Data and Technological Intelligence: The Unspoken Beneficiary of the Modern Age

Katharine Batista

Cranford High School

Advanced Topics in Computer Science

Mr. Bell

June 5th, 2024

Table of Contents

Abstract

Data science and Artificial Intelligence are often coupled together as working together in the technological age. Why is the concept of data in society so hotly debated and what benefits does it present to our society? Research has concluded that the benefits of artificial intelligence and technological advancements outweigh the negatives it so glaringly presents. The research I have completed addresses the inconsistencies in information spread to the general public regarding Artificial Intelligence. I read several titles about data and the technological world that take a more existential look on how technology misunderstands society. My findings include how data bias affects the overall outcome of technological and Artificial Intelligence and what it means to understand the technological world. Ultimately, the significance of my findings extends beyond just knowing what AI is on the surface, but rather how it affects civilization as a whole.

Introduction

Artificial Intelligence, Machine Learning, and Data Visualization are relatively new ways of visualizing the world around us and human-produced data. Professional Data Scientists define them as "the representation of information in the form of a chart, diagram, picture." Society tells you that these things are inescapable. They're being used every single day in daily life whether society realizes it or not. There is an unspoken importance that regular everyday unsuspecting people play in the technological world. Human beings are the biggest source of data on the planet with carbon and digital footprint being more likely to be turned into a visualization or data set than ever before.

The machines and programs are not a direct threat to society, but perhaps the programmers building them are. Undoubtedly, the only difference between AI and a human data analyst is that Artificial Intelligence services such as ChatGPT, OpenAI, Anthropic, and Meta lack intellectual thought about the content it is reviewing. The modern world seizes to question the effectiveness of data analysis despite its benefits and challenges. Before technology, the government was seen as the invisible hand but now Artificial Intelligence and Technological Visualizations are the real invisible hand in civilization. They work behind your back to collect your data without your consent, but it is not always with malicious intent. People see AI as an invasion of privacy, society, and personal autonomy. It is harder to imagine a world where machines will ultimately surpass humans in intelligence and skill, than a world where it won't. Machines can only function on the bias they are built on, they do not have a mind of their own. But what if it did. Engineers certainly do not trust them to perform life changing surgeries or build algorithms that will fuel the next generation of medical devices. Artificial Intelligence is a concept so hotly debated that it remains in limbo, only to be debated over and over again against the same tired morals. Coupled by the cruel inability to trust search engines like Google or Microsoft Edge to make a simple internet connection. By delving deep into the interworking of a technologically reliant society, everyday people begin to understand the complexity, bias, and innovation surrounding the digital world. Addressing the intersectionality between technology and human beings can help us understand AI and Data Visualization, as well as if it is truly beneficial to the psychological state of society.

The importance of human connection in the context of technology will help computer engineers change the way civilization learns and experiences the world as the consumer. This is

also referred to as User Experience, and UX Design; mankind's relationship with emerging technology. Nevertheless, the benefits of the digital world come at a cost. The computational engineers of the whole have been working to find a way to balance the positive and negative effects of technological functions that continue to play the utmost important roles in life. Diversity plays a role in all of this as the technological world is divided between women and men, thus the gender gap increases and allows the other to live in isolation. Machines are biased against gender, race, ethnicity, and many other niche factors of humanity. Ultimately, engineers are trying to address what machines can do and what they can't without human assistance. Subtopics of this analysis will include feminism, critical visualization. Foundation of AI, Big Data, sorting, graphing, Python Machine Learning, and their effects on society as a whole.

