A Discussion With Dr. Alaa Youssef

Introduction:

Welcome to the AI Frontiers Podcast, a Dialogue with Tech Pioneers, hosted by Stanford University Technology Training. I'm Don Cameron.

We are thrilled to have Dr. Alaa Youssef, a postdoctoral fellow in AI Ethics and Governance at Stanford School of Medicine. Dr. Youssef holds a PhD in Health System and Population Health from the University of Toronto and is a leading expert in responsible AI implementation and evaluation.

Dr. Youssef leads several AI educational programs and policy initiatives. She co-directs the Stanford AIMI (acronym for Center for Artificial Intelligence in Medicine and Imaging, also known as AMY) High School Programs, preparing the next generation for careers in AI and medicine. She also serves on several AI policy and education committees at Stanford School of Medicine, shaping the future of AI in healthcare.

Join us as we explore the fascinating world of AI ethics and healthcare with Dr. Alaa Youssef!

Dr. Youssef – thank you very much for joining us today.

Questions

1. To lead things off, please share with us how you got interested in the area of AI and its applications in healthcare.

Artificial intelligence (AI) in clinical medicine and patient care offers numerous benefits, and great potential for improved care, but it also raises several key ethical issues.

a. One of the key issues with AI in Healthcare is Bias and Fairness

AI systems can perpetuate and even exacerbate existing biases in healthcare data, leading to unequal treatment across different populations, especially marginalized groups.

You've worked extensively on evaluating the fairness and bias in large multi-modal models. What methodologies do you recommend to ensure fairness in AI models?

What are some of the ways that you think will help us ensure that AI systems do not favor certain groups over others?

b. **Another key challenge for AI in healthcare is Transparency and Explainability**

Many AI systems operate as "black boxes," making decisions without clear explanations. This lack of transparency can hinder trust and accountability.

What can be done to help Patients and clinicians to understand how AI systems reach their conclusions to maintain trust in the technology?

c. **Privacy and Confidentiality is another area of concern.**

Training AI systems requires vast amounts of personal health data, raising concerns about data breaches and unauthorized access.

Are Patients fully informed about how their data will be used and do they give consent for usage?

You discussed Stanford's use cases in your publication on Organizational Factors in Clinical Data Sharing for Artificial Intelligence in Health Care. What were some of your key findings?

What are some steps Stanford medicine is taking in protecting patient privacy?

d. **There is also the issue of Accountability and liability**

Determining who is accountable when an AI system makes an error is complex. How are Clinicians and healthcare institutions addressing this issue?

What are some of areas of specialty where AI is mostly being applied in healthcare, and what challenges are they presenting?

Do you envision the government stepping in with Regulation and Oversight to help address this issue?

e. **Have there been any studies on the Impact of AI in the Doctor-Patient Relationship?**

f. **Another important area of concern involves the Equity of Access to this new technology.**

The availability of AI technologies may be limited to well-funded healthcare institutions, creating disparities in access to advanced medical care. There will undoubtedly be Global Inequality. How is this issue being addressed?

g. **Failures can have severe consequences for patient health.**

   Are there any quality standards or safety protocols associated with the application of AI in Healthcare?

h. What do you think will be the impact of AI on future Employment in healthcare? Will there be displacement for certain roles?

i. How is AI technology impacting education and Retraining of healthcare professionals?

If there is time ……..

2. What are some of the AI tools you recommend for use in your field?

3. What is your current biggest challenge?

4. Who are some of the leaders in responsible AI implementation in Healthcare?

5. You have been involved in various multidisciplinary research labs. How has working in such diverse environments influenced your research approach?

    a. **Follow-up:** What have some key learnings been from these collaborative experiences?

6. Your study on organizational readiness for health data-sharing for AI highlighted several facilitators and barriers. Could you elaborate on these findings?
    a. **Follow-up:** How can organizations overcome the barriers to effective data-sharing?

7. In your research, you have identified AI ethics and governance issues. What are the key challenges you have come across?
    a. **Follow-up:** How have these frameworks been received by policymakers and healthcare professionals?

8. Your work includes mentoring students in AI research. How do you approach mentorship, especially for students without a background in radiology or AI?
    a. **Follow-up:** What are some success stories from your mentorship experiences?

9. Your research on AI clinical trials addresses ethical and workflow challenges. What are the unique ethical challenges in AI clinical trials compared to traditional clinical trials?
    a. **Follow-up:** How can these challenges be mitigated to ensure ethical integrity in AI clinical trials?

