

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Unveiling the Future of AI in Computer Vision with Sumedh Datar

Welcome to the OpenAI podcast, the podcast that opens up the world of AI in a quick and concise manner.

Tune in daily to hear the latest news and breakthroughs in the rapidly evolving world of artificial intelligence.

If you've been following the podcast for a while, you'll know that over the last six months I've been working on a stealth AI startup.

Of the hundreds of projects I've covered, this is the one that I believe has the greatest potential.

So today I'm excited to announce AIBOX.

AIBOX is a no-code AI app building platform paired with the App Store for AI that lets you monetize your AI tools.

The platform lets you build apps by linking together AI models like chatGPT, mid-journey and 11Labs, eventually will integrate with software like Gmail, Trello and Salesforce so you can use AI to automate every function in your organization.

To get notified when we launch and be one of the first to build on the platform, you can join the wait list at AIBOX.AI, the link is in the show notes.

We are currently raising a seed round of funding.

If you're an investor that is focused on disruptive tech, I'd love to tell you more about the platform.

You can reach out to me at jaden at AIBOX.AI, I'll leave that email in the show notes.

Welcome to the AI chat podcast.

I'm your host, Jaden Schaefer.

Today on the podcast, we are thrilled to have Sumedh Datar on the show.

He is a seasoned computer vision research engineer with over six years experience in applied deep learning and computer vision.

He has designed and delivered impactful solutions in the healthcare and retail sectors, serving thousands of users worldwide.

He's a subject matter expert in developing and deploying cutting edge computer vision technologies and we are excited to dive into his insights today.

Welcome to the show.

Hi.

Thank you so much, Jaden.

So that to know all the views are completely my own and talking purely based from my past experience.

I do not represent any company or anywhere that I work for.

So yeah, thank you so much for having me, Jaden.

Super excited to have you on the show and yes, that sounds fantastic.

The first thing I wanted to ask you about is what kind of got you interested in working in this space in general in the beginning?

Have you always been interested in tech?

What was your kind of journey into this?

Yeah, so my journey was somewhat very different compared to others.

So I actually started my undergrad in biomedical engineering and there I was fortunate enough

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Unveiling the Future of AI in Computer Vision with Sumedh Datar

to just do my specialization in computer vision and medical imaging and my final year thesis was actually on identification of cancerous regions and that's where I got super interested into the imaging world and saw how impactful it is and how useful it is to the society.

That's amazing.

That's so interesting.

And growing up, was this an area that you had been interested in?

I guess what kind of introduced you to it?

Did you have family members or friends or people that were kind of interested in this space?

Oh, no.

Actually, it was nobody.

It was just by myself.

I clearly didn't know initially I thought hardware is where I think I'm more interested in and then later I realized that's not my cup of tea and I think medical imaging was closest to software that I really got exposed to while doing my undergrad because since my undergrad was not in computer science and it was not code intensive, the closest to coding was medical imaging and that's when I was super thrilled that I could actually contribute to a good amount of code and that got me interested in this space.

Very cool.

That's amazing.

I was wondering, can you elaborate a little bit on some of the key challenges you've encountered in applying computer vision to health care and the retail sector?

Oh, yeah, sure.

So the biggest challenge is scarcity of data.

Throughout my experience working in the computer vision space, we have always started where there's absolutely no data.

You still need to have a solution, but you don't have data.

How do you solve a problem?

So you need to find creative ways to actually identify and when it comes to data scarcity, okay, you have data scarcity, but you need to overcome and start somewhere, right?

So the possible ways to start are getting a few actual images that you are going to use later, which is really hard.

So for example, when I'm working on the cancer, working in the cancer space, I would actually go to the patient and actually take an image of the patient and then come back and then train my model and see how that works.

Okay.

There are other techniques as well.

There are some open source data sets that's available.

We can take that initially to just start off and then see how the model performs and then put it out in the wild and then create a platform to collect more data.

So data scarcity is the first thing.

Then the second thing is hardware limitation, right?

Based on the problem that you're solving, you need to exactly get the right hardware as well.

