Analyzing the Adverse Effects that Recommender Systems Have on Humans

STS 4500 Prospectus

School of Engineering and Applied Science

B.S Computer Science

The University of Virginia, Charlottesville

Name: Pawan Jayakumar

Technical Advisor: N/A

STS Advisor: Alice Fox

Projected Graduation Date: December 2023

Submission Date: May 2023

Analyzing the Adverse Effects that Recommender Systems Have on Humans

Overview:

The rise of the internet over the 21st century has led to enormous amounts of data being generated by users and their interactions with other users and websites. Researchers and corporations have been trying to harvest this data to create useful or profitable models such as recommender systems. As the name suggests, these models attempt to learn the user's preferences on various things such as products to purchase on Amazon or videos to watch next on YouTube. By analyzing literature and studies performed on people who interact with these recommender systems, this paper argues that they do more harm than good.

Positionality:

As a computer science student at the University of Virginia, I've learned about the underlying math and computer science that powers models such as Alpha-Fold and Chat-GPT. I also have firsthand experience in research related to the field of machine learning. This experience let me understand at a deeper level where the state of the industry is right now, and where it is likely to head in the next few years. Having seen the rapid improvement in capabilities of state-of-the-art hardware and software, ethics is often not prioritized in the development of new applications using these technologies. Even though I have the privilege of being able to observe these trends due to my schooling experience, I still think it's difficult to understand how these models work, so it would be unfair to expect most people to. That's why extra care must be taken to consider the impact these large models may have on society. As a person of color and the child of two immigrants, I hope that researchers ensure that all social groups are considered when decisions are made, especially those that are historically under-represented. The rate of improvement is only increasing, so more deliberate effort must be applied to ensure all stakeholder interests are being considered.

Problematization:

This paper addresses the adverse effects of recommender systems on users and even society as a by-product. The main system that will be discussed is Twitter's "tweet" timeline recommender system. Twitter is a monetized social media app where users are able to create "tweets" which are accessible to other users. Since millions of tweets are created every minute, the Twitter recommendation system selects which of these are most relevant to the user and displays it in the users' home page, also known as the timeline. This system was chosen because the owner of Twitter has open-sourced the code for it, making it easier to review.

Guiding Question or Main Argument:

This paper argues that Twitter's tweet recommender has harmful effects on its users such as providing bad recommendations as a result of conflict of interests, spreading misinformation, and a lack of privacy/transparency with how user data is collected and stored.

Projected Outcomes:

This paper aims to summarize knowledge about the flaws of Twitter's tweet recommender system from various technical papers and studies published. The findings will educate the reader that knowledge can potentially mitigate any harmful effects these systems may have. Twitter is one of the most used social media websites/apps so this paper can help many people.

Technical Project Description:

Recommender systems (RS) learn to predict what a user is interested in buying or watching using reinforcement learning. By collecting data about the user such as purchase/browsing history as well as feedback from other recommendations, RS will learn user preferences over time. RS also leverage similarities among users and products, which is known as collaborative filtering. For example, if both user A and B love watching cat videos and user B watches a cat documentary, then user A will then be recommended that documentary.

RS use reinforcement learning models to assign a score to each product which represents how likely a user is to interact with it. The more data and feedback the model has from the user, the more accurate this score becomes. The only issue with this approach is the availability of reliable data.

User data is very noisy because the user isn't explicitly being told that they are generating data to train a model. Instead, websites often infer the feelings/preferences through their actions. These data labels, "positive" or "negative" are not the ground truth. Someone could watch five minutes of a video and then quit because they didn't like it. From the models perspective, it is hard to tell whether they liked or didn't like this video since they interacted with it.

Working under professor Hongning Wang, I have helped in formulating techniques to train models with noisy labels. I implemented a special layer in a neural network which adapts to the user who generates the noisy labels and adjusts how the model updates its internal parameters accordingly. Using this approach, the model was around 14% better than without it.

Preliminary Literature Review & Findings:

There are three main findings from the research that I conducted: the first is that content/media RS such as twitter are designed to keep the user "hooked" onto the app or website, rather than actually recommend content the user would benefit from (Seaver, 2019). This results from a conflict of interest between the user wanting personal recommendations and the company wanting the user to keep scrolling to generate more money. With constant bombardment of new tweets that may desirable for us to consume, our brains gravitate towards scrolling on Twitter forever in search of those tweets instead of working on our responsibilities (Evitts, 2022).

The second finding is that RS are biased because they don't consider that by giving recommendations, they are actively influencing the user's preference (Evans, 2022). They also are not particularly good at recommending things when they don't have much data (Zhu, 2021), so it's easy to reach a scenario where the user is being recommended things they never actually needed.

