

**Meijin Samvad aka Expert conversation**

**WEBINAR TRANSCRIPT**

**Webinar Title:**

**You have been (Mis)informed: Technology & Societal Responses to Communication Misperception & Bias**

**Date: June 16, 2023**

**Speakers: 1. Dr. Dhavan Shah**

**2. Dr. Srijan Kumar.**

**Moderator: Mr. Saideep Rathnam**

**ABOUT THE SPEAKERS:**

**Dhavan V. Shah** is the Jack M. McLeod Professor of Communication Research and Louis A. & Mary E. Maier-Bascom Professor at the University of Wisconsin-Madison, where he is Director of the Mass Communication Research Centre (MCRC) and Scientific Director in the Centre for Health Enhancement System Studies (CHESS). Housed in the School of Journalism and Mass Communication with affiliated appointments in Industrial and Systems Engineering, Marketing, and Political Science, his research focuses on the influence of electronic and digital media on social judgments, civic engagement, and health support.

He has developed three major lines of inquiry, with his most recent work extending insights and techniques from his prior examinations of the influence of message construction and online interactions on political life to the development and deployment of digital technologies for individual and community health. These programs of research center on (1) the influence of message construction and processing on social evaluations, judgments, and behaviors; (2) the capacity of mass and interpersonal communication, especially in online communities, to shape civic engagement, political participation, and social trust; and (3) the effects of computer-mediated interactions, especially online support expression, on the management of cancer, aging, and addiction.

**Dr. Srijan Kumar** is an Assistant Professor at the College of Computing at Georgia Institute of Technology. His research expertise lies in developing AI, applied machine learning, and data mining methods. He builds graphs, content (NLP, multimodal), and adversarial learning methods while utilizing terabytes of data from multiple online platforms spanning multiple modalities and languages. He innovates scalable and efficient methods for online safety by detecting and mitigating malicious actors (e.g., ban evaders, sock puppets, coordinated campaigns, fraudsters) and dangerous content (e.g., misinformation, hate speech, fake reviews). At the same time, He develops methods to improve the security and safety of AI methods.

Dr. Kumar have been passionate about the research topics that comprehensively study some of the biggest threats to web safety and integrity from complementary angles:

- **AI for Security:** Multi-modal, multi-lingual, and multi-platform detection methods of harmful content and users utilizing terabytes of graph and content data
- **Security of AI:** Measuring and enhancing the trustworthiness and robustness of AI models
- **Mitigation Solutions:** Developing AI tools for end users (moderators, fact checkers, etc.) to combat cyber harms

#### **ABOUT THE MODERATOR:**

Saideep Rathnam is the Chief Operating Officer of Mizuho India Japan Study Centre, bringing a wealth of 47 years of industry and academic experience to the Centre. An alum of IIM Bangalore, from Hindustan Aeronautics Ltd. to British Aerospace, UK he has spent over 2 decades in the aeronautics industry and over 18 years in the automotive sector in various capacities including president of manufacturing excellence at Anand Automotive Ltd. He is also a Certified Chartered Management Accountant [CMA], UK. He wears many hats and has chaired Anand University, helping companies in the fields of management of change and innovation. Recently, he drives the Visionary Leaders for Manufacturing (VLFM) program as a Senior Advisory Committee Member of CII.

#### **1. OVERVIEW OF THE TOPIC BY THE MODERATOR**

There has been rampant dissemination of digital content in this highly connected society. There is a need to understand various factors that affect the spread of misinformation or bias in our digital ecosystem. The role of social media in shaping and amplifying these issues is crucial.

#### **2. PRESENTATION BY DHAVAN SHAH**

Dr. Shah started his presentation with a famous quote "A lie can travel half world till truth puts on its shoes." Interestingly this quote about misinformation is itself misinformed to be said by many famous personalities. He said falsehood is more likely to be retweeted 70% than the truth. This is an "Information crisis". He gave examples of such rumours regarding the Covid pandemic and US elections.

