Algorithm Bia: Detection and Perspectives of Undergraduate Students

Wei Zakharov

Senay Purzer

Carla Zoltowski

PURDUE UNIVERSITY

ALGORITHM BIAS: DETECTION AND PERSPECTIVES OF UNDERGRADUATE STUDENTS

- SLA MAY 19, 2023

Wei Zakharov, Libraries and School of Information Studies

Senay Purzer, Engineering Education

Carla Zoltowski, Electrical and Computer Engineering

Wei Zakharov, Ph.D
Associate Professor
Engineering Information Specialist
wzakharov@purdue.edu
Liaison areas: AAE, AT, ECE, ENE
49-42872

PURDUE UNIVERSITY | Libraries and School of Information Studies

AGENDA

1. Study Context
2. Algorithm Bias
3. Ethical Reasoning Education

COURSE SETTING

- Information and Library Science (ILS) 103, 1 credit

"Introduction to Data Management"

ILS 103: INTRO TO DATA MANAGEMENT

Foundations of the Data Mind: An Interlocking Modules Approach
Funded by Purdue Integrative Data Science Initiative Education Grant

https://www.purdue.edu/data-science/education/education-proposals.php

Three one-credit courses were created to provide an introduction in three key areas of data literacy:

(i) data organization and management

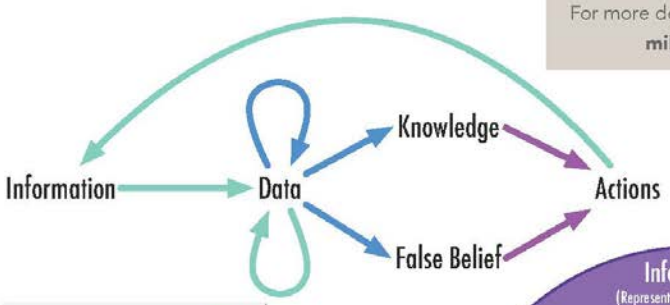
(ii) ethics of data collection and use, and

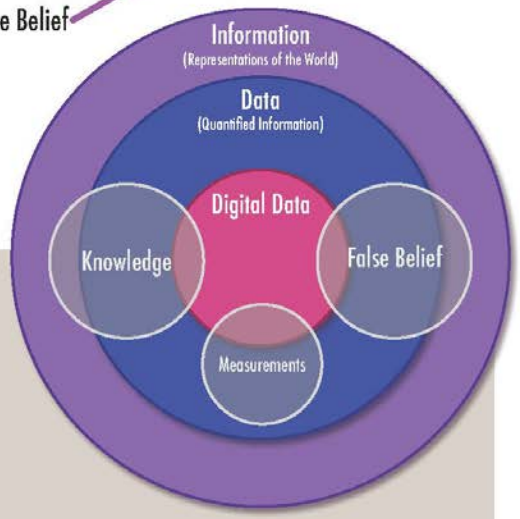
(iii) methods of data analysis

COURSE FLYER

Foundations of Data Science
ECE 295-PHIL 293-ILS 295

For more details, contact Prof. Kulkarni
milind@purdue.edu





- What is data?
- How do we collect and manage it?
- How should it be analyzed?
- How should it be used?

Cutting edge jobs and research are *data centric*: everybody will need to be able to collect, analyze, manage, and use data. But what does that mean? How do you know what questions to ask? How do you present that data to others? What are the right and wrong ways to navigate the world of data?

Spring 2019

- **Mon** ILS 295, Intro to Data Management
- **Wed** PHIL 293, Ethics of Data Science
- **Fri** ECE 295, Intro to Data Science

3:30-4:20

Take all three for the full picture!

ILS 103: INTRO TO DATA MANAGEMENT

Learning Objectives

- Students will be able to understand a data life cycle and identify different types of data.
- Students will be able to identify a variety of data collection techniques and determine the appropriate way to collect data for their questions.
- Students will be able to locate existing data sets for their questions.
- Students will be able to organize data on local storage and in Purdue's PURR.
- Students will apply best practices into data visualizations.
- Students will gain tools to preserve and publish data.

- A minimum of 16 credit hours

- Core Courses: a minimum of 10 credit hours
  - 3 credit hours in Statistical Methods
  - 3 credit hours in Computing
  - 3 credit hours in Data Literacy, Management & Analytics
  - A minimum of 1 credit hour in Data Ethics & Digital Citizenship

- Application Focus: Six (6) credit hours in courses concentrated on applications of data science related knowledge

- A minimum of six (6) credit hours must be in coursework outside the student's program.

DATA ETHICS EDUCATION
-ALGORITHM BIAS

DATA ETHICS

Data ethics is about responsible and sustainable use of *data*. It is about doing the right thing for people and society.