The last finding is related to privacy and transparency. Since RS rely on data, websites/apps gather user data, sometimes without the user's consent. There is a tradeoff between privacy and effectiveness because the more personally identifiable information the RS has, the better recommendations it can make (Friedman, 2015). However, in case of a data breach, users are at the mercy of the perpetrators.

Researchers are trying to improve RS from a technical standpoint which helps with the bias, but there is still an ethical component which only companies can change.

STS Project Proposal:

Science, Technology, and Society or STS is a multi-disciplinary area of study which focuses on how scientific and technological advances impact individuals and the overarching society said advances takes place in. Society's natural stratification means that individuals in different groups are impacted differently by the same technology. Designers need to carefully consider all types of people when creating their product or service. My paper considers ethical and technical considerations that engineers of RS have to make and how these affect the end user and society as a whole.

I approach this STS piece from a ethics and values standpoint. Specifically, my paper investigates the privacy, trustworthiness, and fairness of RS. One of the primary authors I will be referencing is Saad Tariq who is a scholar at the KTH Royal Institute of Technology in Sweden. His work examines the hidden side effects of recommender systems which is in line with my topic. Since his arguments line up with mine, I have found several useful resources from his citations and findings that I can explore further.

I also will be referencing several technical authors including Alex Beutal, Charles Evans, and Miriam Fernandez. These authors have written papers discussing current problems with RS and designed solutions towards these problems while noting their limitations. This is important for my work because it's the most rigorous way to show that current RS are indeed flawed as well as how we can potentially alleviate these problems with more research in that area.

Value sensitive design is a methodology to designing technology that considers the values and ethics of the users and stakeholders who will be affected by the technology. In particular, I will be using Friedman's 2006 work on stakeholder analysis and model for informed consent online. RS have three different stakeholders: end users, advertisers, and the companies using them. The end users want things they would be interested in the most, while companies want them to keep using their product to increase profits which they get from advertisers. This work will focus on the former two.

When it comes to recommender systems, stakeholder analysis and informed consent online can be used to address ethical considerations related to privacy, transparency, and autonomy. Informed consent online can be used to ensure that users have control over their personal data and that their privacy is protected. Stakeholder analysis can be used to design systems that are transparent and explainable, providing users with clear information about how recommendations are generated and allowing users to understand and question the decisions made by the system.

My research method will be a conceptual and technical investigation based on literature review and personal interviews. Literature review can help find gaps in existing research and potential areas for further investigation, as well as providing a solid foundation for the topic. In particular, Twitter released a white paper detailing their recommender system design. Interviews with end users and corporate representatives would provide valuable insight into the values of the stakeholders for this topic. This investigation would support my value sensitive design approach to discussing RS and can help me analyze the impacts it has on society.

Barriers & Boons

Although I have some tangential experience working with RS, I don't have rigorous experience designing one that is actually used in production. My understanding is limited, and I have to rely on reading literature of state-of-the-art RS and this prospectus is my interpretation of those papers. I have a time and money limitation here because I have other responsibilities and debts to pay off which I could spend to better understand RS. Thankfully, many technical papers are available thanks to UVA and ArXiv and Twitter's owner decided to release the RS white paper.

References

Afsar, M. Mehdi, et al. “Reinforcement Learning Based Recommender Systems: A Survey.” *ArXiv.org*, ArXiv, 8 June 2022, https://arxiv.org/abs/2101.06286.

This is a technical paper is a survey discussing various recommender system that use reinforcement learning (RL) to train. RL is a class of machine learning which specializes in problems that involve sequential decision making. The state of the art approach to recommender systems is to model it as a problem where you are sequentially making decisions about what to recommend to the user and receiving reward in the form of user feedback through clicks or ratings. These rewards are then used to adjust the recommender system’s internal parameters to maximize this reward. There are many ways to model both feedback, the internal parameters, and problem formulation as a vector of numbers and this paper goes through various existing methods. This paper is important because it provides a theoretical basis of how exactly recommender systems work which is necessary to understand in order to critique its use.

Beutel, Alex, et al. “Fairness in Recommendation Ranking through Pairwise Comparisons.” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, https://doi.org/10.1145/3292500.3330745.

This paper offers another technical framework of fairness specifically in the “warm start”: when there is previous user data. They then develop an algorithm to correct models to make them more fair. This paper is important because it is important to be aware of different fairness frameworks and their limitations in order to find potential areas to improve upon.