"Course correct " is a system that is developed by various scholars around the world that will help to keep a check on the spread of misinformation by journals. He explained how the process of vigilance initiates, and topics are selected either by journalists, staff monitoring social dynamics, or monitoring experts. Content retrieval is done through APIs and Synthesis i.e. social media monitoring system. Through this process, misinformation claims are identified. This information is analyzed whether it is high credibility information or high bias

URLs. These biases are seen as potential sensors. Ambiguity is avoided while focusing on scientific consensus.

The topics generally are settled ones rather than controversial or emerging ones. The most important topics that are widely considered after Covid and elections are financial. Much more topics that are the new generation's concern. The next step is the detection of such information. Real-time intervention is the most efficient.

Intervention can be of two types Pilot A/B testing i.e. survey update or Field A/B testing i.e. platform testing. Difficult faced during this process is tracking the scalability versus focusing on individualistic misbelief. "Observed correction" is the model developed by Dr.Shah and his colleagues that has gathered evidence that when correction of misinformation is observed by someone, that person is likely to share such correction. It is proved to be effective in most of the topics. The goal is not to reduce the content by cutting it down. The goal is to counter misinformation with corrective information for societal benefit.

### **3. PRESENTATION BY SRIJAN KUMAR**

"Bad actors, fraudsters, and misinformation polarization threaten the safety, equity, and well-being of users on platforms online. This is the biggest threat to democracy, science, and the whole society." It is important to note that people nowadays receive more information from their friends and family online than traditional news organizations. The scale of fake news, reviews, and misinformation is huge.

Generative AIs / Chat GPT can be easily misused to create fake news and misinformation. There are many victims of such misinformation that need to be curbed.

AI for internet security: Experts team has not only developed a detection process but successfully deployed it in various platforms like Flipkart, which is India's one of most popular e-commerce websites. This tool has helped to identify and remove fake reviews and reviewers.

Mr. Srijan and his team also worked on a multilingual model as 74% of the population is not having English as their first or second language. They found the disparity between performance based on languages. Investigation showed that this disparity is due to huge differences of the amount of pre-training data. Collaboration with other experts is working not only successfully in dealing with language but also other issues.

Mr. Srijan further emphasized upon detection part of 'Course correct' which was earlier introduced to us by Dr.Dhavan. It is done on the basis of keywords used. Specifically, which are used? where they are used? The region in which they are used or consumed, etc. This model can be very efficient during elections, health care, natural disasters, etc. Experts have also developed data-generative models that are based on journalistic and fact-checked data. Which will write counter-misinformation responses. This model is empowering consumers and professionals as well.

#### 4. QUESTION AND ANSWERS

**Moderator:** A source check of news in the social media era is necessary. How can we develop a robust mechanism to counter fake news?

**Dr.Dhavan Shah:** We are focusing on settled information but to answer that evidence-based information and verification is required. The real challenge is to decide what is verified misinformation. Firstly, the detection of misinformation and finding the network in which it is traveling. Targeting such information, reducing the confidence in those networks that are most likely to share such misinformation.

**Moderator:** The amount of misinformation generation is high. What can be done to curb that?

**Dr.Dhavan Shah:** it is not always the case that misinformation comes quickly and disappears. Sometimes, it comes in a pattern and spreads for months by the same claimants. Secondly, intervention cannot be done before it is spread to a certain degree. Long-term misinformation correction is the goal of this system.

**Dr.Srijan Kumar:** Echoing with Dr.Shah, One aspect is the transfer of misinformation from one part of the globe to another where such data is translated into different languages. In such cases, each country's fact-checking agencies work differently on the same data. However, our AI system can pull out such data and fact-check more robustly being time efficient at the same time. It provides interfaces that can help professional fact-checkers. As they will need sufficient time to maintain their credibility without which they will not decide upon any misinformation.

**Moderator:** What role can the common educated class play to lessen the spread of misinformation?

**Dr.Srijan Kumar:** Media literacy rate is much low. Simple campaigns like WhatsApp Messenger initiated that do not share the message without reading it can help a lot. A similar effort was made by Twitter. People are likely to stop believing the misinformation when they see the misinformation being corrected.

**Dr.Dhavan Shah:** It is crucial to know where is the information coming from. Media houses or journalists have some agenda or not is to be checked. Moreover, the platforms must have citizen-based fact-checking rather than internal mechanisms alone.

