Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

Digital Reputation Index with People Chain – an AI driven methodology to combat fake news, and create a "Trust worth" for facts and information online

Jai Vishwakarma
Suresh Gyan Vihar University, Jaipur, India

Abstract

"Data may be a Powerful Weapon. Use It Responsibly – And Tell folks." (Elliott, 2018) In his article revealed in 2018, he pointed out some major issues on however information was being mis employed by firms and other people to their profit. Verint International Study of over twenty four thousand customers across twelve Countries Indicates want for Organizations to Strike the correct Balance Between customized Service, information Privacy and Transparency:

• Nearly nine out of ten (89%) customers surveyed say it's important they acumen secure their personal data is.
• 86% need to grasp if their information are passed on to 3rd parties for selling functions.
• Personalized service continues to be vital, with eightieth of customers' locution they find it irresistible once service is ready-made to them and their interests.
• Verint business survey finds that issues are aligned, with businesses putting stress on the importance of information privacy (94%), personalization (95%) and resolution queries quickly (92%). (Verint Systems, 2017) Governments across the globe have already complete these challenges, and one in every of the solutions to safeguard voters in European region is Regulation (EU) 2016/679 of the ECU Parliament and of the Council1, the ECU Union's ('EU') new General information Protection Regulation ('GDPR'), regulates the process by a private, an organization or a company of non-public information concerning people within the EU. It doesn't apply to the process of non-public information of deceased persons or of legal persons. The rules don't apply to information processed by a private for strictly personal reasons or for activities administrated in one's home, provided there's no association to an expert or enterprise. once a private uses personal information outside the non-public sphere, for socio-cultural or monetary activities, for instance, then the info protection law must be revered. (European Commission, 2018)

Hence the need for a mechanism where "**Digital Reputation Index" for individuals' online behavior**, and **People chain to ensure mis-information, rumors are in check, verified and only "trust worth" facts are in circulation**, more importantly a **more responsible social behavior "online" can be expected**.

Keywords: Ethical AI, Digital Reputation Index, Face book, CITIZN, People chain, Trust worth, Fake News, India, Social Media, Mis information, Social Dilemma, GDPR, Anxiety, Depression, Mental Health, Online Behavior, Cognitive Bias, Cognitive Bias Modification, Online Reputation Management

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

Introduction

"**If you're not paying for the product, then you are the product**" (The Social Dilemma, 2020) In today's age and time, "**Social Media is a drug**" (The Social Dilemma, 2020), and while we are realizing it a bit late, the damage that it has done is profound. Social Dilemma[1], a movie released in the year 2020, has brought in some of the facts and the ugly side of social media, human computer interaction, unethical use of AI, and the dark side of mis information to the table. There are some of the dilemma's that it talks about, however the most profound ones being:

Mental Health Dilemma –

A 5,000-person study found that higher social media use correlate with self-reported declines in mental and physical health and life satisfaction.

Face-to-face social interactions enhance well-being. With the omnipresence of social media, necessary queries have arisen regarding the impact of on-line social interactions. within the gift study, we tend to assessed the associations of each on-line and offline social networks with many subjective measures of well-being. we tend to used three waves (2013, 2014, and 2015) of information from five thousand two hundred eight subjects within the across the nation representative Gallup Panel Social Network Study survey, together with social network measures, together with objective measures of Face book use. We tend to investigated the associations of Face book activity and real-world social network activity with self-reported physical health, self-reported mental state, self-reported life satisfaction, and body mass index. Their results showed that overall, the utilization of Face book was negatively related to well-being. for instance, a 1-standard-deviation increase in "likes clicked" (clicking "like" on somebody else's content), "links clicked" (clicking a link to a different website or article), or "status updates" (updating one's own Face book status) was related to a decrease of 5%–8% of a customary deviation in self-reported mental state. These associations were strong to variable cross-sectional analyses, still on 2-wave prospective analyses. The negative associations of Face book use were appreciated or larger in magnitude than the positive impact of offline interactions, that suggests an attainable trade-off between offline and on-line relationships. (Holly B. Shakya, 2017)

Democracy Dilemma –

The # of nations with political misinformation campaigns on social media doubled within the past two years.

