

Congratulations to you Mr. Raghavan for that.

Thank you so much for joining us.

Over to you.

Hi everybody.

How are you?

I am not hearing this at all.

It's like a post-lunch energy downer or something.

Let's hear it.

Are you guys awake?

Alright.

You better be because we have a superstar guest here.

You heard the \$41 million.

I didn't hear honestly anything that she said after that.

So we're going to ask for about \$40 million from him by the end of this conversation.

But let's get started.

I want to introduce Vivek and Pratyush, his co-founder, who's not here.

We wanted to start with playing a video of what Open Hati does.

I encourage all of you to go to the website [server.ai](https://server.ai) and check it out.

But let me start by introducing Vivek.

Vivek is a dear friend and he is very very modest, one of the most modest guys that I know.

But his personal journey Vivek, you got a PhD from Carnegie Mellon, you started and sold a company to Magma and Vivek and I moved back to India, we were both in the valley on the same day actually.

And you've been in India for the last 16 years.

And what most people don't know is your journey at Aadhaar.

He spent 13 years selflessly at Aadhaar.

Nobody would have heard of him.

But he was a pioneering technology visionary behind Aadhaar, which we all take for granted today.

So please give it out.

So honestly, when I think of selfless service, truly selfless service, I always think of the way.

And since then he also was at AI for Bharat, which we're going to touch on with Pratyush's other co-founder.

Pratyush had a PhD from ETH at Zurich.

He was at IBM Research, he was at Microsoft Research playing a key role and a faculty at IIT Madras and at AI for Bharat.

So that's a little brief introduction about them.

These guys are modest, modest engineers.

So they don't toot their own horn.

So forgive me for tooting their horn in this case.

But let's jump right in about the money.

Funding, 41 million bucks man.

That's a lot of money, right?

Every entrepreneur here is saying, what the hell did these guys do?

What did the investors see to write such a big cheque?

I think it's a new trend of what's going on in India I think that for the very first time I think investors have looked at you know let's try and build something deep tech out of the country and let's try to figure out how to build something as a foundational technology out of the country and that's really what's really exciting you know and I think that about you know as

While I was mentioning for the past 15 years, I've been kind of working in both the digital public infrastructure and non-profit kind of things.

But when this whole thing of generative AI came about, we said, OK, how can I actually make a difference in this space?

And I said, maybe this is the opportunity to actually come out and really build something.

And the only way that we realize that we can do this actually

in the private sector and I think that's that and then we when we went out there and we said we want to build something which is a continuation right I mean and fundamentally the question is the reason of what we want to do at Servam AI is we want to basically make generative AI

available and accessible to the people in the country and that's the intent and when we said that we want to do this there was a resonance in the investment community and I think it's a responsibility to really to show that something like this can be built out of India so we see that as confidence and a responsibility and I also hope it's a trend that there are many more people like us who are backed because if you look at it maybe it's a large number in a

you know in the Indian context but in the global context I think there is just there should be many many more entrepreneurs who are back to do things in India.

I'm going to come back to the many more entrepreneurs I'm obviously going to ask you about Bhavesh's Kuthrin so we're going to come back to the question but again \$41 million all of what you said you know \$2 million you know that's a good amount of money for a startup which you know

which has not yet built anything

What are you going to do with all this money?

I can solve the problem.

I can have a perfect solution for the problem.

I think in the last week I've got lots of calls from lots of people telling me how I can do it.

I know you first.

We landed in the country the same day.

I'm in front of the queue.

No, but honestly, I think the key thing in this is, is to putting together an amazing team.

And we actually have an amazing team, but believe that it is talent that will drive this kind of thing.

And so it is, it is to get, get talent.

And of course the other thing is compute.

This is extremely expensive compute wise to actually do these kinds of things.

And I think that those are the primary things that, that, you know, use this one.

Okay.

I'm computing in my own head as an entrepreneur.

Talent, okay, you have like 20, 50 people.

How much are you paying these guys?

You won't touch on that.

But let's talk about what you guys actually built.

What is open Hati?

How would you explain open Hati to many people here who might not have known about it?

So I think OpenHAPI is, so first of all, we come from, I personally come from the open source ecosystem and also from the DPI ecosystem.

So we believe that for this to work, we need the ecosystem to be successful.

And as a result of that, one of the first things we did was

hey, there are these open source large language models that exist, right?

I mean, everybody knows about the Llama family from Meta, there are others like Mistral, there are a bunch of open source, you know, large language models.

And then we said, is there any way that they can existing open source model and teach it language

skills, right?