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Unveiling the Future of AI in Computer Vision with Sumedh Datar

At what speed the camera has to run, what kind of accuracy you need.
When it comes to healthcare, you should be extra careful and it's not healthcare.
When it's like in other space, sometimes if you skip a few frames, it's fine.
But and also these deep learning models are super heavy.
What kind of models do you want to use?
Do you want like super accurate models which has to detect every single frame or it should be like, yeah, you're okay for a few, but it's important to, you know, have the speed rather than accuracy.
So I worked in both areas and both actually have equal importance.
So it's a trade off too.
And last thing is the annotation of data.
You don't get labeled data.
You don't get annotated data.
So when I say annotated data, say if you have a car in an image, so someone has to manually put a box around the car.
And similarly, you need like thousands of images like that.
And that you need to feed that you feed to the model and then model starts doing the prediction, right?
But annotating that data, what strategies can you use?
Apart from doing it manually to annotate these are some challenges that have faced.
Wow, that's very, very interesting.
How do you kind of approach some of those issues?
For example, the issue of, you know, data quality when developing computer vision solutions.
How what's your approach on that?
The first thing is to have something out there, right?
If you sit back in your lab or if you sit back in your office waiting to get the best possible predicted model out there, that's never going to happen.
You need to have a solution first out there because that helps you in getting more data.
And once you have more data, you can build better models.
And the data that you get is the actual data that you'll be dealing with in future.
Right?
That's nice.
So getting a platform out there and having that platform to collect as much data as possible, it'd be wrong or it'd be like, say, 20% accurate and 80% wrong.
That's fine.
But you're getting a lot of data.
You can go back, you can iterate on it, train your model and then redeploy back rather than waiting forever to have the best model out there.
That's one of the strategies that I've learned.
That's really interesting.
What would you say are some of the, for you and all the things that you are working on, right?
You have a lot of knowledge in this space.
What are some real world problems you've solved using computer vision and kind of what impacts

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Unveiling the Future of AI in Computer Vision with Sumedh Datar

have they had?

So I worked in the healthcare space, especially in the oral cancer space, right?

So patients, at least this is in the Indian scenario, in rural areas where people are predominantly smoke

and they suffer from oral cancer, but they're very reluctant to go get themselves tested because the procedure is so painful.

They actually have to go through a biopsy test, which hurts a lot and many people are reluctant to do that.

How do you make it painless, right?

So I basically saw this problem and I was like, can we solve this painless, right?

Non-invasive way can we really solve?

So I actually went and took a photo and then once you have the photo, you can actually say put a box and be like,

yeah, I think this could be cancerous.

Maybe you have to go for diagnosis, right?

So it may not be accurate, but what's happening is you are actually telling the patient that, hey, look, you seriously have a problem, so you have to go for a better treatment to just save yourself, right?

So that way it was more convincing to the people and that helped them to get better care.

So that's when I saw like real value in computer vision wherein you can solve the problem without pain.

So that was one space and the other space was basically like doing like face recognition.

Like an automated attendance management system and rather than using a book and a pen, you can basically take a snapshot and then you can recognize the faces and then do attendance.

That was the other space.

So computer vision, what it has done is the algorithms are not too crazy.

You don't need to have like a bunch of loops or crazy dynamic programming or anything.

It's just about finding that small problem for which you can apply a computer vision model and then the impact is so huge.

That is how computer vision has always been.

So that way I enjoyed computer vision.

Very, very interesting.

And I would be curious to hear your opinion on this.

You've obviously worked a lot with computer vision and ways that this is helping and making big impacts in some really incredible spaces.

For example, like healthcare.

I absolutely love your examples there because you really can see this is something that is helping the patients so much as making such a big positive impact.

I'd be curious from your perspective as you kind of look to the future with this technology.

Where do you see computer vision mixed with artificial intelligence and everything we're developing in these spaces?

Where do you kind of see that in the future?

What kind of changes do you think will happen?

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Unveiling the Future of AI in Computer Vision with Sumedh Datar

What kind of a technology do you maybe not have today, but you think we may have in the future? Yeah, so like what happened is if you take like say a decade ago, right, was when data science became extremely popular.

Data science, I mean tabular data, right, like say transaction data or say in the insurance space, basically it's just what you're doing is you just have rows and columns and you're just having more and more features, right.

And you could build predictive models with statistical techniques like simple things like linear regression, logistic regression, random forest, things like that, right.

But now what's happening is computer vision.

Is almost paired with the tabular data as well.

And the results on the computer vision side of it is so good that you can actually reason out as to why it is good.

And then you can find ways of how you can tag the computer vision features along with the tabular features and make bigger sense and have better decisions or maybe have decisions that probably you might not think, but the model is actually giving you.

Like that's where the future is kind of moving towards.

Okay, because what has happened is 10 years ago you had, for example, like say Facebook, right, but today you have tiktok, Instagram, Reels, lots and lots and lots of visual content, right, so you can get a lot of information from visual content.

So when you pair visual content with tabular, you can get way more insights.