New York Times, mentioned, the researchers compiled data from news organizations, civil society teams and governments to form one in every of the foremost comprehensive inventories of misinformation practices by governments round the world. They found that the quantity of nations with political misinformation campaigns quite doubled to seventy within the last 2 years, with proof of a minimum of one organization or government entity in every of

[1] https://www.thesocialdilemma.com/

these countries participating in social media manipulation. additionally, Face book remains the No. one social network for misinformation, the report aforementioned. Organized info campaigns were found on the platform in fifty-six countries. However, the analysis shows that use of the ways, that embrace bots, faux social media accounts and employed "trolls," is growing. (Davey Alba, 2019)

Discrimination Dilemma –

64% of the people that joined extremist teams on Face book did therefore as a result of the algorithms steered them there.

According to a Wall Street Journal report, Face book determined to require no important action when internal analysis incontestable that its algorithms were stoking political orientation and division. one in all Face book's internal displays from 2018 expressly declared that its algorithms – that boost bound content that targeted users could also be additional probably to move with – are exasperating discordant behavior and would still do therefore, the report aforementioned. "Our algorithms exploit the human brain's attraction to divisiveness. If left unbridled, Face book would feed users additional and additional discordant content in a trial to achieve user attention and increase time on the platform," one slide browse. A separate 2016 study written by Monica Lee, an indoor analysis person, found that sixty-four per cent of individuals WHO had joined associate extremist cluster on the platform did therefore as a result of the cluster was promoted by Face book's automatic recommendation tools. (Editorial, 2020).

Additionally, in another report revealed within the same media, states that Face book has insisted that it'll still allow lies in political advertising on its platform, despite searing criticism, because it announces new options to relinquish users a minimum of some management over political ads bestowed to them. Given the premise, we want to act and produce in a very resolution to those dilemmas. The terribly essence and magnitude of those issues is thus large that we would not be ready to solve it in one go, but we will try and address it via technology, the exact same that created it, has the solution to that, in conjunction with a combination of human intelligence. **"Social media starts to dig deeper and deeper down into the brain stem and takes over kids' sense of self-worth and identity**" (The Social Dilemma, 2020), it's time we brought in **ethical AI, Digital Reputation Index**, and **human intellect** to the mix so that we can create "**trust worth**" information that is not creating undue bias, but only strains out facts, data and leaves the judgment to the individual without orchestrating facts to the platform owner's advantage. In accordance to the challenges, a way to bring moral conduct to online behavior via "Digital Reputation Index" that can be trusted and people adhere to, will be an attempt to solve a lot of defamation, hate speech and more, online.

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

Framework of Study – Research Structure

Based on the what we are trying to achieve, which is going to be an exploratory research with data being the driving force for insights, we are going to take a data science approach for establishing facts and deducing correlation following which the solution will be presented.

The idea of the framework is keeping it repeatable considering the problem; where threats will continue to have newness and the model of people chain and digital reputation index will need to match and follow closely to be relevant, we are proposing the data science methodology as suggested by IBM. (Rollins, 2015).

The premise of this research being the hypothesis as below:

H_0
Rise of social media has made an adverse impact on physical wellbeing, social well-being, and has adversely affected information reliance and its accuracy, along with the people who use them for their analysis.

H_1
While research has extensively been conducted to prove the correlation of the social media vs the wellbeing (Holly B. Shakya, 2017), there is no considerable work being done on how to control this "**eDemic**" of online abuse of facts and mis information, bring in "**responsible online behavior**" by bringing in reputational wellbeing for the individual. Hence the need is immediate to bring in such a matrix and model that can help being a sense of responsibility rather than callous posts, in their online behavior and engagement.