And that is really what we said, can we do something like that?

And is this a relatively frugal way of actually making models work in diverse languages?

Because the truth is still today, if you look at the amount of data and knowledge, still English dominates these things.

And I think that how do you actually take and make it, understand Indian language, understand Indian context, and all of those things in actually an efficient way.

And therefore this was an attempt to do that.

And it's an open-hearted, you know, it's currently based on the LAMA 7 billion model, but we'll be releasing many more models in different languages, different sizes, and things like that as part of this series.

And of course, you know, we will be building further models on those and doing other things to actually, and we'll also have endpoints that people can use, but it's not, it's definitely, you know, something that can use two things.

And that's the essence of what is open happiness.

So what does it mean to people in the audience here who are either doing their own startups or a business or developers?

how should they look at one of the important things that we are doing is we're not just building models we are also building a platform a platform for developers where you can actually use

a combination of various different kinds of models, some which are from us, some which are open source, some which may not be open source, and actually to pull together and figure out how to deploy generative AI applications at scale and understand and evaluate their performance.

And that's something that we are planning to do.

And this platform is, you know, in the next couple of months will be coming out there, be available to developers.

But of course, those who want to start with the open source things and act with that, of course, please go ahead.

That's phenomenal.

How does it compare to OpenAI itself or Google?

See, at least the things that we are doing now, right?

I mean, one of the things that when we thought about building server, we said we want to build a full stack general API company and different people and our understanding of stack is that we need to know how to train models from scratch.

We need to know how to kind of figure out how to deploy models to solve real world mistakes.

And we need to play in the ecosystem to make sure that we can actually deploy population scale applications.

So we were thinking about all of these things.

But still the models we were talking about are fairly small models.

They are fairly small models.

The 7 to maybe up to 70 billion kind of range we're talking about.

While these models like OpenAI and they're obviously much bigger.

But we want to understand the techniques and be able to build that muscle to do all of these things to make it available to people.

Now those models are I mean as I said you know I think that there is space for all of those things and I think as when Sridhar was talking about

earlier in the day, we believe that these smaller models can do many many kind of domain specific tasks extremely well, probably even better than the larger models and that is really one of the key areas and so therefore the value of these kinds of things, we are not aiming in these more set of models to build any AGI, that's not our goal.

Our goal is to make things that work extremely well for domain specific use cases or increase accessibility to language and obviously all of this unique to India.

What is unique about India?

Is there anything special in our ecosystem that makes small models focused with Indian languages more suited for our problems?

So I think that, I mean, there are quite a few things that are unique about India, right?

The first thing is I think that we are a voice first nations.

Therefore, I think voice has to be the core to doing things.

The other thing, of course, India is a cost-conscious country from a cost perspective.

I would say that there are lots of interesting use cases where you can use OpenAI and the cost structure works depending on your application.

But when you want to scale things to a massive level and make it work, then you have to figure out how small amounts work.

So that's something that is also specific to India.

The third thing which is specific to India is really the success that India has had in building all this digital public infrastructure.

When you add the AI layer on top of it, then you can actually get dramatic, you know, dramatic, I think, multiplicative combinatorial effects based on doing things like that.

That's a phenomenal point.

Like, you know, it's like DPI to the power of AI almost in some ways.

It's a part of Adhar, building Adhar, no better person than you.

So, in summary what I am hearing is small models specialized with trained with big specific language data suited for Indian problems at a compelling cost point will be suited for us.

We are not solving some world or some complex problem.

We are solving

basic problems specifically focus with on voice of multiple languages that is what you see as the future am I paraphrasing this correctly no yeah so I think that certainly I mean a voice at Indian languages are an important part of our strategy but we will be building you know custom models to solve various other kinds of problems as well right it's not just limited to I think in different domains

domains making building things based on unique data that enterprises have.

So that's something.

Fair enough.

So coming back to the elephant in the room, no pun intended with open Hati.

What about Bhavesh Atirwal and Kruthiram?

What is your take on it?

I think it's great.

I think it's wonderful, right?

I mean the fact that the technology, AI is so important that we need multiple people working on it.

The fact that there are other people thinking is actually validates that this is an important problem to be solved.

And I think that we need everybody to come together and do that.

So I really welcome that.

I think it's great.

And I think that there will be different people will have different takes as to how to solve this kind of problem.

And hopefully as a result of that, the entire ecosystem benefits.

I have one more question and then I want to talk about some of the predictions that you boldly made.

So Vivek, I usually ask people about what do you think the future will be and everybody usually hedges.