Technological Algorithms: The Dangerous Benefits

Meredith Brousard's 2018 novel, *Artificial Unintelligence: How Computers Misunderstand The World*, a critique on digital culture and how people misunderstand the extent to which computers truly understand the human world. Broussard also highlights the idea that technology is harmful to society but only if it is weaponized. Cathay O'Neil's novel *Weapons of Math Destruction* discusses the intersectionality between math based algorithms and the digital world. O'Neil discusses the idea that mathematical algorithms make incorrect assumptions about people on a day to day basis due to their inconsistencies and harmful outcomes. These harmful outcomes go hand in hand with the idea of bias and data diversity. Ultimately, mathematical algorithms lack the same creativity that humans naturally possess. In the face of naivete regarding the hidden danger of algorithms that are implental in everyday uses of technology. Yet,

algorithms are essential to everyday life. Even the places that humankind visits everyday such as grocery stores, are set up algorithmically. As well as the traffic lights that you stop at every single day, or each time you say the words “Alexa play ‘Happy Birthday’”, you are moving through the world algorithmically. The idea that trigger words activate these technological devices is ingrained in the human mind, ultimately humanizing the devices as a whole. The humanization of technology and the digital world society operates in is the hidden danger of society, especially due to the fact that humans are blind to the truths of data. This is what makes the algorithms used to code the world's newest technological advancements dangerous to the common man. The idea that these programs possess personalities or human abilities is untrue. Computers misunderstand human language consistently despite the use of Natural Language Processing software that has been implemented into these programs. NLP’s are used to ensure a mutual understanding between computers and humans. The design allows for the humanization of artificial intelligence and interactive technological algorithms. Each data driven decision that computers are able to conclude based on datasets, are extremely limited to their algorithms and a unique set of rules that a programmer implements within the system.

Similarly, Relational databases (RDBMS) are datasets that use numeric, alpha, and dates. Any data that does not fall under these categories are significantly more complex for algorithms to process. So, while technological algorithms have the ability to process a lot of the information it is being fed, not everything is completely accurate. Algorithms are specifically programmed to operate under the guidance and coding abilities of humans. Engineering professionals are having a hard time differentiating what is real and what is misinformation which is what makes these technological algorithms potentially dangerous in the future.

Diversifying The Data World

In recent years, data scientists have discovered that is not always diverse. While the ideas and content of data may be diverse, the reality is that it still contains basis, and a lot of it. So much so that it has started to affect the way humans see and navigate the world. So what does it mean when humans are all just data? Civilization can never be broken down into “just data " because each person is incredibly niche and diverse in their own way. But as analysts become more familiar with data sets they are discovering that data is not exactly the most diverse thing known to man, and typically people will pick and choose what they want their graphic based analysis to look like. These same engineers build their own, or use outside sourced programs that also include basis. The big debate lies in whether or not software developers and UX (User Experience) designers can analyze data with as little bias as possible. Machines similar to the one used to produce the graphs in this study transfer data into Python programs Basic machine learning programs can analyze traditional numerical patterns and trends within CSV files. The importance of UX design is critical to the development of user interaction. Creating diversea and well rounded applications is This teachable machine can recognize classes based on images, creating a more niche program that can identify the underlying traits from the user's data. However, machine learning comes with bias in gender, and physical appearance depending on the maker. An example of one sided algorithms includes facial recognition, in which millions of images are being pulled to identify the human features of the face and perform a single action. Facial Recognition poses one of largest sources of bias in society today. Specifically, programs that are involved in criminal investigations, or in court. Data visualization is essential to these projects because it is rooted in the way that humans interact with computers. Evidently, the

importance of diversifying datasets is critical to the general public's perspective of the technological world.

Critical visualization is a term that was coined by Peter Hall in 2008, meaning the technological view of data analysis. Focusing on how power is used and distributed during the design process as well as how engineers can bring more diversity to the design process to push the feminist idea that data must be more diverse to make it more reliable. His textbook reading entitled *Critical Visualization: Rethinking the Representation of Data* discusses what is normal in the context of data and humanity. Hall reflects on the intersectionality between the two worlds and how humans can stay mindful of the world around them without allowing technology to consume them and their world views. The text further approaches the idea of diversity in algorithms and ensuring that the software engineers building the technological interfaces and programs society uses today, appeals to all walks of life.

