpointe software to predict recidivism was seen seen to result in

a. Gender discrimination

b. Race discrimination (Answer)

c. Discrimination based on age

d. Discrimination based on nationality

References:

1. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
2. Columbia Climate School. (2022). Cryptocurrency's Dirty Secret: Energy Consumption. https://news.climate.columbia.edu/2022/05/04/cryptocurrency-energy/
3. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Sci Rev*. https://doi.org/10.1162/99608f92.8cd550d1
4. Lambrecht, A., & Tucker, C. E. (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads.
5. Rashid, M., Ardito, M., & Marco, T. (2015). Energy Consumption Analysis of Algorithms Implementations. *Symposium on Empirical Software Engineering and Measurement*, Beijing, China.
6. Rubel, A., Castro, C., & Pham, A. (2019). Agency Laundering and Information Technologies. *Ethic Theory Moral Prac, 22*, 1017–1041. https://doi.org/10.1007/s10677-019-10030-w