Data feminism: Numbers cannot explain for themselves

Data feminism introduces several problems people may face when failing to achieve the correctness of data uses. Here are important factors for us to notice for data interpretation and applications. It can be misleading if we fail to utilize data properly.

Unveil the Pitfalls of Big Data: applying correctly

First, the essay introduces the 2014 Chibok schoolgirls' kidnapping news in Nigeria, which uses the subsequent misleading data published by FiveThirtyEight. The initial report claimed an alarming increase in kidnappings with the terrifying number of 3608, but later its statement was retracted due to inaccurate application of GDELT's data.

The number of 3608 is the global data of kidnappings instead of being as simply Nigeria's. This report focused on pretentiousness rather than contextual accuracy, and this should be consistent with the term "Big Dick Data". The data is correct, but it should not be here as it cannot match the contents exactly.

Overall, appropriateness and context in data interpretation are important, and people should avoid common mistakes such as disregarding context, fetishizing size, and exaggerating capabilities. These will question the reliability of the database and raise worries about the trust of media reports.

Situating Web Data: correctness, missing part, and biases

Besides applying the right data with an inappropriate context, we face challenges presented by uncontextualized data obtained from web portals and APIs. The lack of accompanying context causes significant confusion in understanding the intricacies of the contents behind the data. Such context-free data may create misinterpretation and incorrect use. Therefore, we should be aware of "zombie data", whose datasets are published without a clear purpose or introduction and thus the incorrect utilization.

As Chris Anderson believed that "the numbers speak for themselves," it highlights the importance of context in interpreting correlations and the potential harm caused by biased algorithms. For instance, a Google search will help reinforce harmful stereotypes. Another example should be sexual assault data, illustrating how imbalances of power in data collection environments can distort the narrative and the necessity of understanding context to reveal the truth hidden in the numbers. Therefore, ignoring necessary data context will also lead to errors.

Context, Creativity, and Ethical Inquiry: Raw Data, Cooked, Data Cooking

Assume that we consider data as raw input and suppose data is already "cooked" due to various social, political, and historical influences. This perspective challenges the conventional idea of "raw data" and introduces the intriguing concept of "data creatives" with the key element of context-hop: proficiency in creatively mining and combining data to discover new insights and the ability to work across diverse domains.

The narrative critiques the common tendency to allow "the numbers to speak for themselves," as Google's Flu Trends and the Sun Sentinel's investigative project highlight the potential pitfalls of interpreting data without proper context.

The emergence of a data creative class is highlighted, but we should also question the reliability of such insights. From social media data's behavioral insights, platforms like Instagram and Reddit introduce biases into the data, which may be due to a tendency of more young users. Therefore, the sampling method (how we choose the sample) may affect the research results.

To be specific, looking for what is missing from a dataset will help gather insight into both the data and the phenomenon it represents. Several case studies, such as historical analyses of Thomas Jefferson's letters and word embeddings exploring gender and ethnic stereotypes, illustrate how the discovery of the cooking process of data reveals hidden biases and structural inequalities.

It seems that the analysis contradicts with the opinion that less theory is required for meaningful data analysis. Relying solely on intuition and "common sense" may bring bias. Moreover, context and computation are not mutually exclusive. SAFE lab's approach to using artificial intelligence to study violence in Chicago highlights the significance of integrating knowledge and sustaining relational human and infrastructure in research projects for social influence and justice.

Communicate text:

The critical roles of context are not only data acquisition but also the important elements of framing and communication of results. Let's turn to the scenario presented involves a data

journalist tasked with creating a graphic and short story about a research study on mental health disparities in the New York City Jail system.

Then we contrast graphics with different titles and framings. The first graphic "Mental Health in Jail," is the usual way of communicating data analysis results, which has neutrality and is a problem in this situation. The second graphic "Racism in Jail: People of Color Less Likely to Get Mental Health Diagnosis" will be more context-aware and accurately represent the research study's findings.

Naming and acknowledging structural forces of oppression is essential in the communication of data. Some fields, like journalism, may resist naming strong biased words due to concerns about bias or opinion. In fact, a specific name such as naming racism is not a matter of bias but a reflection of empirical evidence and facts, and it challenges the notion that such naming is inherently opinionated.

Data journalists should connect the research question to the results for better interpretation. This opinion argues against "letting numbers speak for themselves", stating that placing numbers in context and naming oppression is essential for accurate interpretation and avoiding misinterpretation. To be specific, the subtitle "People of Color Less Likely to Get Mental Health Diagnosis," may contribute to a deficit narrative, reducing social groups to negative stereotypes. If the subtitle is modified to "White People Get More Mental Health Services," attention may be shifted to the dominant group.

Restoring Context: Nurturing Context in Data

People have concerns about power imbalances and inequality in data collection. For instance, WomanStats, incorporating qualitative variables, highlights challenges in capturing complex realities. The question of responsibility for providing context is raised and are seen as vital for cleaning, contextualizing, and verifying data. Besides, we also need significant investment in providing and maintaining context.

Consider Context:

Data feminism's sixth principle means understanding the environment and power dynamics in data collection. It emphasizes preventing numbers from speaking for themselves, particularly concerning people to avoid perpetuating harm. All in all, analyzing power imbalances, acknowledging missing data, and originating contexts are necessary steps. Finally, dedicated funding is crucial to sustain the standard contextual work. We should treat data with dedication and remember those mistakes that may take place to provide the accurate analysis.