The Misunderstanding of Artificial Intelligence

The technological world does not appeal to everyone, but technology is present in the lives of almost every single person living on earth. But for most, technology and more specifically Artificial Intelligence, seems to be an overarching and glaring issue. Artificial Intelligence only possesses the abilities that a human programmes it to inherit. The skills of an interactive AI Large Language Model (LLM) are less than limited because they are unable to think for themselves. Artificial Intelligence bots are only able to answer the most basic questions, the ones most regular people are asking. But, AI and its partners are not going to take over the world. Since these Large Language Processors are still reliant on human development they are

unable to communicate with themselves or change their own program niches. Inter-Artificial Intelligence communication would confuse the program even more because instead of using Natural Language Processors, the computer would only be learning from something that has the exact same information and language patterns as them.

In recent years, they have been working toward coding AI to program itself. However, AI cannot do this as it relies on human interaction and coding softwares to function. This is why humans misunderstand the direct functions of Artificial Intelligence. The regulation of AI is now on the rise with the continued misuses of its services. Hannah Fry's novel, *Hello World: Being Human in The Age of Algorithms,* makes a point to describe the regulations being put in place in the United Kingdom. She writes there has been "a new piece of regulation called GDPR - General Data Protection Regulation "(Fry 46) , that will make most of the data technology collects, unusable and illegal to distribute. Data, algorithms, and software do not have the ability to understand habits, each one is coded to understand that repetition in searches, or times of day creates a method called a habit. The misunderstanding of AI is something that is not new, but it is a critique of society. It directly represents how well people are educated about what Artificial Intelligence truly is and what it truly is not on even the most basic level. While, Ai is extremely beneficial to the way mankind operates, it must be used and understood properly in order for society to not lose its touch on human connection. Artificial intelligence will never possess the same creativity and humanity that humans do and that is the biggest misunderstanding in the technological world today.

Data through The Feminist Lens

Feminist data visualization was first introduced to me during my time with the IT Girls: Summer College Program at Syracuse University. A small group of girls specifically interested in data visualization conducted an experiment in which one singular habit was tracked for 10 days. The goal of feminist data visualization is to expose the assumption made in the choice of teachable machines as well as what the intended audience is for the model, ultimately diversifying the User Experience and data outputs overall. This project emulated 'Dear Data', started by Giorgia Lupi and Stefanie Posavec from opposite sides of the Atlantic Ocean– tracking their habits through visualization. Data visualization focuses on the keywords that are employed to deal with the humanities of how knowledge is transmitted into data analysis. The evolution of feminist theory in data has impacted the way that machine learning is viewed by those in the humanities. Mainly dealing with the ideas of how identity is constructed and or perceived by machines. By just looking at the technological world, you would most likely be able to detect that almost all forms of technology are dreamt, designed, engineered, built, and created by male engineers, white ones. Meredith broussard recounts her time at Harvard University in her novel *Artificial Unintelligence: How Computers Misunderstand The World*, saying that she was "one of six undergraduate women majoring in computer science" and ultimately she saw what was wrong with the system around her, that it was built by men and only men. Regardless of your view on feminism it is important to note that a technological world built solely by the male gender is a biased one. And, unfortunately it was a failing system she was unable to fix, she switched her major and received a Bachelor of Arts instead (Broussard 5). Broussard is just one example of undergraduate programs in computer and data science failing to

support their female students. Unsupported female STEM students have always been an issue since the dawn of time, or since women were allowed to go to college. Only now it is coming to light in the most mundane ways. Gender bias in data has always been in issue, as men are not likely to cover female specific issues when creating their new technology. The little things that are exceedingly important in the life of women, are not as pressing in the lives of their male counterparts. The gender bias can decrease with the awareness of those being and producing these visualizations and algorithms.

Caroline Criado-Perez's novel, *Invisible Women: Data Bias in a World Designed for Men,* explains the discrepancy between technological advances and the gender gap. Perez dives deep into her analysis in the fourth chapter, *The Myth of Meritocracy,* saying "Tech's love affair with the myth of meritocracy is ironic for an industry so in thrall to the potential of Big Data." Perez explains that meritocracy, the idea that any one person can rise to the top of society if they possess the skill, is ironic for an industry so raptured by the idea that big data could take over the world despite its lack of knowledge and individual skill level. Diverse data representation is critical to society and its innermost niche regarding gender and sexuality. Feminist data visualization allows for all data to become equal and for all mankind to possess equal opportunity in the new age of technology.