**Moderator:** It is difficult to change opinions of a certain part of society that is influenced by some personality/ ideology. Is that a challenge?

**Dr.Dhavan Shah:** Such a group of people is very difficult to change but the real problem here is not the personality/ ideology but the influenced people who are being observed by a network.

**Dr.Srijan Kumar:** We are not expected to work against any ideology or any specific group of society at all. We are unbiased in our work.

**Moderator:** Misinformation claims are questionable. Is it objective in nature? Is the truth of data absolute?

**Dr.Dhavan Shah:** Our job is not to raise objections against one group or defend one group. We intervene only when we have scientific consensus on misinformation.

**Dr.Srijan Kumar:** While building machine learning classifiers, we believe only on international fact-checking organizations that provide a ton of evidence about the credibility or non-credibility of any information.

## 5. CONCLUSION

Moderator thanked both panelists for taking their precious time to engage with our audience and making us understand such crucial topics with technical nuances. He also invited them to collaborate with IIMB and its alumni on such important research and their guidance.

### 名人サムバド

#### ウェビナーの文字起こし

#### 発表題目

「あなたは（誤）解されています！コミュニケーションの誤認と偏見を管理するための技術的・社会的アプローチ」

日付: 2023年6月21日

講演者: 1. ダヴァン・シャー教授、博士

2. スリジャン・クマール教授

#### 講演者について:

**ダヴァン・シャー博士:**ダヴァン・シャー教授はウィスコンシン大学マディソン校のジャック・M・マクラウド教授、ルイス・A・とメアリー・E・メイア・バスコム教授です。マス・コミュニケーション研究センター（MCRC）所長、健康増進システム研究センター（CHESS）科学部長です。ジャーナリズム・マスコミュニケーション学部に所属し、産業・システム工学、マーケティング、政治学を兼任しており、電子・デジタルメディアが社会的判断、市民参加、健康支援に及ぼす影響に焦点を当てた研究を行っています。

最近の研究では、政治生活におけるメッセージの構成とオンライン上の相互作用の影響に関する先行研究から、個人とコミュニティの健康のためのデジタル技術の開発と展開へと、洞察と技術を拡張しています。これらの研究プログラムの中心は、(1)社会的評価、判断、行動に対するメッセージの構築と処理の影響、(2)市民参加、

政治参加、社会的信頼を形成するマス・コミュニケーションと対人コミュニケーション、特にオンライン・コミュニティの能力、(3)癌、老化、依存症の管理に対するコンピュータを介した相互作用、特にオンライン支援表現の効果、です。

**スリジャン・クマール教授**は情報工学部、コンピューティング大学で助教授です。

専門は AI、応用機械学習、データマイニング手法の開発。グラフ、コンテンツ

(NLP、マルチモーダル)、敵対的学習法を構築し、複数のモダリティと言語にまたがる複数のオンラインプラットフォームから得たテラバイト級のデータを活用しています。彼は、悪意のある行為者(禁止回避者、操り人形、協調キャンペーン、詐欺師など)や危険なコンテンツ(誤情報、ヘイトスピーチ、偽レビューなど)を検出・軽減することで、オンライン安全のためのスケーラブルで効率的な手法を革新している。同時に、AI 手法のセキュリティと安全性を向上させる手法の開発も行っています。

クマールさんは、ウェブの安全性と完全性に対する最大の脅威のいくつかを補完的な角度から包括的に研究する研究テーマに情熱を注いでいます：

- **セキュリティ用 AI**：テラバイトのグラフとコンテンツ データを利用した、有害なコンテンツとユーザーのマルチモーダル、多言語、マルチプラットフォームの検出方法。
- **AI の安全性**：AI モデルの信頼性と堅牢性の測定と強化。
- **緩和策**：サイバー被害に対抗するためのエンドユーザー（モデレーター、ファクトチェッカーなど）向け AI ツールの開発。

