# Evils of Social Media: Case Study of the Blue Whale Challenge

**Shaan Chopra**
IIIT-Delhi
India
shaan15090@iiitd.ac.in

**Abhinav Khattar**
IIIT-Delhi
India
abhinav15120@iiitd.ac.in

**Karan Dabas**
IIIT-Delhi
India
karan15141@iiitd.ac.in

**Kshitij Gupta**
IIIT-Delhi
India
kshitij15048@iiitd.ac.in

**Ponnurangam Kumaraguru**
IIIT-Delhi
India
pk@iiitd.ac.in

## ABSTRACT

The Blue Whale Challenge is a deadly challenge propagating on online social media and has claimed multiple lives across the globe [5]. This challenge requires the person to indulge in a series of self-mutilating tasks for a duration of 50 days and ultimately commit suicide. The so-called "administrators" or "curators" of the challenge contact users - who express their willingness to take part on social networking websites - via direct messages. We conducted a study to understand the spread of the challenge on social media websites such as VKontakte, Twitter, and Instagram, identify different types of users involved in the challenge, study their demographics, and identify distinguishing features between the users involved in the challenge and those who are not. Through this position paper, we throw some light upon dangerous social media challenges such as the Blue Whale Challenge which lure, engage, and victimize a spectrum of people. We express our interest in studying the harmful effects of technology and social media and elucidate our positionality with respect to the same.

## INTRODUCTION

The Blue Whale Challenge has witnessed numerous cases (including suicide attempts and deaths) around the world [5], with the maximum number of reported cases in Russia, the birthplace of the challenge. This draws attention to the widespread effects of the challenge spread through social networking websites. The Blue Whale Challenge consists of a series of deadly tasks that one must complete within a duration of 50 days [1] with the final challenge being to commit suicide. These tasks, which include waking up at odd hours, self-mutilation, listening to psychedelic music, and watching scary videos among others, are structured in such a way so as to brainwash the participants [6]. "Curators" or "administrators" [7] of the challenge contact users who express their willingness to take part in the challenge by posting on social media using blue whale-related hashtags such as #i_am_whale, #curatorfindme, #wakemeupat420, and #f57. These posts are often accompanied by images of whales as shown in Figure 1.

We see that instead of the "administrators" or "curators", it's usually the other users, who are interested in taking up the challenge, who try to contact the curators and become visible to them. This heightens the difficulty of curbing the impact of the challenge as banning a few accounts no longer remains an effective solution.

In this position paper, we elucidate the findings of our study and throw light upon evils of such challenges prevailing on social networking websites.

## OUR WORK

In [5], we conducted preliminary network, content, and user analysis to gauge how the challenge has spread on social media. This study was conducted across 3 social networks: VKontakte, Twitter, and Instagram. We observed no follower-following link between users posting about the challenge on Instagram, whereas on VKontakte, though sparse, a friends network existed among users posting about the challenge. Moreover, blue whale-related users tended to reveal personally identifiable information like phone number and email addresses so that possible "curators" or "administrators" of the challenge could contact them. This is a major privacy concern. Images of users performing self-mutilating tasks were, as can be seen in Figure 2, found in numerous posts as were a few links to psychedelic music
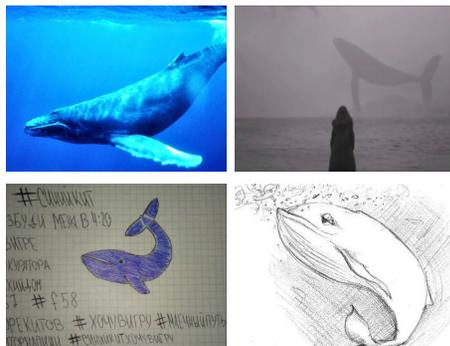


**Figure 1: Images found in posts on VKontakte relating to the Blue Whale Challenge.**
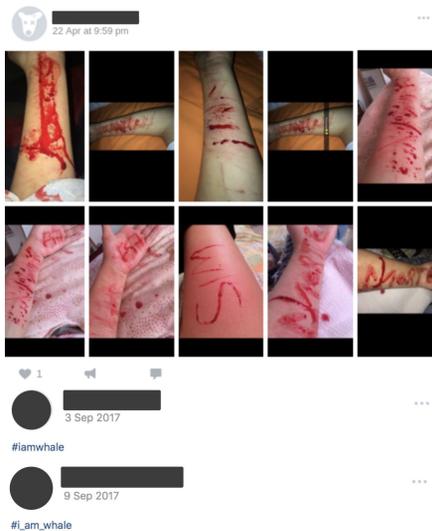
and scary videos. We also witnessed that a major proportion of blue whale-related posts on Twitter were made in Indian languages such as Hindi and Tamil, whereas Welsh and Romanian, Moldavan languages were used in around 35% of the posts on VKontakte; this showed the involvement of the users from different linguistic backgrounds in the challenge. From this preliminary study, we were able to glean information about user behavior with respect to this first-of-its-kind challenge that spread across the world via popular social media.

In a subsequent study, we conducted an in-depth analysis on VKontakte and compared features of blue whale and non-blue whale-related posts and users. For this we collected a blue whale positive, that is users who post about the challenge using the relevant hashtags, dataset and a blue whale negative, users who do not post about the challenge, dataset. We used post features such as n-grams and images and user features such as user location, gender, age, profile picture, and user description/status for the comparison. One interesting finding was that almost 50% of the users from the blue whale positive dataset fell under the age of 20. Further details of the age and gender demographic analysis of the users from the blue whale positive dataset can be seen in Table 1.