Digital, GLP. "Another Modern Myth: Shrinking Attention Spans." *Genetic Literacy Project*, 12 Jan. 2018, https://geneticliteracyproject.org/2016/07/25/another-modern-myth-shrinking-attention-spans/.

This article provides a counter-argument to the idea that attention spans are shrinking. It critiques the Microsoft study which found that our attention span is decreasing by saying that the study wasn't peer reviewed, their concept of attention wasn't capturing the nuances of the topic correctly, and there are some studies which show different results like how video game experts are actually better at tracking objects on a screen and switching tasks. The article then suggests the true problem may lie in multi-tasking. It is important to look at both sides of the argument when trying to conduct research so this article is important.

Evans, Charles, and Atoosa Kasirzadeh. "User Tampering in Reinforcement Learning Recommender Systems." *ArXiv.org*, ArXiv, 2 Nov. 2022, https://arxiv.org/abs/2109.04083.

This paper looks at one way a reinforcement learning based recommender system can fail to provide truthful recommendations to the user. The paper shows that these recommender systems actively manipulate the user's interests by initially recommending things that make it easier to recommend things in the future because they are similar and abundant. I actually observed this phenomenon myself in my Discover Weekly in Spotify. When I first started using it, it gave me some lo-fi or chill music to listen to and I hadn't really listened to it before but I gave it a try and then it kept recommending similar type of music in the ensuing weeks. I don't just want to listen to lo-fi but I don't know how to 'fix' it now. This

paper is important because it suggests that there is a fundamental problem with current recommender systems and researchers will need to figure out a different model to fix it.

Evitts, Jared. “Tiktok-Addicted Students Delete App during Exams.” *BBC News*, BBC, 4 Sept. 2022, https://www.bbc.com/news/uk-wales-62720657.

This news article talks about how students were forced to delete TikTok in order to concentrate on studying for their exams. It then goes on to quote a mental health expert and explain how TikTok is so addicting. It is an endless stream of relatively desirable content thanks to its recommender system and since the videos are short and different, the brain is continually releasing dopamine. This shows how recommender systems can lead to bad outcomes for users.

Fernandez, Miriam, and Alejandro Bellogín. “Recommender Systems and Misinformation.” *Ceur*, University of Madrid, 25 Sept. 2020, http://ceur-ws.org/Vol-2758/OHARS-paper3.pdf.

This article provides an overview of how recommender systems are at the root of when misinformation is spread rapidly. This is another problem with recommender systems that use faulty methods of gauging how ‘good’ content is. Misinformation usually deals with emotionally charged topics, which has been shown to be shared and engaged with more. These increased engagements mean it will get recommended more as well. This article is important because the spread of misinformation is hard to correct and another harmful effect of recommender systems.

Friedman A., Knijnenburg B.P., Vanhecke K., Martens L., Berkovsky S. (2015) Privacy Aspects of Recommender Systems. In: Ricci F., Rokach L., Shapira B. (eds) Recommender Systems Handbook. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7637-6_19

This paper analyzes user privacy risk and explores various methods for balancing the user privacy/accuracy tradeoff when it comes to designing recommender systems. It concludes that there is still more work to be done in order to create systems that can balance both of these needs. This paper was written by a UK research group and it supports my argument that recommender systems have the potential to cause privacy related issues

Harris, Jeremie. "Ethical Problems with Recommender Systems." *Towards Data Science*, Medium, 27 Jan. 2021, https://towardsdatascience.com/ethical-problems-with-recommender-systeems-398198b5a4d2.

This podcast summary written by an AI safety expert talks about the ethical concerns she has seen concerning recommender systems. He first identifies 4 different stakeholders with competing interests: users, the system itself, the companies, and society at large. He then describes how these competing interests lead to several problems. For example, the recommender system for twitter or TikTok is designed to monopolize user attention for as long as possible by constantly feeding them new information. This blog serves as prior work for my topic and is a first step to finding more information about the subject and areas to explore further.

Jan Kleinnijenhuis, Anita M J van Hoof, Wouter van Atteveldt, The Combined Effects of Mass Media and Social Media on Political Perceptions and Preferences, Journal of

Communication, Volume 69, Issue 6, December 2019, Pages 650–673, https://doi.org/10.1093/joc/jqz038

This paper explores the effects of social media on political views. I chose to explore this because social media is increasingly becoming the most prevalent way people hear about news and what others think about news as well. Social media websites like Twitter use recommender systems to control what you see on your timeline which is pretty scary because if it keeps showing tweets with a certain type of content, it can lock a user into an echo chamber, and this leads to more polarization and even extreme views such as people who believe the Earth is flat. This paper is useful in my thesis because it shows how recommender systems are capable of affecting more than just the user, but society itself.

