

Mizuho India Japan Study Centre (MIJSC) & Centre for Software and Information Technology Management (CSITM)

WEBINAR TRANSCRIPT

Webinar Title:

“PROMISES & PERILS OF AI: MITIGATING BIAS, EXPLORING OPPORTUNITIES AND ACHIEVING FAIRNESS”

Date: 30th of April 2024

Speakers: Dr. Jaideep Srivastav, Dr. Sharath Chandra Guntuku, Dr. Nikhil Malik

Moderator: Saideep Rathnam

About the Speakers:

Dr. Jaideep Srivastav: *Dr. Jaideep Srivastava is a professor of Computer Science at the University of Minnesota. He joined the Department of Computer Science & Engineering in 1988. He currently serves as the Data Science Director of Undergraduate Studies. Dr. Srivastava is the co-founder and president of Ninja Metrics, Inc., a start-up company that focuses on analytics for social systems and social media. He spent three years at Qatar Computing Research Institute (QCRI), a flagship institute of Qatar Foundation, as Research Director and Chief Scientist. Additionally, he has spent significant time in the industry, including stints as the Chief Technology Officer for Persistent Systems, Director of Data Analytics for Yodlee, Inc., and as a Data Mining Architect for Amazon. He has received a Bachelor's Degree from the Indian Institute of Technology (IIT) Kanpur in India. He has done his Master's and Ph.D. from the University of California, Berkeley.*

Dr. Sharath Chandra Guntuku: *Dr. Sharath Chandra Guntuku is an Assistant Professor in the research-track in the Department of Computer and Information Science at the University of Pennsylvania and a Senior Fellow at the Leonard Davis Institute of Health Economics. His research focuses on building predictive models using user-generated data for and uncovering insight into health outcomes and psychological states of individuals and communities. This research aims to supplement clinical diagnoses and facilitate early and personalized interventions for improving treatments and well-being. His research has raised over \$ 4.5 million from the National Institutes of Health, Penn Global, the World Bank Group, and other agencies. His work has been covered by the Times of India, American*

Psychological Association, WIRED, Canadian Broadcasting Company, The Atlantic, US News, and other venues.

Dr. Nikhil Malik: *Dr. Nikhil Malik is an Assistant Professor of Marketing at the Marshall School of Business, University of Southern California. Dr. Nikhil studies the economic implications of Machine Learning (ML) and other disruptive technologies. His recent work touches on Zillow's ML-based pricing for homes and Bitcoin's pricing for payments. His research highlights issues of pricing, adoption, bias, and fairness. Dr. Nikhil's research has been supported in the past by PNC Bank, Ripple (payments network), and Henry J. Gailliot Presidential fellowships. He serves on the Editorial Review Board of Marketing Science and received the Distinguished Service Award for peer review work at Management Science. Dr. Nikhil received his Ph.D. and Master's from Carnegie Mellon University and his Bachelor's Degree from IIT Delhi. He has also spent time designing Big-Data Technology for Goldman Sachs. Given this background, Nikhil brings expertise at the intersection of AI and Marketing.*

About the Moderator:

Saideep Rathnam: *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. He drives the Visionary Leaders for Manufacturing (VLFM) program as a Senior Advisory Committee Member of CII.*

INTRODUCTION

Dr. Sharath begins by highlighting the importance of mental health and the potential of AI to provide insights in this area. He polls the audience to gauge how many people experience loneliness. He emphasizes that loneliness can be felt even when surrounded by people. Dr. Sharath acknowledges the challenges associated with mental health, including scepticism about its legitimacy and the negative side effects of medication. Dr. Malik shifts the focus to the financial markets and housing markets, examining the potential for bias in AI algorithms used by platforms like Zillow. He argues that these algorithms are designed to maximize profits for the platform, not necessarily to ensure fair and accurate pricing for consumers. Prof. Jaideep facilitates a discussion about the issues raised by

the panellists. He highlights the misaligned incentives between platforms and consumers, and the potential dangers of bias in AI algorithms, particularly in areas like healthcare and finance.

The panellists highlighted the immense potential of AI to revolutionize various sectors. From tackling complex challenges to automating tasks and making data-driven predictions, AI offers the potential to reshape processes, enhance efficiency, and improve lives around the world.

Setting the context:

Artificial intelligence (AI) has emerged as a powerful force, rapidly transforming numerous aspects of our world. Akin to the rise of the steam engine or the internet, AI is ushering in a new era of intelligent machines capable of tasks once thought to be the exclusive domain of human cognition. From self-driving cars navigating bustling city streets to complex algorithms making stock market predictions, AI is making its presence felt across industries.