Harmful Effects of Sorting in Data Visualization

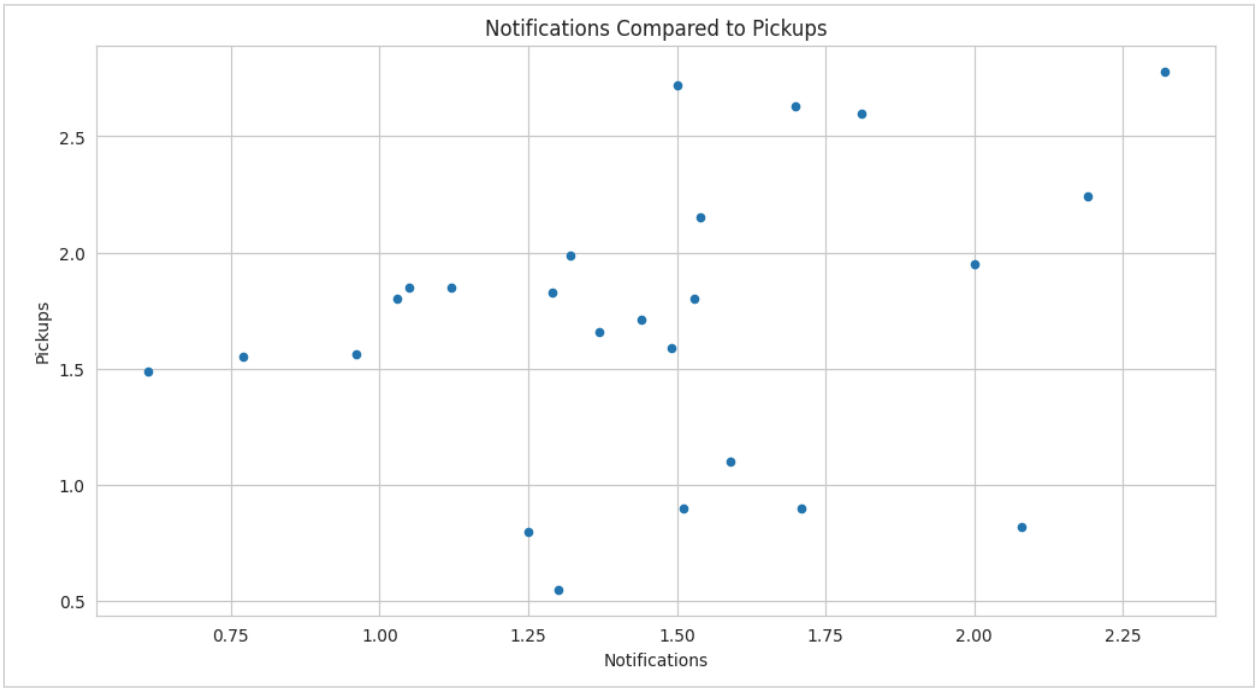
Data Visualization deals with diversifying the data that technology collects over time. Computer algorithms can only be as diverse as the data they are testing. Class sorting is used by companies like Google and Pinterest, in services such as Reverse Image or Google Images. But

what kind of results do analysts acquire based on the bias of search history, region, location, and human interests? Teachable machines or algorithms are only trained on the data they consume. as well as what the intended audience is for the model. Companies are looking to marginalized communities to identify the design and user experience to diversify their products. This also allows them to analyze what knowledge is being excluded from the current knowledge of their visualizations. The categorizations of different images and phrases produces a bias against these visualizations as the CSV files become more specific to one entity.

Sorting is critical to the developmental discrepancies between data bias and technological advances. Services such as Google Image developed their own bias toward the images they are viewing. Hence, if you were to put the word ‘ women’ into the search bar, versus the word ‘men’ the search result would vary. Particularly, their algorithm is based on impending bias toward men and women that the algorithm picks up from the most popular Google searches. While every search process must use sorting to determine the importance, content, or size of an image it still is biased toward things such as race, gender, and ethnicity. These harmful effects are the product of inconsistent algorithms and methods, engineers are working to alleviate some of these issues to make data visualization beneficial to all. Data visualization is a critical part of the way regular people can understand the world around them. So it is essentially that engineers work to understand the biases behind their own algorithms that are being implemented into programs society uses everyday. Despite the growing database and diversity presented in such datasets, no one person can possibly fit into this average blueprint created by these algorithms. And still, humans interact with this software every single day, on their phones, on the street, and anywhere they least expect it.