I think kind of that's where the future is heading towards.

Very, very interesting.

Something that you mentioned earlier also struck me as fairly interesting.

You kind of mentioned the importance of like labeled high quality visual data.

What strategies do you recommend for organizations to source and utilize that type of data?

Yeah, so the thing is, as a machine learning engineer or as a data scientist, you are not just doing a simple model training or changing the architecture of the model, but you should also be okay with doing a lot of legwork like the data quality.

Why I'm coming to this is say, for example, you have a vendor was actually doing the data quality for you, right?

Once they do it and when it comes back to you, you have to actually check every single data that you have.

And once you feed it to the model, see how the response is like.

And then if there are issues, which there will always be, you have to go back to the vendor and tell them because you have the maximum context and they don't have it, right?

The context that you have, they don't have it.

So the high quality data starts with you and ends with you.

Other guys are just supporting you, but you should never assume that they are doing your job and you don't have to.

It's like, it's your job in the end.

So that's where high quality data comes into picture and the involvement of the stakeholders here.

Very interesting.

You know, earlier you, you know, you talked about the fact that you've done some really impressive

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Unveiling the Future of AI in Computer Vision with Sumedh Datar

things, interesting things in the healthcare space.

I'd be curious if you can kind of discuss some of the ethical considerations that come into play when, you know, deploying computer vision technologies, especially in sensitive areas like healthcare.

Yeah, so what has happened and even I struggle on a day to day basis, right?

When a model gives a certain output, it's very hard to understand why it is giving a certain output, right?

They take, for example, Tesla, they are purely running on visual sensors, that's how they, right?

And then they had the, they had an image of a pickup truck which had like wood logs or something and that was being shown like a traffic signal, right?

On the display itself, on the display itself.

Visually as a human, you see it like, yeah, it's a wood log and it's a truck, it's not a signal, right?

Since you, since it's predicting as a signal, the decisions are according to that, like, oh, is it green? Is it right? Right?

It's, right.

So deep learning models have this particular problem wherein it's so hard to interpret why the model gave a certain output, right?

Oh, coming back to the healthcare angle, it's very hard to say someone that, hey, I think you have cancer, right?

It's, it's very hard and you need to have like ample amount of information to back it up, right?

Yeah.

So, a model interpretability is something that there's a lot of research going on in this space wherein you can clearly visualize why the model actually

gave a certain output, like you can visualize the layers of the neural network and you can be like, hey, I'm not sure, but this is what the neural network says.

And this is the reason I feel the output is somewhat like this, like the reasoning like this helps.

And second thing is it should be always backed by subject matter experts, especially in the healthcare space, right?

Three people use three different answers, right?

And the most specialized the doctor is the answers are more different.

But how do you get that, right?

So it's always good to have an AI solution with the support of a doctor wherein you're actually helping the doctor, but you're not really taking over the doctor because it's like way far in the line.

I don't think AI is there yet, but some areas where when you have a solution, you should have explainable techniques along with it.

This is why the model told this is the answer.

That's when I think you'll be in a better place rather than, yeah, okay, this is the probability point nine and this is what it is.

Okay.

Yeah, that makes a lot of sense.

What from your experience do you think are some kind of emerging trends in computer vision that you find particularly exciting or promising today?

The emerging trends are one is like the self supervised learning wherein you just give the data and the model does the labeling for you.

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Unveiling the Future of AI in Computer Vision with Sumedh Datar

That emerging field and of course, the JNA AI is the next big thing, right?

So you already have the chat GPT wherein you give a question and it gives you an answer, right?

And there are a lot of vision models wherein you just give an image and it gives the description of it, right?

Can you pull that with chat GPT and build a better model, better models and eventually better applications, right?

Of course, this is the emerging trend.

Very, very interesting.

Yeah, that's it's exciting stuff.

It's exciting to be kind of watching is unfold so fast.

Everything's advancing.

I'm wondering if you share some insights into some of the specific challenges and solutions associated with item recognition in retail.

I know there's a lot of different kind of challenges and nuances there.

Right, right.

So one thing is in the retail space, just the sheer number of products, right?

You have so many products.

It's just so hard to do the recognition.

So how do you do the recognition for that?

And when it comes to visual recognition, when it comes to supervised learning, what you're doing is you're basically doing prediction from the prior data you have.

Now, what is happening is when you run some kind of a campaign, when the packaging changes, right, the model results also change automatically.

Interesting.

Oh, man, that's so complex.

How do you handle these situations?

And then the third thing is like the sizes, right?