END

Thank you so much, Dr. Alaa Youssef, for joining us today and sharing your insights into AI ethics and governance. Your work is truly inspiring and vital to improving the future of healthcare.

That brings us to the end of this episode of AI Frontiers Dialogue with Tech Pioneers. We hope you enjoyed our conversation with Dr. Youssef and gained a deeper understanding of AI healthcare's ethical challenges and innovations.

Thank you to Dr. Youssef for sharing your expertise with us. To learn more about her research, check the links in the show notes.

Thank you for listening, and until next time, stay curious and keep exploring the frontiers of AI!

Transcript

Note: Any views, discussions, or opinions expressed in this podcast do not represent in any way the opinions or positions of Stanford University, its staff, or its employees.

Don Cameron: Welcome to the *AI Frontiers* podcast, a dialogue with tech pioneers hosted by Stanford University Technology Training. I'm Don Cameron. We are thrilled to have Dr. Alaa Youssef, a postdoctoral fellow in AI Ethics and Governance at Stanford School of Medicine.

Dr. Youssef holds a PhD in Health Systems and Population Health from the University of Toronto and is a leading expert in responsible AI implementation and evaluation. She leads several AI educational programs and policy initiatives. She co-directs Stanford AIMI—the Center for Artificial Intelligence in Medicine and Imaging—also known for its high school programs preparing the next generation for careers in AI and medicine.

She also serves on several policy and education committees at Stanford School of Medicine, shaping the future of AI in healthcare. Join us as we explore the fascinating world of AI, ethics, and healthcare.

And Dr. Youssef, thank you very much for joining us.

Dr. Alaa Youssef: Thank you. It's such a pleasure to be with you today.

Don Cameron: We want to kick things off with a question: Can you share with us how you got interested in the area of AI and its applications in healthcare?

Dr. Alaa Youssef: Yeah, this really goes back to my PhD at the University of Toronto where I was using machine learning to build algorithms that could help predict long-term patient outcomes to improve population health. Healthcare has always struggled with improving clinical outcomes, and one of the core areas where AI can help is in prediction and prevention—catching patients before they need hospitalization rather than treating them afterward.

I initially envisioned building an algorithm to help bariatric patients by predicting outcomes using electronic health records. I spent six months of my PhD cleaning the data and started questioning if this was the norm. Through both quantitative and qualitative research, I realized patient stories often reveal things not captured in health records. As we tested models, it became clear they had limitations, which led me to think more broadly about the healthcare system's readiness to adopt AI.

Healthcare is like an upside-down funnel. What happens at the macro level translates to the clinic and ultimately impacts patient outcomes. If we get the system right, much else follows. That motivated me to pursue postdoctoral training focused on implementation, combining it with my understanding of health systems.

Don Cameron: Thank you. AI systems can perpetuate—and even exacerbate—existing biases in healthcare data, leading to unequal treatment, especially for marginalized groups. You've worked extensively on evaluating fairness and bias in large multimodal models. What methodologies do you recommend to ensure fairness?

Dr. Alaa Youssef: That's a really complex question with no single answer. The field is still grappling with what bias and fairness mean in clinical data. Some biases are visible and measurable, while others are not. For example, a model might underrepresent specific groups simply because they're not well-represented in a geographic region or dataset.

Then there are implicit biases—subtle patterns that exist in data and are unknowingly learned by algorithms. As researchers, we ask: What proxy measures are we using? How do we define fairness? For instance, if one algorithm improves access to care but performs worse in sensitivity and specificity compared to another, is it fair to use it?

Efforts like the Coalition for Health AI (HAI) are working to develop standards and best practices. But we’re still studying how models perform in real-world settings, where they go wrong, and what fairness really means.

Don Cameron: What are some ways we can ensure these AI systems don't favor certain groups over others?

Dr. Alaa Youssef: That’s nuanced. We assume underrepresentation leads to poorer model performance, which can be true—but not always. For example, in diabetic retinopathy, our study showed surprising results. Despite differences in skin pigmentation, the model performed well across different test sites. So generalizability varies.

We must examine the data's representativeness and assumptions around how data is encoded. Bias isn't only in development; it's also in usage. For instance, a heatmap in an image might lead clinicians to overlook other areas. Bias appears across the pipeline—from data capture to clinical interpretation. We need both quantitative and qualitative evaluations, and we must consider downstream effects and ethical implications.

Don Cameron: Many AI systems operate as “black boxes,” making decisions without clear explanations. This can hinder trust and accountability. What can be done to help patients and clinicians understand how AI systems reach conclusions?