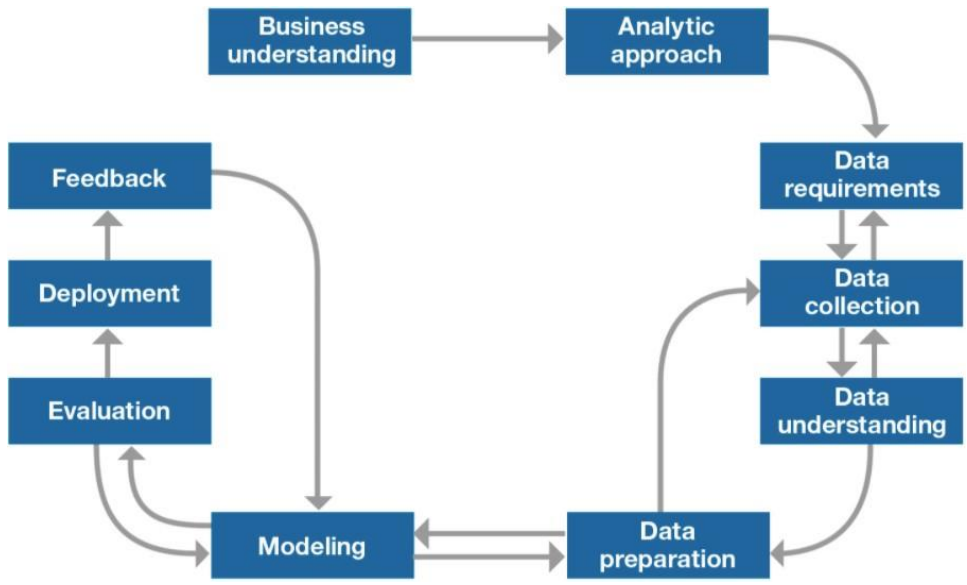

Figure 1: Foundational Methodology for Data Science by IBM (Rollins, 2015)

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

Literature Review

In terms of some of the research around work of importance around the topic of digital reputation, and cognitive bias modification there are some papers as below:

In their research (Eryarsoy, 2015) they called out the need for a digital reputation of companies as the internet and social media presence dominates the reputational factor. Their pivotal research question was on how to represent digital reputation of any organization in a digital paradigm. Their approach was a quantitative methodology to collect data from social sites, company web pages, blogs, wikis etc about a company and rank them against each other. However, what their research wasn't addressing was, how do we manage the same construct for an individual, fact check and use of AI to diffuse fake news.

In other research, focused on cognitive bias, decision styles, and how they impact an individual (Gloria Phillips-Wren, 2019) it was discussed that systems and the platforms must help the individual decision makers think rationally. They brought out the point that Cognitive biases are closely related to how we human beings decide. Hence the idea to ensure that the biases are not fueled by mis information, and fake news, it's important to build a system that can easily help remove un verified data and only let the verified, checked and trust worth information reach the end users.

Harvard Business Review, also published an article (Jack B. Soll, 2015) where the authors explained how one can overcome their own biases. Most of the times online users are over confident about the information they already have and give it more weight than anything else, and we cannot see the future that clear hence we believe whatever is being presented. The below image is a good representation of their work:

Figure 2: How to Prevent Mis weighting (HBR)

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

Objective

The objective of the research is to develop, and implement a working solution / model that can:

- Detect False News from various online medias
- Detect various Biases
- Gamify user's collaboration to give them reputational scores
- Higher the score = Better the individual's online behavior and participation to strain out false rumors and news
- Create a People Chain to map along with AI to confirm the authenticity of the news / information online
- Be a benchmark method and mode for "Digital Reputation Index" for Online media
- Provide a Trust worth score to news and information online

With the objectives as a premise the way we have structured the logical flow of the algorithm is illustrated as below:

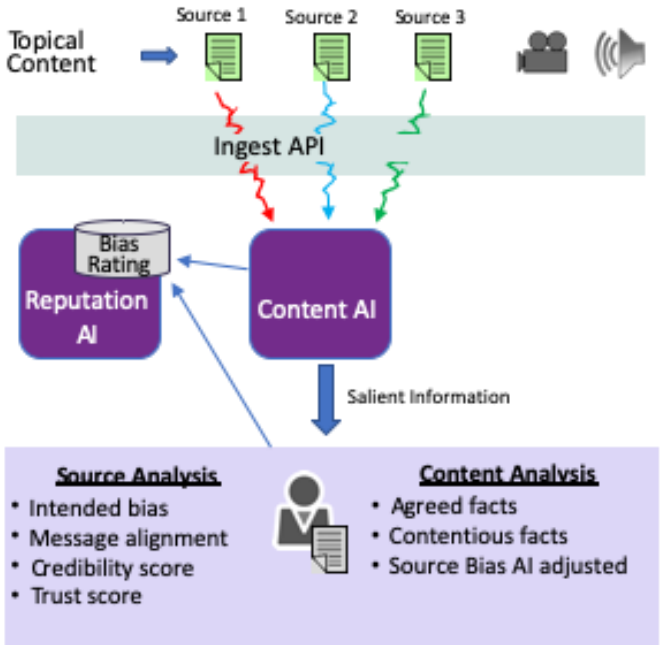

Figure 3: Content AI driving the Reputation AI - Conceptual Model (CITIZN)

Content AI reduces media distortion by distilling topical news into salient information with confidence, while also identifying slant and alignment of each content sources.

- Ingest of topical 'news' from select sources
- CITIZN (a societal platform)[2] analysis and comparison of content, adjusted for source bias score.
- Presentation of simplified salient information

[2] www.citizn.world

- Source content rating for bias and alignment with political agenda

In addition to the above we will then decipher the activities of the user, the type of content s/he is engaged in, and accordingly provide a "Digital Reputation index". The conceptual view of the architecture is as below

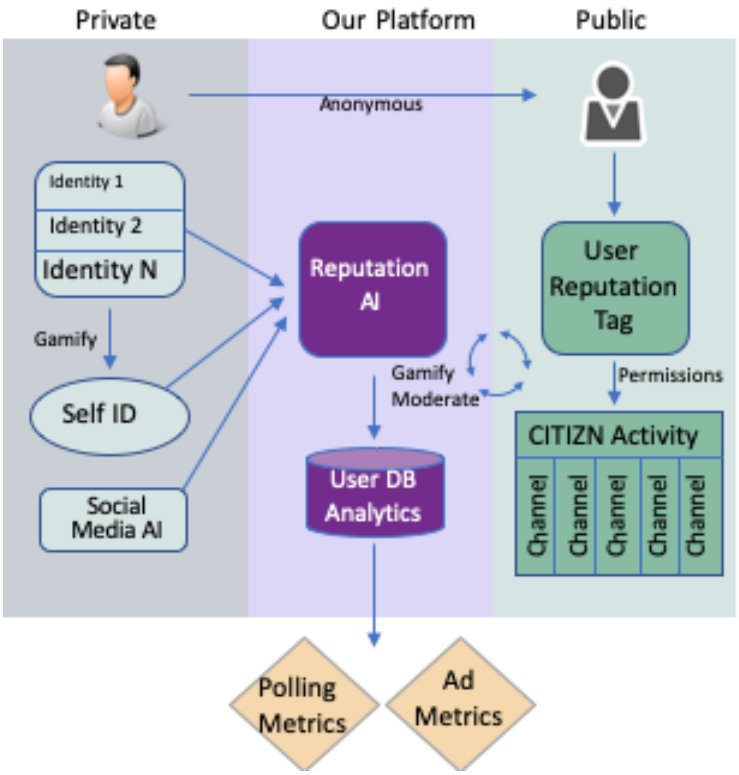

Figure 4: Conceptual Architecture of Reputation AI flow (CITIZN)

Reputation AI is core to user engagement, delivering a managed, safe and positive experience, while also building depth/validity in user data:

- Validate user identity, location, person
- Maintain public anonymity
- Gamify user engagement
- Create a 'safe-space'
- Enable free expression of credible opinions
- Maximize depth of user data

An overview of the platform to explain how the flow of information will be enabled by AI.