This panel discussion, held on April 30, 2024, by two Centres of Excellence at IIM Bangalore - Mizuho India Japan Study Centre (MIJSC) and the Centre for Software and Technology Management (CSITM), aimed to delve into the multifaceted world of AI. Much like a scientist peering into a kaleidoscope, we sought to examine AI from various angles, exploring both its immense potential and the inherent challenges it presents.

AI: A Double-Edged Sword

The panellists highlighted the immense potential of AI to revolutionize various sectors. From tackling complex challenges to automating tasks and making data-driven predictions, AI offers the potential to reshape processes, enhance efficiency, and improve lives around the world. Dr. Srivastava emphasized AI's ability to tackle complex challenges by processing massive datasets, automating tasks, and making data-driven predictions. This has the potential to streamline processes, enhance efficiency, and improve the quality of life across the globe.

Dr. Nikhil shed light on the economic benefits of AI. He discussed how AI-powered applications can optimize processes, leading to increased productivity and economic growth. "Because of the feedback loop that's present with AI predictions in the housing market, the home prices will become more erratic both in positive and negative directions over a long run," Prof. Nikhil Malik explained.

However, the discussion also acknowledged the inherent dangers associated with AI. Dr. Guntuku raised concerns about bias in algorithms, which can lead to discriminatory outcomes. He emphasized the importance of developing fair and unbiased AI models to ensure ethical implementation.

Dr. Malik echoed these concerns, highlighting the potential for AI to exacerbate existing social inequalities. He discussed the challenges of bias in areas like algorithmic pricing and loan approvals. Issues such as algorithmic bias, privacy concerns, data security risks, and potential job displacement were identified as key ethical challenges demanding immediate attention.

One area of particular concern is the use of AI in domains that directly affect individual well-being, such as mental health and housing. The potential for AI algorithms to perpetuate or amplify existing biases, leading to unfair treatment or discrimination, has raised alarms among experts and policymakers alike. Additionally, the opaque nature of many AI systems and the lack of transparency surrounding their decision-making processes have fuelled concerns about accountability and trust.

"Digital data x AI is a great low-cost resource to measure mental health in real time, especially in under-resourced communities. Heterogeneity is real! AI models need to account for cultural differences and correct for bias. If anyone says they are feeling lonely, perhaps they are; being there helps!"

Furthermore, the increasing reliance on AI systems in critical decision-making processes has highlighted the need to align their objectives with societal values and ensure their outputs are fair, ethical, and beneficial to all stakeholders. The misalignment of incentives between AI developers, deployers, and end-users has the potential to create unintended consequences, eroding public trust and hindering the responsible adoption of these technologies.

Navigating the Complexity of AI

The panellists emphasized the need for a multi-pronged approach to harness the transformative power of AI while mitigating its risks. The discussion explored strategies for:

- **Reducing bias:** Techniques to identify and address bias in AI algorithms were explored.
- **Identifying opportunities:** The panellists discussed how AI can be leveraged across various sectors for positive impact.

- **Achieving fairness:** Strategies for ensuring fairness and responsible development of AI applications were addressed.

Dr. Srivastava suggested employing diverse datasets to train AI models, ensuring they accurately reflect the real world.

Dr. Guntuku proposed the use of fairness metrics to evaluate AI models during development, identifying and rectifying potential biases.

Dr. Malik emphasized the importance of regulations and ethical guidelines to ensure responsible AI development and deployment.

Options for recourse:

Speakers stressed the need for a multidisciplinary approach to navigate this complex landscape. The conversation then shifted towards solutions for mitigating bias and achieving fairness in AI development. Dr. Srivastava suggested employing diverse datasets to train AI models, ensuring they reflect the real world. Dr. Guntuku proposed fairness metrics to evaluate AI models during development, enabling the identification and rectification of potential biases. Dr. Malik stressed the importance of regulations and ethical guidelines for responsible AI development and deployment.

Building Fairness Metrics and Fostering Transparency

Dr. Guntuku introduced the concept of fairness metrics, which can be likened to a compass guiding us towards equitable AI development. These metrics would enable us to evaluate AI models during their development phase, identifying areas where bias may be creeping in. By proactively addressing these biases, we can ensure fairer outcomes from AI systems.