#### 司会者について:

サイディー・ラスナムさんは、みずほ印日研究センターの最高執行責任者であり、47年にわたる豊富な産業界と学術界の経験を生かして、同センターの運営に携わっています。IIM バンガロールの卒業生であり、ヒンドスタン航空から英国ブリティッシュ・エアロスペースまで、航空業界では20年以上、自動車業界では18年以上、アナンド・オートモーティブの製造エクセレンス社長など様々な立場で経験を積んできました。英国公認管理会計士[CMA]でもあります。アナンド大学の学長も務め、変化と革新のマネジメントの分野で企業を支援しています。最近では、CIIの

上級諮問委員会メンバーとして、VLFM「ビジョナリー・ラーニング・コミュニティ」プログラムを推進しています。

## 6. 司会者によるトピックの概要説明

この高度に接続された社会では、デジタルコンテンツの拡散が横行しています。私たちのデジタル・エコシステムにおいて、誤った情報や偏見の拡散に影響するさまざまな要因を理解する必要があります。このような問題の形成と増幅におけるソーシャルメディアの役割は極めて重要です。

## 7. ダヴァン・シャーのプレゼンテーション

ダヴァン博士は、「真実がその靴を履くまで、嘘は世界の半分を旅することができる」という有名な言葉でプレゼンテーションを始めました。興味深いことに、誤報に関するこの引用は、それ自体が多く有名な人物によって言われた誤報であります。彼は、虚偽は真実よりも 70% リツイートされる可能性が高いと言いました。これは「情報の危機」です。コビトパンデミックやアメリカの選挙に関するそのような噂の例を挙げました。

「コース修正」とは、世界中のさまざまな学者が開発したシステムで、ジャーナルによる誤った情報の拡散をチェックするのに役立ちます。彼は、警戒のプロセスがどのように開始され、トピックがジャーナリスト、ソーシャルダイナミクスを監視するスタッフ、または監視の専門家によって選択されるかを説明しました。コンテンツの取得は、API と合成、つまりソーシャルメディア モニタリング システムを通じて行われます。このプロセスを通じて、誤った情報の主張が特定されます。この情報は、信頼性の高い情報であるか、偏りの高い URL であるかが分析されます。これらのバイアスは潜在的なセンサーと見なされます。科学的な合意に重点を置きながら、曖昧さは回避されます。

介入には、パイロット A/B テスト「調査更新」とフィールド A/B テスト「プラットフォームテスト」の2つのタイプがあります。このプロセス中に直面する困難は、スケーラビリティを追跡することと、個人主義的な誤解に焦点を当てることです。

「観察された修正」は、ダヴァン博士と彼の同僚によって開発されたモデルで、誤った情報の修正が誰かによって観察された場合、その人はその修正を共有する可能性が高いという証拠を収集しました。ほとんどのトピックで効果的であることが証

明されています。目的は、コンテンツを削減して内容を減らすことではありません。目標は、社会の利益のために、正しい情報で誤った情報に対抗することです。

## 8. スリジャン・クマールのプレゼンテーション

悪質な行為者、詐欺師、誤った情報の偏向は、オンライン上のプラットフォームにおけるユーザーの安全性、公平性、幸福を脅かしています。これは民主主義、科学、そして社会全体に対する最大の脅威です。現在、人々が伝統的な報道機関よりも、ネット上の友人や家族から多くの情報を受け取っていることに注意することが重要です。フェイクニュース、レビュー、誤報の規模は非常に大きいです。

Generative AI ・ Chat GPT は、フェイクニュースや誤った情報を作り出すために簡単に悪用される可能性があります。そのような誤報の被害者は多く、抑制する必要があります。

インターネットセキュリティ用の AI： 専門家チームは、検知プロセスを開発しただけでなく、インドで最も人気のある電子商取引サイトの一つである Flipkart のような様々なプラットフォームへの導入に成功しました。このツールは、偽のレビューや査読者を特定して削除するのに役立ちます。

スリジャンさんと彼のチームは、人口の 74% が英語を第一言語または第二言語としていないことから、多言語モデルにも取り組みました。彼らは、言語によるパフォーマンスの格差を発見しました。調査の結果、この格差は事前学習データ量の大きな違いによるものであることがわかりました。他の専門家とのコラボレーションは、言語だけでなく、他の問題にもうまく作用しています。