U = User. The image below illustrates some of the cognitive bias. (Gloria Phillips-Wren, 2019)

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

The idea has been influenced by Cognitive Bias challenges, and is an attempt to counter the biases by facts.

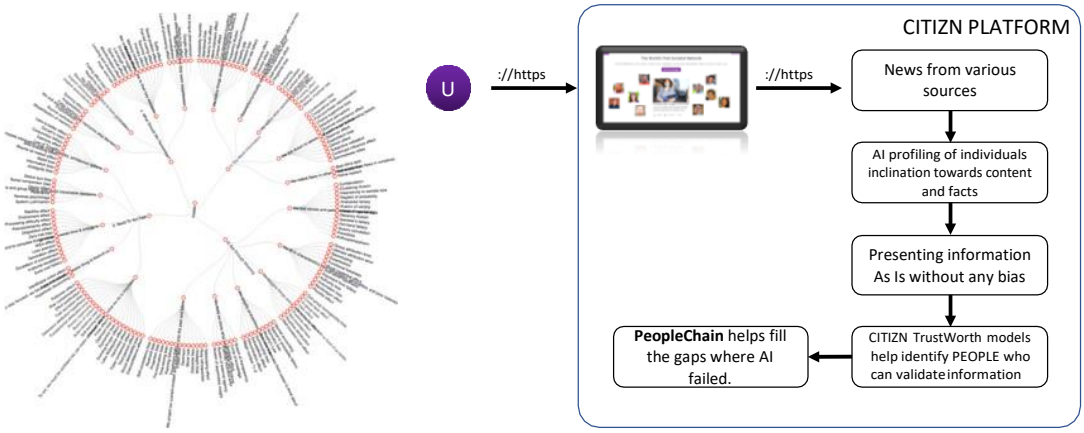

Figure 5: Enabling Information and Facts via AI and People chain (CITIZN)

Research Methodology

Given this is a Data Science (Cleveland, 2019) challenge and there would be a need to establish the facts in order, bring in the right process to structure the data, aggregate data for individuals from various sources, eliminate biases, and more, then be able to give scores which will lead to an index of scoring for individuals, we will approach this problem with an established data science approach (Rollins, 2015), however with twist of our own approach.

The approach that we propose is as illustrated below, where the focus is not just research, but bringing in a lot of AI and Data Science approach to solving a real-world problem that is very prominent at this age and time.