He envisioned AI algorithms that are not shrouded in secrecy, but rather designed with transparency in mind. This transparency would allow users to understand the rationale behind the model's decisions, fostering trust and enabling the identification of any biases that may be present.

A Multi-Faceted Approach: Collaboration, Education, and Regulation

The panellists underscored the importance of a multifaceted approach to address the challenges of AI bias. Dr. Malik stressed the need for collaboration between various stakeholders, including

researchers, policymakers, and industry leaders. By working together, we can develop comprehensive solutions that address the ethical considerations surrounding AI.

Recommendations to Policymakers:

1. Monitoring and Detection of "Feedback" and "Excess Fluctuations" in housing market prices.
2. Enforce A1 "Model Control" on Platforms offering A1.
 - Similar in spirit to "Model Control" instituted on financial asset pricing models after 2007-08 financial crisis.

Public education also emerged as a critical element in the fight against bias. Dr. Malik, echoing the need for media literacy emphasized in "Digital Dialects," proposed educational initiatives to raise awareness about AI and its potential pitfalls. An informed public is better equipped to advocate for responsible AI development and use.

Regulation was another crucial aspect highlighted during the discussion. Just as traffic laws ensure order on the roads, regulations can provide a framework for ethical AI development. Dr. Misra advocated for adaptable regulations that can keep pace with the rapid evolution of AI technology.

The Road Ahead: A Call for Continuous Vigilance

The solutions outlined by the panel offer a roadmap for mitigating bias and ensuring fairness in AI. However, the fight against bias is an ongoing endeavour, demanding continuous vigilance. By adopting these solutions and fostering a culture of open dialogue, we can harness the power of AI for the benefit of all. As Albert Einstein aptly said, "Not everything that counts can be counted, and not everything that can be counted counts." The human element in AI development and oversight remains paramount in ensuring this technology serves humanity towards a brighter future.

A Call for Collaboration

The discussion concluded with a call for collaboration between various stakeholders, including academics, industry professionals, policymakers, ethicists, and risk assessment experts. The panellists emphasized the importance of:

- **Government Oversight and Regulation:** To ensure responsible AI development and implementation in critical sectors, panellists advocated for increased government oversight,

and the consideration of multiple independent pricing models or credit scoring systems to mitigate the risks associated with relying on a single source.

- **Transparency and Openness:** To foster scrutiny of AI algorithms and mitigate bias, it was proposed that platforms could be mandated to provide interpretable AI explanations alongside their predictions, even if the underlying models are not fully disclosed. The suggestion of governments providing open-source, non-profit pricing models as alternatives to commercial AI pricing platforms was raised, as these would not be driven by profit motives and could serve as a neutral benchmark.
- **Investment in Expertise:** Recognizing the challenges faced by regulators in auditing and understanding complex AI systems, panellists emphasized the need for equipping governments with the necessary expertise in AI and data analysis to effectively monitor and audit AI algorithms.

Q AND A

***Q:** How can Technologies be ethically integrated to address mental health disparities and also promote interventions while upholding the principles of the patient privacy and autonomy?*

Dr. Sharath Chandra Guntuku: "I think informed consent is very much required and it shouldn't be like the terms of service that we usually just scroll away and then click accept; it has to be communicated in a very plain and simple language". Dr. Guntuku emphasizes the importance of transparency surrounding patient privacy and autonomy. "There are also multiple stakeholders involved so we often think of AI driven mental healthcare as a tired triaging system, so AI cannot do the Diagnostics as it is not there yet. We definitely need the clinicians, but at the same time, AI could close the gap between how many people need treatment and how many people are actually being treated," he said, highlighting the potential of AI in bridging the mental healthcare gap to ensure effective addressing of mental health disparities. "AI could potentially play a role in recommending or giving a sort of a prediction for individuals who can then follow up with clinicians or their primary care provider".

***Q:** Dr. Malik, your discussion on bias in AI algorithms is concerning. How can we ensure fairness in areas like housing markets where AI is increasingly used?*

Dr. Nikhil Malik: "Algorithms used by platforms like Zillow can perpetuate existing biases in housing prices based on factors like race or socioeconomic status. There are several steps we can take to mitigate this," Dr. Malik said. "Firstly, data collection practices need to be scrutinized. If the training data for AI algorithms is biased, the resulting models will be biased as well. We need diverse and representative data sets to ensure fairness." He continued, "Secondly, we need to develop transparent and explainable AI models. This allows us to understand how algorithms arrive at decisions and identify potential biases. If an algorithm consistently undervalues homes in certain neighbourhoods, we need to know why." And at last, he concluded, "Finally, regulatory frameworks need to be established to hold platforms accountable for bias in their AI algorithms. Governments can play a vital role in ensuring fair and ethical use of AI in critical sectors like housing."