Kanokporn Sriwilai, Peerayuth Charoensukmongkol. "Face It, Don't Facebook It: Impacts of Social Media Addiction on Mindfulness, Coping Strategies and the Consequence on Emotional Exhaustion." *Stress and Health : Journal of the International Society for the Investigation of Stress*, U.S. National Library of Medicine, 30 Mar. 2015, https://pubmed.ncbi.nlm.nih.gov/25825273/.

This article explains the impact of social media on mindfulness which is a combination of one's mental health, discipline, and wellbeing. It conducted a survey and found that addiction to social media is prevalent among those who have low mindfulness and use emotion-focused coping techniques to deal with stress. This article is important because recommender systems may be worsening the negative effects of social media when it comes to mindfulness.

Koene A. et al. (2015) Ethics of Personalized Information Filtering. In: Tiropanis T., Vakali A., Sartori L., Burnap P. (eds) Internet Science. INSCI 2015. Lecture Notes in Computer Science, vol 9089. Springer, Cham. https://doi.org/10.1007/978-3-319-18609-2_10

This paper describes the privacy concerns that recommender systems bring up. These algorithms work on user data, and there is a tradeoff where better recommendations require more personal user data. The need to collect and store this data leads to instances where users are unaware that their data is being collected, stored, and potentially even shared. There are also concerns that said data can be accessed by malicious parties through data breaches. The paper is written by academics in the field of computer science who are concerned about the risks of recommender systems and surveyed current solutions to find solutions to strike a healthy balance between meaningful recommendations and protecting user privacy. This paper helps support my argument that recommender systems have the potential to cause privacy related issues

"Microsoft Attention Spans." *Microsoft*, Microsoft, 2015, https://dl.motamem.org/microsoft-attention-spans-research-report.pdf

This report by Microsoft studies how consumer attention span has changed with the widespread adoption of digital technologies. They found that the average attention span dropped from 12 seconds to 8 seconds in 2013. They conclude that even though a shorter attention span makes it difficult to retain consumer attention, consumers are still hungry for new things to jump to instead of focusing on what they are currently doing and so it provides an opportunity for short form advertisements to effectively capture their attention.

This report is useful because it shows how companies are reacting to a decline in attention span and offers evidence to support that our attention spans are much worse these days because of technology.

Seaver N. Captivating algorithms: Recommender systems as traps. Journal of Material Culture. 2019;24(4):421-436. doi:10.1177/1359183518820366

This paper describes the 'trapping' effect that recommender systems have. The author was able to interact closely with designers of many recommender systems employed in the US such as Willow's radio song recommender and Netflix. Seaver concludes that these systems are designed to hook the user into increased consumption of the product that the system is recommending. This leads to more profits for the company. This article is helpful because it gives an insider view of recommender systems.

Sparr, Matthew. "Explicit User Manipulation in Reinforcement Learning Based Recommender Systems." *ArXiv.org*, Arxiv, 20 Mar. 2022, https://arxiv.org/abs/2203.10629.

Recommender systems try to learn a good representation of the user's interests in order to give recommendations. The problem is that while learning to find this preference, it is actively influencing them. This paper discusses in depth why and how this is detrimental. This is useful because user manipulation is one of the areas I wanted to explore in my thesis.

Tariq, Saad. "The Hidden Side Effects of Recommendation Systems." *KTH ROYAL INSTITUTE OF TECHNOLOGY*, KTH ROYAL INSTITUTE OF TECHNOLOGY, 2021, https://kth.diva-portal.org/smash/get/diva2:1634235/FULLTEXT01.pdf.

This article from Sweden's largest technical institute gives an overview of the various harmful aspects of recommender systems such as privacy, behavioral manipulation, transparency, and unfairness. This article is important for my paper because it is a prior work in my area and gives a starting point to form arguments.

Zhu, Ziwei, et al. "Fairness among New Items in Cold Start Recommender Systems." *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, https://doi.org/10.1145/3404835.3462948.

This paper looks at problems with recommender systems when tested on users who just joined the service (cold start). Since the system has no information on the user, it is hard to make a good decision on what to recommend. One problem with current models is that it is very unclear whether they recommend things fairly. For example, are products from big companies being promoted at the same rate as products from small businesses assuming they are of the same quality? Does an AI-resume screener disproportionately recommend users of a particular race or gender? This paper argues that recommender system are not being fair in the cold start scenario. It then describes fairness formally and then a theoretical model which would be fair under their described fairness framework. This paper is important because models need to be fair or else people will suffer from injustice.