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

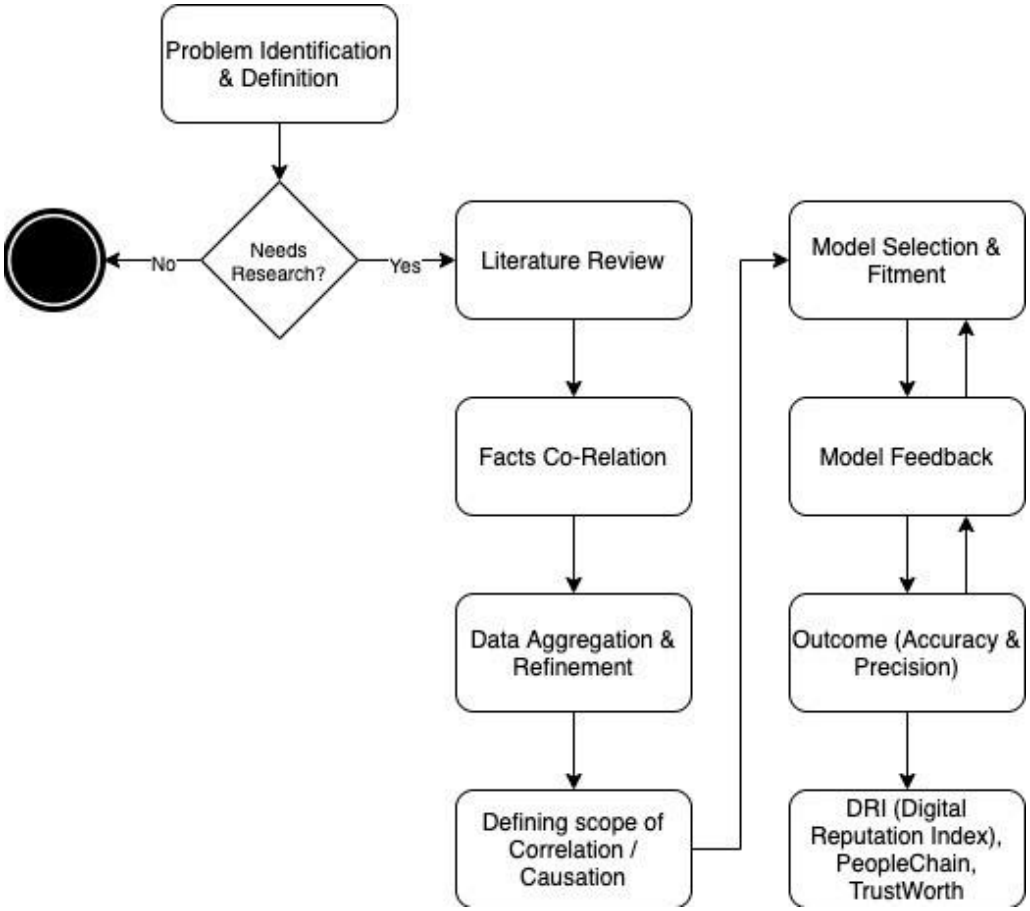

Figure 6: Research Methodology for Digital Reputation Index, People Chain and Trust worth (CITIZN)

Design of Experiment

To set the basis of this research I have taken the liberty to use the construct of America in One Room (CDD, 2019), a national experiment which was conducted by Stanford CDD, where by 523 voters were sampled (stratified) from across the country and stationed at Dallas for discussions on civil matters from Immigrations, Healthcare, and Foreign Policy, to Tax & Economy. The highlight of the experiment was to showcase how biases cloud your judgement and availability of perspectives and information can help you steer right.

The mass experiment was conducted during September 19-22, 2019, where sample met in small groups in a moderated session and shared their views against experts and politicians from both Democrats and Republicans. There was a control user group of 844 participants who addressed the exact same questions as the sample, and was recruited by NORC (University of Chicago).

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

Results & Findings

The findings are showcasing the impact of how informed people gets through biases. The table below compares the amount of correct responses of deliberation participants before deliberation (T1) and after deliberation (T2). Correct answers are mentioned in parenthesis and the statistical significance was reached using popular paired t-tests. Percentage represents sample who answered right. Numbers in parenthesis are the size of the sample. +p≤0.1, *p≤0.05, **p≤0.01, ***p≤0.001.

Participants

	T1%	T2%	Change
Which political party holds the majority in the Senate? [Republican]	74.4 (389)	72.3 (378)	-2.1
Which political party holds the majority in the House? [Democrat]	71.5 (374)	69.6 (364)	-1.9
About how many undocumented immigrants are in the US? [10 million]	20.5 (107)	74.0 (387)	53.5***
Which of the following countries is NOT part of the Paris Agreement on the environment? [All of the above]	32.1 (168)	48.6 (254)	16.5***
The Affordable Care Act allows which of the following? [All of the above]	57.0 (298)	60.0 (314)	3.0
What percentage is the highest tax rate for capital gains taxes? [20%]	24.5 (97)	46.8 (93)	21.8***
Which of the following organizations dealing with trade has the most countries? [WTO]	40.5 (212)	49.7 (260)	9.2***
Knowledge Index	**45.8**	**60.1**	**14.3***