Q: *Can government offer an alternative which is more sort of open source?*

Dr. Nikhil Malik: *"Realistically for policy makers to create enough awareness that as a population we want to fund multi-billion-dollar projects creating a sort of like venture capital investment into the government creating such big AI, I find that might be challenging to pull off," Dr. Malik pointed out.*

Q: *What role can governments play in regulating AI?*

Governments can establish regulations that promote transparency, accountability, and fairness in AI development. These regulations should be adaptable to keep pace with the rapid evolution of AI technology," Dr. Srivastava admitted.

Q: *Professor Srivastava, you mentioned the importance of government oversight. However, wouldn't this stifle innovation in the field of AI?*

Prof. Jaideep Srivastav: *"Not necessarily. Regulation can actually foster responsible innovation. Clear guidelines can help developers focus on creating AI applications that are beneficial and ethical," Prof. Srivastav reassured the audience. "Additionally, government oversight can help address public concerns about AI. Transparency and accountability are crucial for building trust in this rapidly evolving technology. Of course, regulations need to be designed carefully to avoid hindering innovation. A collaborative approach involving industry, academia, and policymakers is essential to strike the right balance."*

CONCLUSION

In closing, the panel topic "Promises & Perils of AI: Mitigating Bias, Exploring Opportunities and Achieving Fairness" underscored the potential of AI for mental health applications, particularly in under-resourced communities, where digital data analysis by AI could offer a low-cost solution for real-time monitoring. However, concerns were raised regarding bias in AI models, emphasizing the need to account for cultural differences and actively address potential biases in training data. It became evident that the promises and perils of AI are inextricably linked, and navigating this landscape requires a delicate balance between innovation and responsible deployment. This sentiment aligns with the broader discussion on the importance of fairness and mitigating bias in AI development across various sectors.

The way forward lies in fostering collaboration among stakeholders, promoting transparency, and establishing robust governance frameworks. Governments must take proactive steps to invest in AI auditing capabilities, consider alternative pricing models, and mandate interpretable AI explanations. Moreover, multidisciplinary policymaking involving diverse expertise, is crucial to ensure that AI systems are developed and deployed with fairness, accountability, and the best interests of society at heart.

As AI continues to permeate every aspect of our lives, it is imperative that we remain vigilant, embrace innovation responsibly, and work collectively towards mitigating biases, exploring opportunities, and achieving fairness in the age of artificial intelligence.

みずほ印日研究センター（MIJSC）とソフトウェアおよび情報技術管理センター（CSITM）

ウェビナーの文字起こし

ウェビナー題名

「AIの約束と危険：偏見を緩和し、機会を探り、公正さを達成する」

日付：2024年4月30日

講演者：ジャイディーブ・スリヴァスタヴ博士、シャラート・チャンドラ・グントック博士、ニキール・マリク博士、司会者：サイディーブ・ラスナム

講演者について:

ジャイディーブ・スリヴァスタヴ博士: ミネソタ大学コンピューターサイエンス教授。1988年にコンピュータサイエンス&エンジニアリング学科に入学。現在、データサイエンス学部ディレクター。社会システムとソーシャルメディアの分析に特化した新興企業 **Ninja Metrics, Inc.**の共同設立者兼社長。カタール財団の主要機関であるカタール・コンピューティング・リサーチ・インスティテュート (QCRI) にリサーチ・ディレクター兼チーフ・サイエンティストとして3年間勤務。また、**Persistent Systems**社で最高技術責任者、**Yodlee**社でデータ分析ディレクター、**Amazon**社でデータマイニングアーキテクトを務めるなど、業界での経験も豊富。インドのインド工科大学 (IIT) カンパール校で学士号を取得。カリフォルニア大学バークレー校で修士号と博士号を取得。

シャラート・チャンドラ・グントウク博士: シャラート・チャンドラ・グントウク博士は、ペンシルバニア大学コンピュータ・情報科学科の研究課程で助教授を務め、レナード・デイビス医療経済研究所のシニアフェローでもある。研究テーマは、ユーザー生成データを用いた予測モデルの構築と、個人やコミュニティの健康アウトカムや心理状態に関する洞察の解明。この研究は、臨床診断を補完し、治療と幸福を改善するための早期かつ個別化された介入を促進することを目的としている。彼の研究は、米国国立衛生研究所、ペン・グローバル、世界銀行グループ、その他の機関から450万ドル以上を集めている。彼の研究は、**Times of India**、**American Psychological Association**、**WIRED**、**Canadian Broadcasting Company**、**The Atlantic**、**US News**などで取り上げられている。