ACTIVITY

How would you program machine intelligence for driverless cars?

https://www.moralmachine.net/

ETHICS VS LAW

1. Ethics are **rules of conduct**. Â Laws are rules developed by governments in order to provide balance in society and protection to its citizens.
2. Ethics comes from people's awareness of what is right and wrong. Â Laws are **enforced** by governments to its people.
3. Ethics are **moral codes** which every person must conform to. Â Laws are codifications of ethics meant to regulate society.
4. Ethics does not carry any punishment to anyone who violates it. Â The law will punish anyone who happens to violate it.
5. Ethics comes from within a person's moral values. Â Laws are made with ethics as a guiding principle.

Read more: Difference Between Law and Ethics | Difference Between http://www.differencebetween.net/miscellaneous/politics/difference-between-law-and-ethics/#ixzz6cer1x9ZE

DATA OWNERSHIP

"In legal theory, ownership is determined by who is entitled to sell a particular item of property."*

"The rights associated with ownership provide a powerful basis for control. A data owner can provide access to data or can restrict access partially or entirely. Data owners can impose conditions on access or use, including charging fees. Yet, if rights of ownership are recognized, the power to control information resources must have limits as well…any data ownership rights must similarly include exceptions that are appropriate for the public interest."**

*Moore GW, Berman JJ. Anatomic pathology data mining. In: Cios KJ, editor. Medical data mining and knowledge discovery. Heidelberg: Springer, 2000. p. 61–108 [chapter 4]. As cited by Cios, Moore. 2002. Uniqueness of medical data mining. Artificial Intelligence in Medicine 26(1-2). pp. 1-24

**Scassa T. 2018. Data ownership. CIGI Papers No. 187. https://www.cigionline.org/publications/data-ownership

INTELLECTUAL PROPERTY AND COPYRIGHT

- In general raw data on their own are considered facts and thus can not be copyrighted

- Datasets can be copyrighted or licensed

- Using a dataset

When re-using existing data, be sure to clarify ownership, obtain permissions if needed, and understand limits set by licenses.

Be sure to provide appropriate attribution and citation. If licensing restricts sharing of the data, providing detailed information about where the data were obtained and how the data were analyzed can help with reproducibility.

DATA CITATION

Citing research data in a manner similar to traditional scholarly works:

- Full citations should be included in the reference list
- Give permanent identifiers, such as DOIs, in the form of a linked URL if possible.
- Cite data sets at the most detailed level possible and provide a version if appropriate.

https://libguides.lib.msu.edu/citedata

https://data.research.cornell.edu/content/data-citation#examples-data-citation-styles

DATA PRIVACY

User privacy is the right to open inquiry without having the subject of one's interest examined or scrutinized by others.

Data Classification

Confidential Information

The term “Confidential” is often used interchangeably with other security terminology. It is **not** a data classification like sensitive or restricted. It describes how information should be treated.

For example, a conversation between an individual and their supervisor may be confidential if the employee requests the supervisor not share that information with anyone else.

DATA ETHICS EDUCATION SUMMARY

Data ethics issues

- Ownership
- Intellectual property and copyright
- Data citation
- Privacy
- Confidentiality

Newly added topic:
Algorithm bias & related assignment

ALGORITHM BIAS

Algorithm Bias

occurs when computer algorithms that systematically discriminate against certain content, individuals or groups in favor of others without a sound basis.

Bias= preference, commonly assume to be an unfair preference.

CLARIFICATION

- Algorithm bias does not apply to computer programs that have known correct solutions. If we as humans know how to always get a correct solution, a computer program should just do that.

- Algorithm bias applies to problems where humans themselves do not have methods to always be correct.

DISCUSSION QUESTIONS

- Which laws and regulations might be applicable to your project?
- Which types of data need to be handled with care?
- How are we achieving ethical accountability?
- How might the legal rights of an individual be impinged by our use of data?
- How might individuals' privacy and anonymity be impinged via our aggregation and linking of data?

DISCUSSION QUESTIONS - CONT

- How do we know that the data is ethically available for its intended use?
- How have we identified and minimized any bias in the data or in the model?
- How was any potential modeler bias identified, and if appropriate, mitigated?
- How transparent does the model need to be and how is that transparency achieved?
- What are likely misinterpretations of the results and what can be done to prevent those misinterpretations?

ALGORITHM BIAS ASSIGNMENT

Data ethics is about the responsible and sustainable use of *data*. It is about doing the right thing for people and society. Identify one algorithm bias example in your discipline or interest area. The source needs to be a news article or a journal article. Please answer the below questions.