Thus, we analyzed and found differences between post and user characteristics of blue whale positive and blue whale negative datasets. These features, we believe, can be used to build a classifier to detect potential victims who are likely to be involved in the challenge. While analyzing textual features, code-mixing in posts is a major hurdle we faced. We further wish to find a way to effectively exploit textual features present in different languages and combine them with the aforementioned features.

## LIMITATIONS

Due to moderation being done by social networking websites, we faced a lack of data. A number of user accounts were either banned or deactivated and a lot of posts were either deleted by the user himself/herself or removed by the social network. VKontakte was temporarily banned in India so that too posed a problem during our data collection process [2]. Our data was also restricted by the API limits of social networking websites. Another issue we faced was the presence of code-mixed posts, especially on VKontakte where a majority of the posts are not in English; we were unable to generate n-grams for content present in other languages such as Russian.

## PRIVACY CONSIDERATIONS

Since this research focuses on analyzing user profiles and posts, it leverages certain data about users that might be considered sensitive, thus bringing privacy considerations into the picture. In our datasets, we used only publicly available data on VKontakte, Twitter, and Instagram and did not retain any personally identifiable information about users. The dataset of the number of blue whale-related cases as proposed in [5] was collated using publicly available information from news sources and no new information about the victims was dug out.



Figure 2: Some example posts from the collected data posted on VKontakte - containing only hashtags or images and no text.

## STUDYING EVILS OF SOCIAL MEDIA CHALLENGES

The research on ill-effects of gamification [3, 4] and dangerous online social media challenges such as the Blue Whale Challenge [6] is rather circumscribed. It would be an interesting and challenging task to study the harmful effects of game-based challenges which affect populations across the globe and design possible interventions to reduce their impact. Studying such a social media evil might bring out visible patterns among users who are spread throughout the world, separated through different cultural contexts but still engaging in and getting victimized by similar challenges.

## SELF-DISCLOSURE

Being of Indian origin, the authors' background and personal experiences of being raised in India have shaped their motivation and experiences of working in the field of HCI and social computing. The authors have conducted research in HCI and social networks across multiple domains. Also familiar with ICT and social media usage in India, the authors have an understanding of the "curiosity" driven behavior of users in light of such social media challenges.

## BENEFIT FROM THE WORKSHOP

Attending this workshop and engaging in a healthy discussion would help us better understand the vices of technology and social media which is widely used and accepted in today's world. We look forward to interacting with researchers in the domain and hearing out their experiences of working with or witnessing the evils of technology. Moving forward with the learnings from this workshop, we would like to be able to spot and provide interventions to possibly curb harmful effects of technology in our domain of study and possibly design solutions for our identified problem. For four out of the five authors who are undergraduate students and are yet to get a complete taste of the domain, this workshop would be an unmissable opportunity.

## AUTHORS

**Shaan Chopra, Abhinav Khattar, Karan Dabas, and Kshitij Gupta** are fourth-year undergraduate students at IIIT Delhi. Though still exploring, their research interests lie broadly in the domain of HCI and social computing. They are currently working on multiple projects including analysis of the social media spread of the Blue Whale Challenge. **Ponnurangam Kumaraguru** is an associate professor at IIIT Delhi where he heads the Precog research group.

**Table 1: Gender and age of users present in blue whale positive dataset**

| Age/Gender | 10-20 yrs | >20 yrs | Unknown | TOTAL |
|---|---|---|---|---|
| Male | 63 | 65 | 170 | 298 |
| Female | 64 | 41 | 137 | 242 |
| Unknown | 16 | 41 | 0 | 57 |
| TOTAL | 143 | 147 | 307 | 597 |

## REFERENCES

[1] 2017. What are the exact 50 challenges in the "blue whale challenge"? *Reddit* (March 2017). https://www.reddit.com/r/morbidquestions/comments/5xsnpq/what_are_the_exact_50_challenges_in_the_blue/

[2] Kim Arora. 2017. Russian social network VKontakte temporarily blocked in India for Blue Whale threat. https://timesofindia.indiatimes.com/.

[3] Kim Bohyun. 2015. *Understanding Gamification*. American Library Association, Chicago.

[4] Sami Hyrynsalmi, Jouni Smed, and Kai Kimppa. 2017. The Dark Side of Gamification: How We Should Stop Worrying and Study also the Negative Impacts of Bringing Game Design Elements to Everywhere. In *1st International GamiFIN Conference*. Pori, Finland.

[5] Abhinav Khattar, Karan Dabas, Kshitij Gupta, Shaan Chopra, and Ponnurangam Kumaraguru. 2018. *White or Blue, the Whale gets its Vengeance: A Social Media Analysis of the Blue Whale Challenge*. Technical Report.

[6] Ashutosh Kumar, Sada Nand Pandey, Vikas Pareek, Muneeb Faiq, Nazar Khan, and Vivek Sharma. 2017. Psychobiological determinants of 'Blue Whale Suicide Challenge' victimization: A proposition for the agency mediated mental health risk in new media age. (Sept. 2017). https://doi.org/10.17605/OSF.IO/8XH92

[7] Aisling Moloney. 2017. What is the Blue Whale Challenge? *Metro.co.uk* (2017). http://metro.co.uk/2017/08/31/what-is-the-blue-whale-challenge-6893410/