ニキール・マリク博士: 南カリフォルニア大学マーシャル・スクール・オブ・ビジネス助教授。機械学習 (ML) やその他の破壊的技術の経済的影響を研究。最近の研究では、**Zillow**のMLベースの住宅価格設定やビットコインの決済価格設定について触れている。彼の研究は、価格設定、採用、バイアス、公平性の問題に焦点を当てている。ニキール博士の研究は、過去に**PNC**銀行、**リップル**社 (決済ネットワーク)、**ヘンリー・J・ガリオット**大統領フェローシップの支援を受けている。**Marketing Science**誌の編集査読委員を務め、**Management Science**誌の査読で特別功労賞を受賞。カーネギーメロン大学で博士号と修士号を、**IIT デリー**で学士号を取得。また、**ゴールドマン・サックス**でビッグデータ・テクノロジーの設計に携わった経験もある。このような経歴から、ニキールはAIとマーケティングの交差点で専門知識を発揮している。

司会者について

サイディーブ・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る。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た。英国公認管理会計士[CMA]でもある。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る。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る。

はじめに

シャラート博士はまず、メンタルヘルスの重要性と、この分野における AI の可能性を強調する。シャラス博士は、聴衆にアンケートを取り、孤独を経験する人がどれくらいいるのかを測る。彼は、孤独は人々に囲まれていても感じる事ができると強調する。シャラート博士はまず、メンタルヘルスの重要性と、この分野における AI の可能性を強調する。彼は聴衆にアンケートを取り、どれだけの人が孤独を経験しているかを測る。彼は、孤独は人に囲まれていても感じられるものだとして強調する。彼は、メンタルヘルスの正当性に対する懐疑的な見方や、薬物療法による悪影響など、メンタルヘルスにまつわる課題を認識している。マリク博士は、金融市場と住宅市場に焦点を移し、Zillow のようなプラットフォームが使用する AI アルゴリズムに偏りがある可能性を検証する。彼は、これらのアルゴリズムはプラットフォームの利益を最大化するように設計されており、必ずしも消費者にとって公正で正確な価格設定を保証するものではないと主張する。ジャイディーブ教授は、パネリストが提起した問題について議論を進める。彼は、プラットフォームと消費者の間のずれたインセンティブや、特にヘルスケアや金融のような分野における AI アルゴリズムのバイアスの潜在的な危険性を強調する。

パネリストたちは、さまざまな分野に革命をもたらす AI の計り知れない可能性を強調した。複雑な課題への取り組みから、タスクの自動化、データに基づく予測まで、AI は世界中のプロセスを再構築し、効率を高め、生活を改善する可能性を提供する。

文脈の設定

人工知能 (AI) は強力な力として台頭し、私たちの世界のさまざまな側面を急速に変化させている。蒸気機関やインターネットの台頭と同様に、AI は、かつては人間の認知の独占領域と考えられていたタスクをこなすインテリジェント・マシンの新時代の到来を告げている。にぎやかな街並みをナビゲートする自動運転車から、株式市場の予測を行う複雑なアルゴリズムに至るまで、AI はあらゆる産業でその存在感を示している。

2024 年 4 月 30 日、IIM バンガロールの 2 つのセンター・オブ・エクセレンス、みずほ印日研究センター (MIJSC) とソフトウェアおよび情報技術管理センター (CSITM) が開催したこのパネルディスカッションは、AI の多面的な世界を掘り下げることが目的としていた。万華鏡を覗き込む科学者のように、私たちは AI を様々な角度から検証し、その計り知れない可能性と内在する課題の両方を探ろうとした。

AI：両刃の剣

パネリストたちは、さまざまな分野に革命をもたらす AI の計り知れない可能性を強調した。複雑な課題への取り組みから、タスクの自動化、データに基づく予測まで、AI は世界中のプロセスを再構築し、効率を高め、生活を改善する可能性を提供する。スリヴァスタヴァ博士は、膨大なデータセ

ットを処理し、タスクを自動化し、データ駆動型の予測を行うことで、複雑な課題に取り組む AI の能力を強調した。これは、世界中のプロセスを合理化し、効率を高め、生活の質を向上させる可能性を秘めている。

