*.... In conversation with*

## Prof. Neeraj Kumar Sharma,

Professor, Mehta Family School of Data Science & Artificial Intelligence
The SPIN (Sensing, Perception & Intelligence) Lab
IIT Guwahati

**MFCEM, IIT Kanpur:** Hello Prof. Sharma, it is great to connect and get familiar with your work. Your SPIN (Sensing Perception & Intelligence) lab is carrying out some fantastic work on sound and perception. What attracted you to this domain of research, could you broadly touch upon some of the ongoing work in your lab?

**Neeraj Kumar Sharma:** Thank you for the kind words. My journey to pursuing a PhD was somewhat serendipitous and nonlinear. What I was certain about, however, was the need to deepen my technical understanding so I could contribute meaningfully to engineering solutions. This early motivation, combined with a broader curiosity about how to probe and understand intelligent behavior in living systems, shaped my research direction.

Intelligence, as we observe in humans and other species, is not a static property. It emerges from the interaction of sensory inputs, context, memory, and adaptation, and evolves over time across an individual's lifetime as well as at a broader evolutionary scale. This perspective led me to explore how machines can perceive and interpret the world in ways similar to humans, and how they can be engineered to even surpass human performance in specific, well defined scenarios. Pursuing this direction required drawing upon and integrating ideas from multiple disciplines, including psychology, computer science, electrical engineering, and mathematics

Within this broader context, sound emerges as a particularly compelling modality. Many aspects of behavior and interaction, such as speech, movement, and physiological processes, naturally produce acoustic signals. As a result, sound provides a rich and non-invasive way to capture information about both the environment and living systems, and to study how intelligent behavior manifests in real world settings. In many ways, if there is a sound, there is an underlying process or "story" behind it, and with advances in audio analysis, data science, and machine learning, we can begin to uncover that story and translate it into actionable insights.

> .....our research focuses on building intelligent systems that can extract meaningful insights from audio and multimodal data.

I was first introduced to the beauty of signal processing for understanding sound during my PhD, working with T. V. Sreenivas (TVS), Professor at the Indian Institute of Science Bangalore. Even now, in our discussions, he continues to amaze me with ideas on how sound may hold answers to many fascinating phenomena we observe around us. At the same time, intelligence is inherently multimodal. It arises from the integration of multiple sensory inputs and contextual cues. This motivates us to go beyond audio and explore approaches that combine sound with other modalities such as vision and physiological signals, enabling a more comprehensive understanding of perception and behavior.

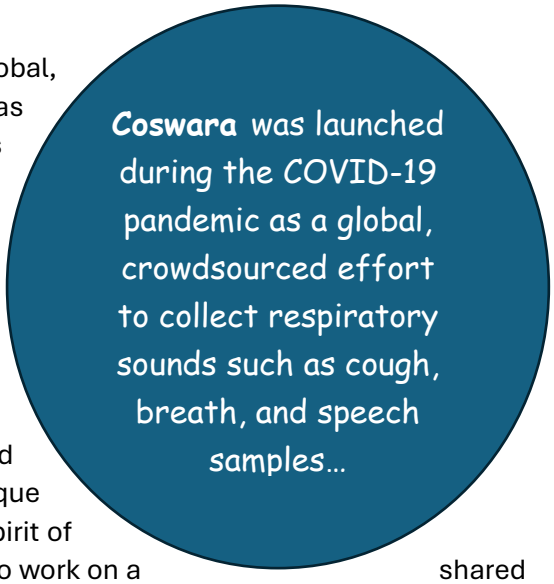
At **the SPIN (Sensing, Perception and Intelligence) Lab** at Indian Institute of Technology Guwahati, our research focuses on building intelligent systems that can extract meaningful insights from audio and multimodal data. We work at the intersection of signal processing, data science, and AI.

Broadly, our work spans creating datasets, building benchmarks, and developing demonstrable AI based solution methodologies. The problems we address include acoustic based respiratory health monitoring, uncovering acoustic patterns in spoken language, speech and speaker recognition, source separation, conversational AI, atypical speech processing, and developing neuro-inspired signal processing and AI models.

Being in one of the most linguistically diverse regions of the world, a key focus for us is to document spoken languages and build technological applications for diverse low resource languages of North East India. In addition, we explore multimodal learning frameworks that integrate audio with other data sources, extending beyond human AI interaction to broader engineering applications.

