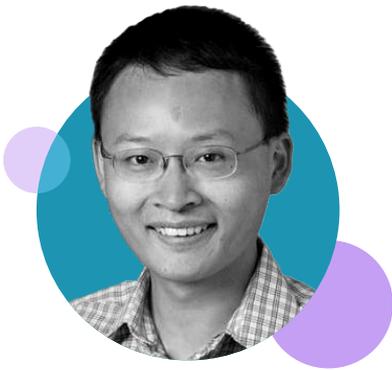




Interview

James Zou, Associate Professor of Biomedical Data Science at Stanford University, California, USA, joined EMJ for a discussion on AI and its role in transforming clinical research. He discusses the shift towards AI systems that can generate hypotheses, the challenges of deploying AI tools in real-world healthcare, and the growing role of real-world data and wearable technologies in predicting disease and improving trial design.



James Zou

Associate Professor of Biomedical Data Science, Stanford University, California, USA

Citation: EMJ. 2026;11[1]:28-31.
<https://doi.org/10.33590/emj/OR9BWSEN>

Q1 When did you first become interested in applying AI to clinical research?

I've been working at the intersection of AI and biomedicine for about 15 years now, since my PhD. My initial work was more focused on using AI to enable biomedical discoveries. When I came to Stanford, California, USA, in 2016, that's when I started doing more work on AI for clinical research.

Q2 Your work spans clinical prediction, genomics, and large-scale AI models. When you look at the field today, what's changed the most since you first started working in healthcare AI?

I think the biggest change we've seen is the shift from viewing AI as a tool to viewing AI as more of an agent or co-scientist.

In the past, when I first started in the field, we would begin with a specific problem in medicine or healthcare and then apply or develop an AI tool to tackle that problem. But in the last 2

years, we've seen the emergence of much more autonomous AI driven by language models and AI agents. Because these agents are more autonomous, they can start to come up with their own problems. They don't have to wait for us to define the problem. They can generate hypotheses and even create their own tools.

Q3 From a clinician's perspective, for example, in dermatology, when diagnosing melanoma, if AI provides a conclusion, a clinician still has to review it. Do you think that will always be necessary, or could we step aside in the near future?

Currently, it's still necessary for human experts and clinicians to provide final judgment and oversight over these AI systems. Not necessarily for every step, but for the most important decisions. For example, the final diagnosis or treatment decision. I think that oversight is still necessary.

In the last 2 years, we've seen the emergence of much more autonomous AI driven by language models and AI agents

Q4 Some studies suggest AI can make fewer mistakes than clinicians. So why do we still need to revise AI decisions?

AI is still a relatively new and emerging technology. The algorithms might work very well on the specific data and hospitals they were trained on, and they could be very reliable there, but if they're applied in the wild, in new clinics or with different patient populations, there's still uncertainty about how the models will perform. They might work very well and be robust, but there's also a chance they could make mistakes.

Because of that uncertainty, at least for now, it's still useful to have human experts involved in the assessment.

We can computationally generate digital or *in silico* cohorts that mirror patients from different kinds of clinical trials

Q5 In your work on Trial Pathfinder, you showed that eligibility criteria can significantly impact both trial enrolment and generalisability. What was the most unexpected insight from that research?

One interesting finding was that we could actually broaden the eligibility criteria. Many existing criteria that determine which patients are eligible for clinical trials are quite restrictive, perhaps overly so.

In our paper, we showed that we could substantially relax some of these criteria. This allows much more diverse and larger patient populations to enrol, including more women, more minorities, and older patients.

Even though we relaxed the criteria to some extent, we found that we could still maintain safety and not incur more adverse events.

Do you think this approach will be implemented more widely in clinical trials?

Yes. Since our research, there's growing recognition that clinical

trials are often overly narrow and restrictive. There are increasing efforts from drug developers, pharmaceutical companies, and regulators like the FDA and EMA to diversify clinical trial populations.