ニキール博士は、AI の経済的利益に光を当てた。彼は、AI を活用したアプリケーションがいかによりプロセスを最適化し、生産性の向上と経済成長をもたらすかについて語った。「住宅市場における AI の予測にはフィードバックループが存在するため、住宅価格は長期的にプラス方向にもマイナス方向にも不安定になる」とニキール・マリク教授は説明した。

しかし、議論では AI に内在する危険性も認識された。グンツク博士は、差別的な結果を招きかねないアルゴリズムのバイアスについて懸念を示した。彼は、倫理的な実装を確実にするために、公正で偏りのない AI モデルを開発することの重要性を強調した。

マリク博士はこうした懸念に共鳴し、AI が既存の社会的不平等を悪化させる可能性を強調した。彼は、アルゴリズムによる価格設定やローン承認などの分野におけるバイアスの課題について議論した。アルゴリズムによる偏見、プライバシーへの懸念、データ・セキュリティ・リスク、潜在的な離職といった問題が、早急な対応が求められる重要な倫理的課題として挙げられた。

特に懸念されるのは、メンタルヘルスや住宅など、個人の幸福に直接影響する領域での AI の利用である。AI アルゴリズムが既存の偏見を永続させたり増幅させたりする可能性があり、不当な扱いや差別につながるとして、専門家や政策立案者の間で警鐘が鳴らされている。さらに、多くの AI システムの不透明な性質と、その意思決定プロセスをめぐる透明性の欠如が、説明責任と信頼に関する懸念を煽っている。

「デジタルデータ×AI は、特に恵まれないコミュニティにおいて、メンタルヘルスをリアルタイムで測定するための素晴らしい低コストリソースである。異質性は実在する！ AI モデルは文化の違いを考慮し、バイアスを補正する必要がある。孤独を感じていると言う人がいれば、おそらくそうなのだろう」。

さらに、重要な意思決定プロセスにおける AI システムへの依存の高まりは、その目的を社会的価値観と一致させ、そのアウトプットが公正で倫理的であり、すべてのステークホルダーにとって有益であることを保証する必要性を浮き彫りにしている。AI の開発者、導入者、エンドユーザー間のインセンティブの不整合は、意図しない結果を生み出し、社会的信頼を損ない、これらの技術の責任ある採用を妨げる可能性がある。

AI の複雑さを乗り越える

パネリストは、AI が持つ変革の力を活用しつつ、そのリスクを軽減するためには、多方面からのアプローチが必要であることを強調した。ディスカッションでは、以下のような戦略が検討された：

- **偏見を減らす：** AI アルゴリズムにおけるバイアスを特定し、対処するための技術が探求された。
- **機会の特定：** パネリストたちは、AI をさまざまな分野で活用し、ポジティブなインパクトをもたらす方法について議論した。
- **公平性の実現：** AI アプリケーションの公正さと責任ある開発を確保するための戦略が取り上げられた。

スリヴァスタヴァ博士は、AI モデルの訓練に多様なデータセットを採用し、現実世界を正確に反映させることを提案した。

グントック博士は、開発中の AI モデルを評価し、潜在的なバイアスを特定し、修正するために、公平性の評価基準を使用することを提案した。

マリク博士は、責任ある AI の開発と配備を保証するための規制と倫理的ガイドラインの重要性を強調した。

求償の選択肢

講演者は、この複雑な状況を乗り切るためには、学際的なアプローチが必要であることを強調した。そして話題は、AI 開発における偏見を緩和し、公平性を実現するための解決策へと移った。スリヴァスタヴァ博士は、AI モデルの訓練に多様なデータセットを採用し、現実世界を反映させることを提案した。グントック博士は、開発中の AI モデルを評価し、潜在的なバイアスの特定と是正を可能にする公平性の指標を提案した。マリク博士は、責任ある AI の開発と展開のための規制と倫理的ガイドラインの重要性を強調した。

公平性の指標の構築と透明性の育成

グントック博士は、公平な AI 開発へと導く羅針盤に例えられる、公平性評価基準の概念を紹介した。このメトリクスによって、開発段階で AI モデルを評価し、バイアスが忍び込んでいる可能性のある領域を特定することができます。このようなバイアスに積極的に対処することで、AI システムから公平な結果を得ることができる。