Dr. Alaa Youssef: This is something I think about often. From qualitative interviews with clinicians, I’ve found trust often hinges on consistent, accurate output. But clinicians also want to know how a model reaches its decisions.

Explainability differs by role. What’s understandable to an engineer may not be to a clinician or a patient. Dr. Nigam Shah writes about this—interpretability isn’t universal. Just because a model is explainable doesn’t mean it’s correct.

In medicine, clinicians are trained to think in terms of biological processes and symptoms. AI models may use pixel data or other features that clinicians don’t see as meaningful. And with the rise of large language models, we’re seeing a shift: people trust these systems more, even without explanations, because they seem convincing.

But we need to be cautious. Automation bias is real. Explainability should not replace critical thinking.

Don Cameron: What kinds of applications or tools are clinicians currently using?

Dr. Alaa Youssef: Radiology leads the way, with over 700–800 AI-powered medical devices. These models can detect things in images that radiologists might miss, raising ethical questions: if a tool can improve care, shouldn’t we use it? But validation is crucial.

Other specialties like radiation therapy, dermatology, and pathology are also seeing adoption. AI is helping reduce radiation doses, identify brain aneurysms in emergency settings, and even predict outcomes.

With large language models, we're seeing tools that assist with documentation, billing, and policy—boosting efficiency and productivity.

Don Cameron: Let’s talk policy. Training AI requires vast amounts of personal health data, raising concerns about consent, breaches, and unauthorized access. How is this being handled?

Dr. Alaa Youssef: Patients must consent for their data to be used. In most cases, data is de-identified—stripping out 18 key variables. That process is time-consuming and a barrier to widespread AI development.

Hospitals are protective of data, which is good for privacy but limits research. Secure GPT platforms—like Stanford’s internal one—offer a safer way to use language models within institutional firewalls. That’s critical, as we know these models can leak data.

Don Cameron: In your publication on organizational factors and clinical data sharing for AI, you highlight Stanford’s use case. What were your key findings?

Dr. Alaa Youssef: Stanford AIMI has led in sharing datasets for algorithm development. But not all institutions are as open. Why? It comes down to organizational readiness, resources, and motivation.

Stanford's vision of precision medicine aligns AI with its mission. Leadership supports it, infrastructure exists, and privacy governance is strong. In contrast, other health systems may lack resources or hesitate to share before completing their own research.

Motivation is key. Government incentives—like during COVID-19—helped accelerate data sharing. But for-profit companies may act differently, driven more by financial interests.

Don Cameron: What about tools like ChatGPT in healthcare? How is Stanford addressing potential data breaches?

Dr. Alaa Youssef: Stanford Medicine uses a secure GPT environment to protect patient data. That ensures anything processed stays within Stanford's system and isn't used to train external models. It's been a game-changer, enabling us to use the tools responsibly.

Don Cameron: What steps is Stanford taking to protect patient privacy?

Dr. Alaa Youssef: Stanford's privacy and governance teams evaluate requests for data sharing. They don't just look at de-identification—they also assess re-identification risks and vendor compliance. The focus is on responsible collaboration, particularly with for-profit companies.

Don Cameron: Determining who's accountable when AI makes a mistake is complex. How are institutions handling this?

Dr. Alaa Youssef: This remains unresolved. Clinicians are expected to use these tools, but they didn't develop them. If something goes wrong, who's liable—the doctor or the developer?

We're asking physicians to be clinical experts *and* tech evaluators, which is unfair. We need clearer regulations around vendor accountability. Faculty at Stanford and elsewhere are exploring legal frameworks for this.

Don Cameron: Do you envision government stepping in with oversight?

Dr. Alaa Youssef: Yes. We've seen federal executive orders promoting responsible AI. But impact varies by region and institution. Implementation will take time, and we’re still waiting to see how effective these regulations will be.

Don Cameron: Have there been studies on AI’s impact on the doctor-patient relationship?

Dr. Alaa Youssef: Yes. One study showed patients preferred AI-generated responses over doctors’ because they seemed more empathetic. This is a call to action—physicians need to ensure their interactions remain patient-centered.

Don Cameron: A final question: AI access is often limited to well-funded institutions, creating disparities. How do we address global inequality in healthcare AI?

Dr. Alaa Youssef: Unfortunately, this issue isn’t being adequately addressed. Everyone wants to build AI tools, but few invest in the infrastructure that supports them.

Healthcare systems without the resources can’t build the pipelines necessary for effective AI implementation. It requires serious investment—and many organizations are already stretched thin. Disparities will persist until we prioritize infrastructure and collaboration.