Like you have so many different sizes, they all look the same.

And with vision, you can do only so much.

So that's when you need to think about other signals.

Like how do you handle like the sizes?

Do you want to bring in like point clouds and things like that, which helps you in making better decisions?

These are all the challenges that we face in the retail space.

Very interesting.

Yeah.

And for people listening that are trying to understand like kind of what we're talking about, there's a number of different retailers and stores where essentially they have like a little shelf and you can come and put all of your products on the shelf.

They'll scan them, analyze them.

Then they'll just tell you what the price is.

It's kind of like self-checkout that, you know, you do a Walmart, except instead of scanning all the items, you can just put all the items on a shelf.

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Unveiling the Future of AI in Computer Vision with Sumedh Datar

So there's really cool things happening with computer vision right now.

But yeah, like you're imagining, I imagine that just throws everything for a loop when, you know, for example, you have a can of Coca-Cola, you know what that is.

But if they do a new campaign and there's a new packaging on there or they change the size of the can or the bottle, all of a sudden, like you got to re-figure out what every single item is as it's constantly changing with thousands of variations.

I can't even imagine the headache, but I'm happy.

Right.

You're tackling it because it's a big challenge.

When you do something like this, have you kind of measured the success or the effectiveness of a computer vision project?

Yeah, this is something that's super important, right?

Say in your lab or in your environment, it's working really well.

Once it goes in the wild, it's not working well, right?

So how do you handle the AI adaptability, right?

How many people are ready to accept the technology?

How many people are using it regularly, right?

So this is when MLOps and model monitoring, they come into picture and you need to have metrics, you need to have dashboard and continuously keep monitoring and seeing what's going on.

Is it doing really well, are the models off, are people not happy, like what's going on, right?

So data monitoring, model monitoring, these actually play a big role and the metrics are very different.

It's not just the regular accuracy metric, but it's more on a private level and the metric should be translated to business, right?

So the metrics are very subjective and the definition of the metric is not very straightforward.

So it's very tailor made to the use case that relates to business.

So tracking these metrics actually help you know how good or how bad the models are doing.

And yeah, AI adaptability is not easy and we need to find different ways of monitoring to make sure we have a successful product.

Very interesting.

Really appreciate you coming on the show today.

I would love to ask you as kind of a question as we're wrapping up here.

You know, what advice would you give to aspiring engineers who want to specialize in computer vision and deep learning today?

Yeah, based on my experience, coming from a non computer science background to a computer science background, what I really noticed is the algorithms and data structures play a big role because you are coding the equations, right?

So you need to know your basics well, having an expertise in one language is enough, but being an expert in that language is the most important thing to know the syntaxes.

That's the most important thing.

And when it comes to deep learning, machine learning, I think the Stanford CS231N course is the starting point because they actually teach how the algorithms work by actually coding from scratch without using any libraries.

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Unveiling the Future of AI in Computer Vision with Sumedh Datar

So in the process, you learn all the internal components that is hidden when you just use a library. So that's where your learning is maximum.

I think these two are most important thing.

And last thing is the software engineering aspect once.

So nowadays what's happening is it's not just on the research level.

It's on it's also on the application level, right?

Like, in the end, it should go to a customer or it should go to someone who has to use it or he she whoever it is, they have to use it.

Right?

So how do you see it from a customer's lens?

And I think there's a software engineering comes into picture having strong software engineering skill set which involves like a full stack development, a rough knowledge of the front end back end kind of gives you an end to end view of how the product looks like.

I think these are good starting points to get into this video.

Really interesting, amazing and great advice.

So really appreciate you giving that.

You know, if people want to contact you or ask you questions or, you know, connect with you, what's a good way for people to find you?

Yeah, I think they can find me on LinkedIn and ask me any questions.

That's totally fine.

Thank you so much, Sumed for coming on the podcast.

I really appreciate all of your insights and everything you've shared to the audience.

Thank you so much for listening to the AI chat podcast.

Make sure to rate us wherever you get your podcasts and have an amazing rest of your day.

If you are looking for an innovative and creative community of people using chat GPT, you need to join our chat GPT creators community.

I'll drop a link in the description to this podcast.

We'd love to see you there where we share tips and tricks of what is working in chat GPT.

It's a lot easier than a podcast as you can see screenshots, you can share and comment on things that are currently working.

So if this sounds interesting to you, check out the link in the comment.

We'd love to have you in the community.

Thanks for joining me on the open AI podcast.

It would mean the world to me if you would rate this podcast wherever you listen to your podcasts and I'll see you tomorrow.