彼は、秘密のベールに包まれた AI アルゴリズムではなく、透明性を念頭に置いて設計された AI アルゴリズムを構想している。この透明性によって、ユーザーはモデルの判断の背後にある根拠を理解することができ、信頼を育み、存在する可能性のあるバイアスを特定することが可能になる。

多面的アプローチ：協力、教育、規制

パネリストは、AIの偏見という課題に対処するためには、多面的なアプローチが重要であることを強調した。マリク博士は、研究者、政策立案者、産業界のリーダーなど、さまざまな関係者が協力する必要性を強調した。協力することで、AIを取り巻く倫理的配慮に対処する包括的なソリューションを開発することができる。

政策立案者への提言

1. 住宅市場価格の「フィードバック」と「過剰変動」の監視と検出。
2. AIを提供するプラットフォームでAI「モデルコントロール」を実施する。

o 2007-08年の金融危機後に金融資産価格モデルに導入された「モデル・コントロール」と精神が似ている。

偏見との闘いにおいて重要な要素として、公教育も浮上した。マリク博士は、『デジタル方言』で強調されたメディア・リテラシーの必要性に共鳴し、AIとその潜在的な落とし穴についての認識を高める教育的イニシアチブを提案した。十分な情報を得た一般市民は、責任あるAIの開発と利用を擁護するためのより良い環境を整えることができる。

規制は、議論の中で強調されたもうひとつの重要な側面だった。交通法が道路の秩序を保証するように、規制は倫理的なAI開発の枠組みを提供することができる。ミスラ博士は、AI技術の急速な進化に対応できる適応可能な規制を提唱した。

前途多難：継続的な警戒の呼びかけ

パネルが概説した解決策は、AIにおける偏見を緩和し、公平性を確保するためのロードマップを提供する。しかし、偏見との戦いは継続的な取り組みであり、継続的な警戒が必要である。これらの解決策を採用し、オープンな対話の文化を醸成することで、我々はすべての人の利益のためにAIの力を活用することができる。アルバート・アインシュタインが適切に言ったように、「数えられるものすべてが数えられるわけではなく、数えられるものすべてが数えられるわけでもない」。AIの開発と監視における人間の要素は、このテクノロジーが明るい未来に向けて人類に役立つことを保証する上で、依然として最も重要である。

共同作業の呼びかけ

ディスカッションの最後には、学者、産業界の専門家、政策立案者、倫理学者、リスク評価の専門家など、さまざまなステークホルダー間の協力が呼びかけられた。パネリストは以下の重要性を強調した：

1. **政府の監督と規制：**重要な分野での責任あるAIの開発と実装を確実にするために、パネリストは政府の監視を強化すること、そして単一の情報源に依存することに伴うリスクを軽減

するために、複数の独立した価格設定モデルや信用スコアリングシステムを検討することを提唱した。

2. **透明性と開放性**： AI アルゴリズムの精査を促進し、バイアスを緩和するために、たとえ基礎となるモデルが完全に開示されていないとしても、予測とともに解釈可能な AI の説明を提供することをプラットフォームに義務付けることが提案された。商業的な AI 価格設定プラットフォームの代替として、政府がオープンソースで非営利の価格設定モデルを提供する案も提起された。これは、利益動機に左右されず、中立的なベンチマークとして機能する可能性があるからだ。
3. **専門知識への投資**： パネリストは、複雑な AI システムの監査と理解において規制当局が直面する課題を認識し、AI アルゴリズムを効果的に監視・監査するために必要な AI とデータ分析の専門知識を政府に装備させる必要性を強調した。

質疑応答

質問: 患者のプライバシーと自律性の原則を守りながら、メンタルヘルス格差に対処し、介入を促進するために、テクノロジーは倫理的にどのように統合できるのだろうか。

シャラート・チャンドラ・グントック博士： 「インフォームド・コンセントは非常に必要なもので、利用規約のように、ただスクロールして承諾をクリックするようなものであってはならない」と思う。グントック博士は、患者のプライバシーと自律性を取り巻く透明性の重要性を強調する。「AI を活用したメンタルヘルスケアは、疲弊したトリアージ・システムのように思われがちだ。「臨床医が必要なのは確かですが、同時に AI は、どれだけの人が治療を必要としているか、どれだけの人が実際に治療を受けているかのギャップを埋めることができる」と彼は述べ、メンタルヘルス格差への効果的な対処を確実にするために、メンタルヘルスケアのギャップを埋める AI の可能性を強調した。「AI は、臨床医やプライマリ・ケア提供者のフォローアップができるような個人を推薦したり、ある種の予測を与えたりする役割を果たす可能性がある」。