**MFCEM, IIT Kanpur:** One cannot help but notice Prof. Neeraj the constant endeavor of yours to translate fundamental insights from your lab to possible applications. Your earlier and ongoing work has resulted in unique resources for the scientific community and public at large. Could you tell us a little about the one-of-its kind projects that you spearheaded such as **Coswara** and **Boli**?

**Neeraj Kumar Sharma:** Thank you for pointing that out. Translating research into real world impact has always been an important guiding principle for our work. Two initiatives that reflect this philosophy are Coswara and Boli.




Coswara was launched during the COVID-19 pandemic as a global, crowdsourced effort to collect respiratory sounds such as cough, breath, and speech samples. The project was conceptualized in April 2020, when many of us were confined to our homes during the lockdown. A key member of the team, and indeed the one who initiated the effort, was Sriram Ganapathy, Associate Professor at Indian Institute of Science Bangalore, with whom I had the joy of working as a postdoctoral researcher. The idea was to create an open access respiratory audio dataset that could support the development of AI based tools for early screening and monitoring of respiratory conditions. What made Coswara unique was the audio modality, the scale of participation, and the spirit of open collaboration. It enabled researchers across the world to work on a shared resource during a critical time. We curated the dataset and released it as an open access resource, accompanied by a publication in Nature Scientific Data, along with validation studies and baseline models. We also conducted two machine learning challenges at flagship conferences to encourage wider participation from the research community.

On the other hand, the **Boli project focuses on creating a stuttered speech dataset**, with a particular emphasis on the Indian context. The project is led by my PhD scholar, Ashita Batra, who originally proposed this idea to me. **The project also aims to understand the acoustic patterns associated with the atypical nature of stuttered speech.** It further aims to capture aspects of the lived experience of people who stutter, including how they adapt, if at all, in everyday spoken interactions. The larger goal is to build technological solutions that can assist people who stutter, just as they do for everyone else.

Both projects emphasize openness, inclusivity, and societal relevance. They also demonstrate how fundamental research, when combined with collaborative efforts, can lead to resources that benefit not just the academic community but the public at large.

**MFCEM, IIT Kanpur:** We are living in interesting times, there is a palpable shift—now more than ever—in which the lines separating engineering, technology, and medicine as distinct domains are fading. Instead, addressing societal and medical challenges has taken center stage through a more integrated approach, driven by active collaboration and collective effort. Would you like to share some thoughts for those who are contemplating getting started in research career.

**Neeraj Kumar Sharma:** Indeed, it is a very exciting time for research. As you noted, the boundaries between traditional disciplines are increasingly porous. Addressing healthcare challenges, for instance, is no longer a solitary endeavor; it requires engineers, clinicians, and scientists to move beyond their silos and learn to speak each other's professional languages. At the SPIN Lab, we often emphasize that the quality of the question defines the research journey. For those starting out, curiosity is your most reliable compass. It is perfectly normal to begin a PhD without a clearly defined problem; research environments are designed to help you discover and refine such questions. What matters is the honesty, dedication, and a systematic approach to pursue them once they emerge.

...Addressing healthcare challenges, for instance, is no longer a solitary endeavor; it requires engineers, clinicians, and scientists.....

While emerging AI tools can significantly accelerate research workflows, they are most effective when built upon a strong foundation in computational methods and domain knowledge. I continue to find immense value in landmark papers and classic texts, as they offer a depth and structure that AI generated summaries can sometimes miss, depending on how one engages with and guides the interaction. These tools are best used to navigate the landscape, while critical thinking and rigorous validation ensure that your work remains original and robust.

Finally, we need to remember that research is a collective and deeply human effort. A PhD is not just about individual output, but also about contributing to a healthy ecosystem where curiosity, both yours and that of your peers, can flourish. Stay curious, focus on depth, and aim for impact that reaches beyond the immediate problem.

MFCEM: Prof. Sharma thank you much for speaking to us, it was a pleasure. Wishing you and your team the very best.

Neeraj Sharma with his team of PhD Scholars at SPIN Lab, IIT Guwahati. Left to right: Prakhar Kumar Sonkar, Ashita Batra, Timothy Alex John, Sneha Ray Barman, Lammibert Sumer, Neeraj Kumar Sharma, and Aamir Hamid Dar.