スリジャンさんはさらに、ダヴァンさんから以前に紹介された「コース修正」の検出部分について強調しました。これは使用されているキーワードに基づいて行われます。具体的には、どれが使われているのか、どこで使われているのでしょうか。それらが使われたり消費されたりする地域などです。このモデルは、選挙、医療、自然災害などの際に非常に効率的です。専門家たちは、ジャーナリスティックなデータやファクトチェックされたデータに基づくデータ生成モデルも開発しています。これは、誤報に対する反論を書くものである。このモデルは消費者や専門家にも力を与えています。



## 9. 質疑応答

**司会者:** ソーシャルメディア時代のニュースはソースチェックが必要です。フェイクニュースに対抗する堅牢なメカニズムを開発するにはどうすれば良いでしょうか。

**ダヴァン:** 私たちは定説的な情報に注目しているが、それに答えるためには、証拠に基づいた情報と検証が必要です。本当の課題は、何が検証された誤情報なのかを決めることです。まず、誤報を検出し、その誤報が流れているネットワークを見つけること。そのような情報をターゲットとし、そのような誤情報を共有する可能性の高いネットワークの信頼性を低下させます。

**司会者:** 誤報の発生量が多いです。それを抑えるにはどうすれば良いでしょうか。

**ダヴァン:** 誤報はすぐに消えてしまうとは限りません。時にはパターン化され、同じ主張者によって何カ月にもわたって広まることもあります。第二に、ある程度広まる前に介入することはできません。長期的な誤報是正がこのシステムの目標です。

**スリジャン:** 一つの側面は、誤った情報が世界のある地域から別の地域へ伝わり、そのようなデータが異なる言語に翻訳されることです。このような場合、各国のファクトチェック機関は、同じデータに対して異なる作業を行います。しかし、私たちの AI システムは、そのようなデータを取り出し、より強固にファクトチェックすることができます。また、プロのファクトチェッカーを支援するインターフェースも提供しています。彼らは信頼性を維持するために十分な時間が必要であり、それがなければ誤報を判断することはできないからです。

**司会者:** 誤った情報の拡散を抑えるために、一般的な教育を受けた層はどのような役割を果たすことができるのでしょうか。

**スリジャン:** メディアリテラシーが低いです。WhatsApp メッセンジャーが始めたような、メッセージを読まずにシェアしないようにするシンプルなキャンペーンは、大きな助けになります。ツイッターも同様の取り組みを行っています。誤った情報が訂正されるのを見れば、人々は誤った情報を信じなくなるでしょう。

**ダヴァン:** 情報の出所を知ることは極めて重要です。メディア企業やジャーナリストが何らかの意図を持っているかどうかは、チェックする必要があります。さらに、プラットフォームは内部的なメカニズムだけでなく、市民ベースのファクトチェックを持たなければなりません。

**司会者:** ある種の個性やイデオロギーに影響された社会のある部分の意見を変えるのは難しいです。それは課題ですか。

**ダヴァン:** このような集団を変えるのは非常に難しいが、ここで本当に問題なのは、人格やイデオロギーではなく、ネットワークによって観察されている影響された人々です。

**スリジャン:** 私たちは、いかなるイデオロギーや社会の特定のグループに対して働くことはまったく期待されていません。私たちは不偏不党で仕事をしています。

**司会者:** 誤報の主張は疑わしいです。それは客観的なものなのですか。データの真偽は絶対なのですか。

**ダヴァン:** 私たちの仕事は、あるグループに対して異論を唱えたり、あるグループを擁護したりすることではありません。私たちが介入するのは、誤った情報について科学的なコンセンサスが得られたときだけです。

**スリジャン:** 機械学習分類器を構築する際、私たちは、あらゆる情報の信憑性や非信憑性について多くの証拠を提供する国際的なファクトチェック機関のみを信頼する。

## 10. まとめ

司会者は、両パネリストが貴重な時間を割いて聴衆と対話し、このような重要なトピックを専門的なニュアンスも含めて理解させてくれたことに感謝しました。また、このような重要な研究やその指導について、IIMB やその卒業生と協力してほしいと呼びかけました。