Figure 7: Knowledge Gain Table on Participants

Figure 8: Democrat and Republican polarization view on all proposals before factual discussion

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

Figure 9: Democrats and Republican views after open discussions

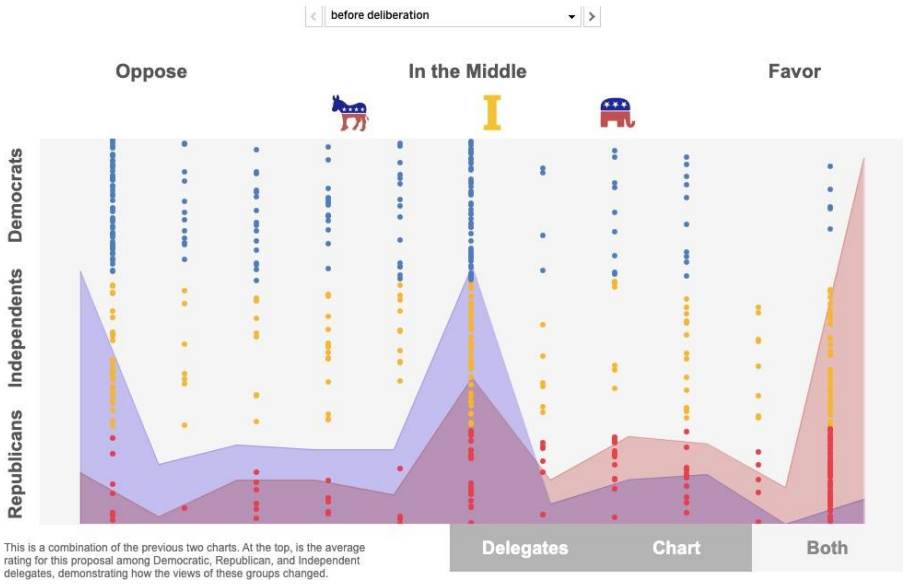

Figure 10: Distribution before the factual discussions on issues

Suresh Gyan Vihar University Journal of Engineering & Technology
(An International Bi - Annual Journal)
Vol. 7, Issue 1, 2021, pp 10 to 24
ISSN: 2395 - 0196