- What's your discipline or interest area?
- What are your perspectives on fairness and ethics in regard to algorithm bias?
- What's the algorithm bias example you located? Describe the problem in your own words.
- What are your thoughts about the identified example guided by the data ethics session?
- Source citation (APA format) https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/general_format.html

ALGORITHM BIAS DETECTION

- Steve Wozniak says Apple Card discriminated against his wife. – CNN Business, 2019
- Credit model (Women's World Banking, 2022)
- Google interfere with its search algorithms and changes your results. – WSJ, 2019
  Google Translate (Edmondson, 2022)
- Reading apps algorithm feedback to heighten human bias
- Facial recognition used in law enforcement – potential source of both race and gender bias (Wilson, et al., 2019)

ALGORITHM BIAS DETECTION- CONT

- Dissecting racial bias in an algorithm used to manage the health of populations (Obermeyer, et al., 2019; Grant, 2022; & McKeon, 2022)
- Hiring practices (Dastin, 2018 & Lawton, 2022)
- Autonomous vehicles (Booth, et al., 2021)
- Twitter- young, thin females; rightwing politicians and news outlets (Knight, 2021 & Milmo, 2021)
- TikTok's algorithm shows unintentional racial bias. – BuzzFeedNews, 2020

ALGORITHM BIAS DETECTION- CONT

- Facial recognition used in law enforcement – potential source of both race and gender bias (Wilson, et al., 2019)
- Equitable loan decisions (Daniels, 2022)
- Financial services (Klein, 2022)
- OpenAI's DALL·E image generator(Traylor, 2022)
- Algorithmic bias in education(Baker & Hawn, 2022)
- …

CHALLENGE

Ethical Reasoning Development

ETHICAL REASONING PRINCIPLES

The reflexive principlism framework lists four core moral principles as follows:

1. Preventing harms and providing benefits (beneficence)
2. Avoiding the causation of harm (non-maleficence)
3. Supporting and respecting autonomous decisions (respect for autonomy)
4. Fairly distributing benefits, risks, and costs (justice)

- Beever & Brightman, 2016

ETHICAL REASONING FRAMEWORK

Centering on the negotiation of risk and benefits

QUESTIONS?

Wei Zakharov
wzakharov@purdue.edu

PURDUE UNIVERSITY | Libraries and School of Information Studies

REFERENCES

- Baker, R. S., & Hawn, A. (2022). Algorithmic Bias in Education. International Journal of ArtificialIntelligence in Education, 32(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9
- Beever, J., & Brightman, A. (2016). Reflexive Principlism as an effective approach for developing ethical reasoning in engineering. Science and Engineering Ethics, 22(1), 275-291. doi: 10.1007/s11948-015-9633-5
- Booth, B. M., Hickman, L., Subburaj, S. K., Tay, L., Woo, S. E., & D'Mello, S. K. (2021). Bias and Fairness in Multimodal Machine Learning: A Case Study of Automated Video Interviews. 268–277. https://doi.org/10.1145/3462244.3479897
- Daniels, C. (2022). The financial sector is adopting AI to reduce bias andmake smarter, more equitable loan decisions. but the sector needs to be aware of thepitfalls for it to work. Business Insider. Retrieved from https://www.businessinsider.com/how-the-finance-sector-uses-ai-to-make-loan-decisions-2022-8
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Retrieved from https://www.reuters.com/article/usamazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-thatshowed-bias-against-women-idUSKCN1MK08G
- Edmondson, A. (2022). Study Finds Google Translate Is Sexist. The Reykjavik Grapevine. Retrieved November 15, 2022, from https://grapevine.is/news/2022/06/08/study-finds-google-translate-is-sexist

REFERENCES

- Klein, A. (2022). Reducing bias in AI-based financial services. Brookings. Retrieved November 15, 2022, from https://www.brookings.edu/research/reducing-bias-in-ai-based-financial-services
- Knight, W. (2021). Twitter's photo-cropping algorithm favors young, thin females. Wired. Retrieved from https://www.wired.com/story/twitters-photo-cropping-algorithm-favors-young-thin-females
- Lawton, G. (2022, August 29). Ai hiring bias: Everything you need to know. SearchHRSoftware. Retrieved November 14, 2022, from https://www.techtarget.com/searchhrsoftware/tip/AI-hiring-bias-Everything-you-need-to-know
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342
- Stahl, C. (2022). No quick fix: How OpenAI's Dall·E 2 illustrated the challenges of bias inAi. NBC News. Retrieved from https://www.nbcnews.com/tech/tech-news/no-quick-fix-openais-dalle-2-illustrated-challenges-bias-ai-rcna39918
- Women's World Banking. (2022). In a world of gender bias, Lendingkart's AI-based credit model stands apart. Women's World Banking. Retrieved from https://www.womensworldbanking.org/insights-and-impact/in-a-world-of-gender-bias-lendingkarts-ai-based-credit-model-stands-apart