質問: マリク博士、AI のアルゴリズムにおける偏りについての議論は気になる。住宅市場のように AI がますます活用される分野では、どのようにすれば公平性を確保できるのでしょうか。

ニキール・マリク博士： 「Zillow のようなプラットフォームが使用するアルゴリズムは、人種や社会経済的地位のような要因に基づく住宅価格の既存の偏りを永続させる可能性がある。これを緩和するために、私たちはいくつかの手段を講じることができます」とマリク博士は語った。「まず、データの収集方法を精査する必要がある。AI アルゴリズムのトレーニングデータに偏りがあれば、結果として得られるモデルにも偏りが生じる。公平性を確保するためには、多様で代表的なデータセットが必要だ」。第二に、「透明で説明可能な AI モデルを開発する必要がある。これによって、アルゴリズムがどのように判断に至るかを理解し、潜在的なバイアスを特定することができる。ア

ルゴリズムが一貫して特定の地域の住宅を過小評価している場合、その理由を知る必要がある」。そして最後に、彼はこう締めくくった。「最後に、AI アルゴリズムに偏りがあった場合、プラットフォームに責任を負わせるための規制の枠組みを確立する必要がある。政府は、住宅のような重要な分野における AI の公正かつ倫理的な利用を確保する上で、重要な役割を果たすことができる」。

質問: 政府はもっとオープンソースに近い代替案を提示できるのか。

ニキール・マリク博士:「政策立案者が、国民として数十億ドル規模のプロジェクトに資金を提供し、ベンチャーキャピタルのような政府への投資を行い、このような大規模な AI を開発したいという意識を持つことは、現実的には難しいでしょう」とマリク博士は指摘する。

質問: AI 規制において政府はどのような役割を果たせるのか。

政府は、AI 開発の透明性、説明責任、公平性を促進する規制を設けることができる。これらの規制は、AI 技術の急速な進化に対応できるものでなければなりません」とスリヴァスタヴァ博士は認めた。

質問: スリヴァスタヴァ教授、あなたは政府の監視の重要性に言った。しかし、これでは AI 分野のイノベーションが阻害されるのではないのでしょうか。

ジャイディーブ・スリヴァスタヴァ博士:「必ずしもそうではない。規制は実際に責任あるイノベーションを促進することができます。明確なガイドラインは、開発者が有益で倫理的な AI アプリケーションの作成に集中するのに役立つ」とスリヴァスタヴァ教授は聴衆を安心させた。さらに、「政府の監視は AI に対する国民の懸念に対処するのに役立つ。透明性と説明責任は、この急速に進化するテクノロジーに対する信頼を築く上で極めて重要だ。もちろん、規制はイノベーションを妨げないよう慎重に設計する必要がある。適切なバランスを取るためには、産業界、学界、政策立案者を巻き込んだ協力的なアプローチが不可欠である」。

結論

最後に、「AI の約束と危険: 偏見を緩和し、機会を探り、公正さを達成する」というパネルディスカッションでは、メンタルヘルスへの AI の応用の可能性が強調され、特に恵まれないコミュニティでは、AI によるデジタルデータ分析がリアルタイムのモニタリングに低コストで対応できる可能性があることが示された。しかし、AI モデルのバイアスに関する懸念が提起され、文化の違いを考慮し、訓練データの潜在的なバイアスに積極的に対処する必要性が強調された。AI の約束と危険は表裏一体であり、この状況を乗り切るには、イノベーションと責任ある展開の間の微妙なバランスが必要であることが明らかになった。この思いは、様々な分野にわたる AI 開発における公平性の重要性と偏見の緩和に関するより広範な議論と一致している。

進むべき道は、利害関係者間の協力を促進し、透明性を促進し、強固なガバナンスの枠組みを確立することにある。政府は、AI 監査能力への投資、代替価格モデルの検討、解釈可能な AI 説明の義務化など、積極的な措置を講じる必要がある。さらに、AI システムが公正さ、説明責任、社会の最善の利益を念頭に置いて開発・導入されることを確実にするためには、多様な専門知識を含む学際的な政策立案が極めて重要である。

AI が私たちの生活のあらゆる側面に浸透し続けるなか、私たちは警戒を怠らず、責任を持ってイノベーションを受け入れ、人工知能の時代における偏見の緩和、機会の探求、公平性の実現に向けて一丸となって取り組むことが不可欠である。