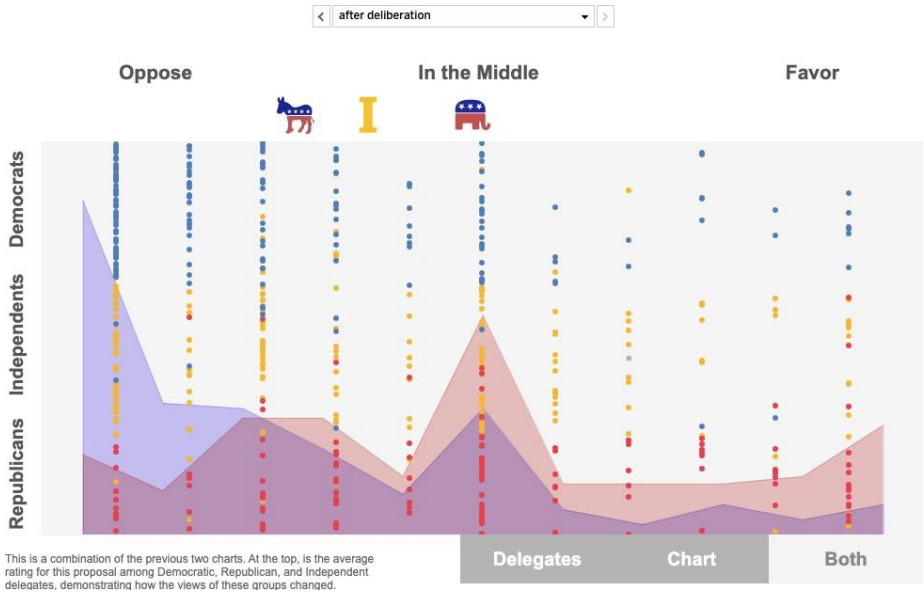

Figure 11: Distribution of views after open discussions

Conclusion

The idea of America in a room is what I have taken to a technology construct to address biases, via data driven approach. Instead of the people in a room, I intend to bring anonymous users in a debate room, discuss their views, like, comment, share, and more of their views and facts online, thereby I bring all their activities data to a Trust worth and People Chain model. I have built a very basic level user gamification that tracks a user, and eventually scores their activities over time leading to an elevated user whose actions are seen with respect and this is where DRI will be driving the online behavior.

Works Cited

Verint Systems, 2017. *Business Wire.* [Online]
Available at: https://www.businesswire.com/news/home/20170307005123/en/Global-Study-Ten-Consumers-Concerned-Data-Security
[Accessed 23 December 2020].
CDD, S.,2019. *America in One Room.* [Online]
Available at: https://cdd.stanford.edu/2019/america-in-one-room-results/#overall-results [Accessed 6 January 2021].
Cleveland, W. S., 2019. *Wiki books.*
[Online] Available at:

https://en.wikibooks.org/wiki/Data_Science:_An_Introduction/A_History_of_Data_Science#:~:text=The%20term%20%22Data%20Science%22%20was,is%20attributed%20to%20William%20S.
[Accessed 25 December 2020].

Davey Alba, A. S., 2019. *New York Times.* [Online]
Available at: https://www.nytimes.com/2019/09/26/technology/government-disinformation-cyber-troops.html [Accessed 20 December 2020].
Editorial, E., 2020. *E&T.* [Online]
Available at: https://eandt.theiet.org/content/articles/2020/05/facebook-did-not-act-on-own-evidence-of-algorithm-driven-extremism/#:~:text=According%20to%20a%20Wall%20Street,were%20stoking%20extremism%20and%20division.&text=He%20is%20thought%20to%20be,to%20fact%2Dchec [Accessed 20 December 2020].
Elliott, T., 2018. *Digital list Mag.* [Online]
Available at: https://www.digitalistmag.com/future-of-work/2018/02/05/data-is-powerful-weapon-use-it- responsibly-tell-people-05826064/
[Accessed 23 December 2020].
Eryarsoy, S. E. S. a. E., 2015. Generating Digital Reputation Index: A Case Study. *Procedia - Social and Behavioral Sciences,* Volume 195, pp. 1074-1080.
European Commission, 2018. *EC EUROPA EU.* [Online]
Available at: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-does-general-data-protection- regulation-gdpr-govern_en
[Accessed 23 December 2020].
Gloria Phillips-Wren, D. J. P. a. M. M., 2019. Cognitive Bias, Decision Styles and Risk Attitudes in Decision Making and DSS. *Journal of Decision Systems,* 28(2), pp. 63-66.
Gloria Phillips-Wren, D. J. P. a. M. M., 2019. Cognitive bias, decision styles, and risk attitudes in decision making and DSS. *Journal of Decision Systems,* 28(2), pp. 63-66.
Holly B. Shakya, N. A. C., 2017. Association of Face book Use With Compromised Well-Being: A Longitudinal Study. *American Journal of Epidemiology,* Volume 185(Issue 3), p. Pages 203–211.
Jack B. Soll, K. L. M. a. J. W. P., 2015. *Outsmart Your Own Biases.* [Online]
Available at: https://hbr.org/2015/05/outsmart-your-own-biases [Accessed 6 January 2020].
Rollins, J., 2015. *IBM.* [Online]
Available at: https://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science [Accessed 21 December 2020].
The Social Dilemma. 2020. [Film] Directed by Jeff Orlowski. United States: Exposure Labs, Argent Pictures, The Space Program distributed by Netflix.