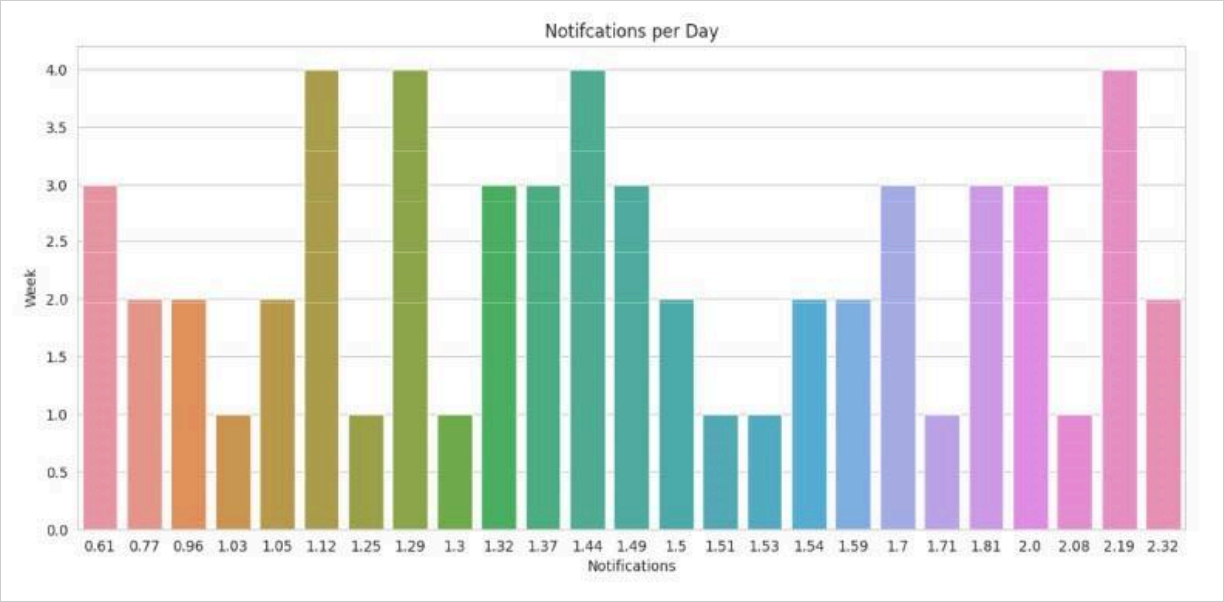
Analyzing Visualizations: How to Understand Your Data

Despite all of the efforts to change the biases in visualizations, they are inevitable as seen in the ones produced from my own Natural Language Processing program based in Python built in Kaggle. Started by forking a basic NLP (Natural Language Processing), supported by accommodating the different types of data. (Strings, Standard Numerical, and Decimal Values). The data tracked included how many times in 24 hours I received a notification, picked up my cell phone, and went on iMessage immediately after pickup. I used add-ons such as Matplotlib, a comprehensive plotting library that serves as an object-oriented API (Application Programming Interfaces) for the data imputed by the user in a dataset (CSV Files). Application Programming Interfaces allow for overlays such as Matplotlib to communicate with the CSV and create interactive, comprehensive, and flexible visualizations. Seaborn, similar to Matplotlib, simplifies the process of data visualization while allowing the user to create a more complex graph such as a heat map, or pie chart. Using a Python based machine learning tool, a dataset is implemented



and adjusted for the NLP model. However, as the process continued throughout it became apparent that large numbers would not run smoothly through the program. Thus, the dataset changed from whole values to decimal values in order to represent the amount of pickups in comparison to messages received on a singular given day.