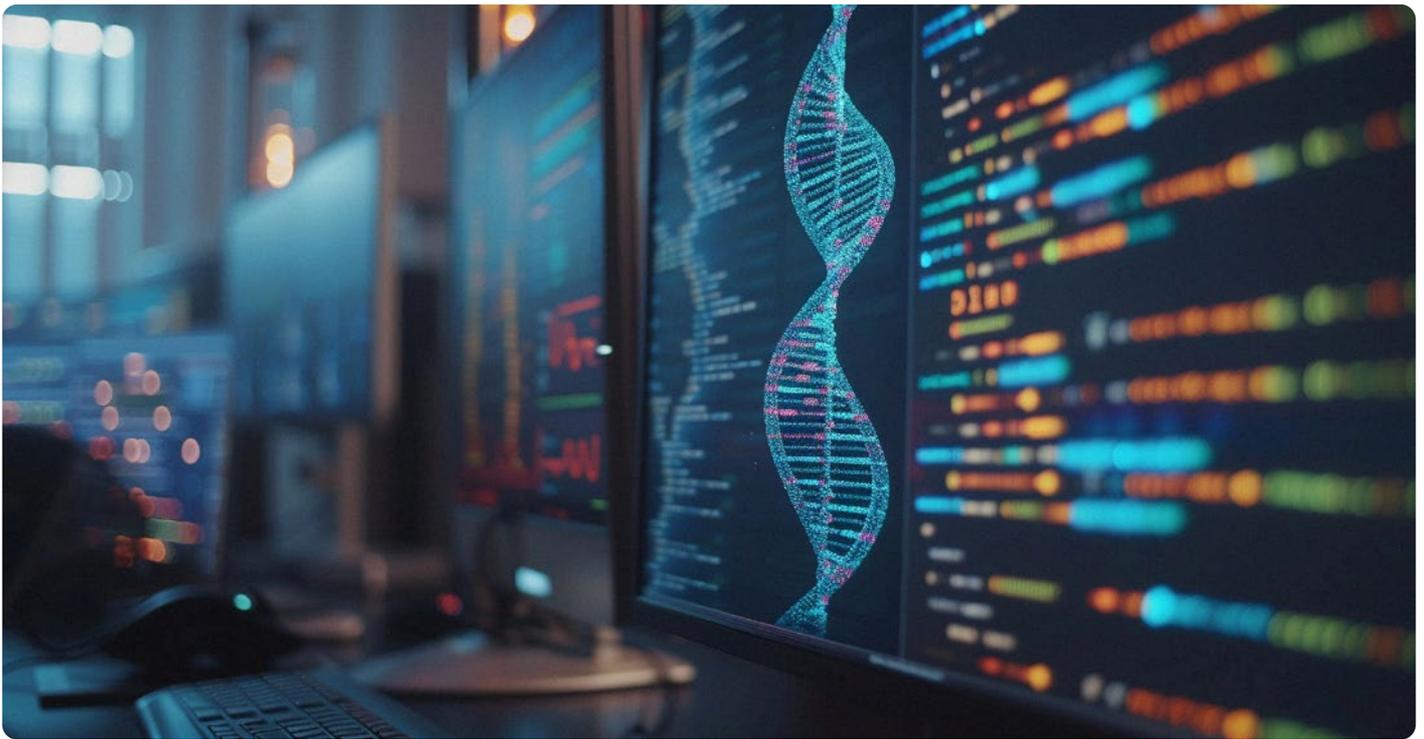
Q6 Your work relies heavily on real-world clinical data. What are the biggest opportunities and limitations of using real-world data to inform trial design and clinical decision-making?

There are many opportunities to leverage real-world data. One area I'm particularly excited about is using it to create digital twins or to simulate clinical trials synthetically.

Real-world data is very diverse and linked to medical records, so we know the outcomes. We can computationally generate digital or *in silico* cohorts that mirror patients from different kinds of clinical trials.

This is faster and can provide valuable insights. It doesn't fully replace actual clinical trials, but it helps us design trials more efficiently and make them more inclusive.





One limitation is that real-world data is often messy and noisy compared to curated clinical trial data. Electronic health records often contain missing information, and there are biases in what gets recorded and what doesn't. We need computational and statistical methods to account for that when leveraging this data.

Q7 There's a lot of excitement around AI in medicine, but also frustration. Where are we genuinely making progress, and where are we getting stuck?

We published a study looking at how many medical AI devices have been cleared by the FDA. There are already over 1,000 AI medical devices cleared by regulation.

That shows significant technological progress. We have AI-driven devices across many imaging modalities and indications that have gone through FDA review.

Where we're getting stuck is deployment. When we looked at those 1,000 FDA-cleared devices and measured adoption, for example, through insurance reimbursement data, we found that only a handful are being widely deployed.

A major bottleneck is the economics. How do you reimburse AI algorithms? How do you quantify their value? Headline metrics are often context dependent. You might see excellent performance in a controlled setting like Stanford, but results could differ in a rural clinic or another country.

Companies need a sustainable financial model. The economics of deployment remains a key challenge.

Q8 Fairness and bias are central topics in clinical AI. Has the discussion matured in recent years?

Fairness and bias are critical, especially when AI systems are making important decisions like diagnoses. We need these models

to be robust and work well across diverse populations and settings.

Bias and fairness are components of robustness. I think we've made significant progress in understanding how to evaluate models rigorously and test them across different sites and distributions.

There is greater awareness and better techniques now to assess and mitigate potential biases.

Q9 How can clinicians recognise when a system may not generalise to their patient population?

Vendors need to provide transparent statistics. For example, if an AI system for dermatology was trained mainly on European populations, that should be clearly stated.

Clinicians in other regions may then decide to conduct additional testing to ensure the algorithm performs well on different skin tones or populations.

Q10 When health systems adopt AI tools, what should they look for beyond headline performance metrics?

Headline metrics are often context dependent. You might see excellent performance in a controlled setting like Stanford, but results could differ in a rural clinic or another country.

Healthcare systems shouldn't just look at the metrics themselves, but also at the context in which they were generated. They need to ask whether that context is generalisable to their own setting.

Q11 Do clinicians today have enough understanding to ensure AI tools are used safely?

There's still an educational process underway. There are many AI systems available, and there's also a lot of noise in the space.

We need more experience and better support to help clinicians evaluate and use these AI devices appropriately.

Q12 Looking ahead, where is your research headed next?

We've been fortunate that several of our research projects have gone through FDA clearance. For example, we developed an algorithm for diagnosing cardiovascular diseases from ultrasound videos. It was evaluated in a clinical trial, cleared by the FDA, and is now being deployed.

Looking ahead, we're excited about leveraging consumer-accessible wearables to predict health status and disease.

In a recent publication, we showed that by analysing one night of sleep recording, our AI model could predict over 100 different diseases.

Participants from sleep clinics were linked to their medical records, so we knew what future diseases they developed. The model could predict 130 diverse diseases, including cardiovascular disease, dementia, chronic kidney disease, and stroke.

This demonstrates the potential value of data that people can collect even during a single night of sleep.

We developed an algorithm for diagnosing cardiovascular diseases from ultrasound videos