Figure 1, the scatter plot pictured above, is a representation of how many notifications an iPhone receives on a given day. This plot is basic enough for an untrained eye but complex enough for someone who is familiar with data visualizations to understand the objective of the



model. While this NLP can only build basic heat maps, bar graphs, and dot plots, the program is still able to process large amounts of data and output these simple visualizations. As previously discussed, data contains major bias against humans, which is apparent in these visualizations as the data is only being collected through one source and does not pull from multiple different places and or walks of life. In a professional experiment, multiple sources would have been used but for the purpose of this exploratory analysis it was deemed unnecessary.

Figure 2, a graph depicting which day and week a certain amount of notifications were received to a singular iPhone. This graph is easy to understand by the untrained easily due to its soothing colors and simple bar graph. The graph directly explains why data visualization is a great way for people to understand the complexities of natural language processors and python machines that produce these graphs. Ultimately, understanding these visualizations is not

```python
# Add title
plt.title("Notifcations per Day")

# bar chart shwoing the nofications for everyday, sorted by week.
sns.barplot(x=yelp['Notifications'], y=yelp['Week'])

# Add label for vertical axis
plt.ylabel("Week")
```

Python

particularly difficult despite the daunting code that operates to output them. The code segment above is adjusted to create figure 2, the multicolor bar chart. It contains the plot title, the sns barplot that will be displayed, and lastly the y axis label. By using systems such as Matplolib and Seaborn, the code is able to pull from these massive libraries in order to create these visualizations.

Although bias is inevitable in the algorithms and datasets used to create these visualizations it is still particularly important to the process of data diversity and digital advancement. Analyzing these visualizations is something that anyone in all walks of life can understand.

Acknowledgments

Thank you to my parents, they never knew that I was so nerdy about this kind of thing until they gave me the amazing opportunity to attend Syracuse University during the summer before my junior year. My parents always encouraged me to do whatever I wanted and be whomever I wished to be, except an astronaut or a soccer goalie. But this joy, the one I so passionately write about-- is all thanks to them. Women go into STEM all the time, people don't bat an eye. I wanted to make people really dig deep about what technology means to them and how much they really know about the world that is so diligently tracking their every move. They always taught me to dig deep and ask people what it meant to *feel*. This task was no easy feat and I would like to thank Mr. Bell for being nothing short of supportive despite my ideas hitting a wall at times. Thank you to all the *female* scientists, mathematicians, and authors I have accredited in this paper. Without women in technology, and women writing about technological innovation I would not have been able to find my niche. I started working with data and Artificial Intelligence more and more as Advanced Topics in Computer Science progressed and it only got me more excited to bring my skills to the real world. I spent a lot of time researching and narrowing down the chapters in this paper to what I believed was truly important for people to know. I know that not everyone is interested in the things that I find interesting and truly fascinating about the technological world. The world at large interacts with technology everyday and whether it is beneficial or not it is changing the daily functions of humanity at a rate so rapidly even the best computer, data, life scientists in the world can keep up with everything.

There is so much to explore in this world and everyday that I have spent exploring what it means to be human in the intersection of technology and humanity has left me starstruck.

References

Broussard, M. (2019). *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press.

Criado Perez, C. (2019). *Invisible Women: Data Bias in a World Designed for Men*. Harry N. Abrams.

Deiglmeier, K. (2023, Spring). Seize the Future by Harnessing the Power of Data. https://ssir.org/articles/entry/seize_the_future_by_harnessing_the_power_of_data

Fry, H. (2018). *Hello World: Being Human in the Age of Algorithms*. W.W. Norton.

Hall, P. A., & Dávila, P. (2022). *Critical Visualization: Rethinking the Representation of Data*. Bloomsbury Academic.

Kartikey Pillay, T. (2023, March). *basic NLP Yelp review Analysis* [Python]. Kaggle. Retrieved January 21, 2023, from https://www.kaggle.com/code/tmkartikeypillay/basic-nlp-yelp-review-analysis

Shane, J. (2019). *You Look Like a Thing and I Love You: How Artificial Intelligence Works and Why It's Making the World a Weirder Place*. Little, Brown.