I ask Vivek, what do you think is going to happen by December 2024?

What do you think sitting in this room one year later we can expect?

And he made three bold predictions.

So I want to talk about that.

Before that I have one last question.

What are the top three applications that you think are relevant for India?

You heard Sridhar talk about medical.

Quick summary, what do you think the top three apps are for India for AI?

So I mean I think that as I said things like education and medical are clearly areas where I think that can be leveraged.

The whole idea of

or these kind of the DPI aspect of it is another major application where things can happen.

And here I'm talking about country specific work.

And I think the whole idea which you also talked about was the concept of software, right?

And I think that and clearly we have a very large software industry and how to reimagine those things.

This context is also something that's coming.

Fair enough.

Are you guys ready for Vivek Radwan's bold predictions?

Yeah?

no I'm not hearing any yes this is like a big deal he's like one of the smartest guys that I know he wants to make three predictions you don't want to hear it all right so I asked him what do you think you know here later what what do you think we can expect and he came up with three things and usually people give very blah answers when they ask questions like this because they don't want to be caught wrong not be vague, vague is bold so he basically said three things and I want to list out the three things and then

ask him about it.

So number one he says I would prefer to talk to an automated customer service than a real person because they give me a better answer.

That is Vivek Ragwan's prediction number one.

So number two is that when everybody is talking about a GPU shortage, Vivek predicts that there will be a GPU glut in India.

He thinks there will be too much GPU.

So, if you want a short in-media stock, it's a good time.

And number 3, which was extremely unexpected, he said some companies will suddenly die.

not what I expected.

So do you want to quickly talk about each of them, why you just came up with these and then we'll throw the audience questions.

So I don't think I quite said it the way that Walai was kind of pointing, but it's interesting.

But I think the first thing that we said is I think that, and I don't think that this, I think there will come a time

when you know in areas of customer service etc.

when you want to do something very specific.

Today, when you call some kind of a bot, you actually end up, you mostly try to disconnect the call or you're extremely upset that you're talking to a bot.

But I think that there will come a time, and I'm predicting this sooner than later, that you will actually get better responses from the bot than what the human representative, at least the average human representative that people talk to,

could give and I think that that's just I just said that there will come a time where you know it's not a human you're talking to but it's probably more likely to solve your intent that's just something that that I think that could happen okay definitely controversial but we let it go what about the GPU glut

I think that the fact that there is a tremendous shortage right now, I think that shortage will ease because that is how the cycles of things go.

I think the fact that there was such a severe shortage last year basically caused a number of different players to ramp up in various kinds of forms and I think that will always go in cycles.

But we may find out that there are many more interesting problems that people will be able to solve.

I still remember, you know, we were at a JNEI event in Bangalore and we were talking to people and we said, you know, how many people have access to, you know, four A-hundred?

This was the question that I had asked.

And nobody in the room.

And these are all extremely enthusiastic JNEI people.

And nobody had access.

And I think that thing is going to change.

You will be able to get these kinds of things.

And people who want to hack and do things will have access to these things without having to write a check.

Vivek is also a semiconductor guy.

Before he went into Adar.

So I would take his predictions very seriously.

So I don't know what I would do.

I would sell my previous stock.

I would not do that.



But that's not what I said.

I don't want to blame you for this.

But the third one is pretty strange.

Companies are born, companies die.

But you said some companies will suddenly die.

What does that mean?

No, I think, see, I think the interesting thing is, and I think that it comes back to the fundamental nature of AI.

AI is a tool, right?

And you have to use that.

And you have to use that within your business process, right?

And how AI is being used.

And so, and what's going to happen is that, I mean, I think this is true with, you know, when someone said in terms of, you know, people, they said that the people who leverage AI

will be more effective than those who don't leverage AI.

And that will speak for organizations also.

Organizations that leverage AI fundamentally in their core business processes will be more effective than those who don't.

And I think that's the thing.

And you won't know the difference until one day it becomes too obvious and it will be too late.

And I think that's the reason why everybody needs to think about what it means for your business.

Because everything will be fine.

Everything will be fine and one day somebody in your either competitor in your space or something brand new coming into your space will be reimagining your business process completely.

And at that stage you will find that it's a very big, very tall business.

you know mountain to climb and that's why I think it's important for both people and entities to think about how they will you know they will upgrade themselves or they will modify their first system.

That's a very nuanced answer everybody here running a business should really think about it because life will be the same

And then suddenly, suddenly something will be a step change.

Vivek, I have a few more questions, but I'm sure the audience has a lot of questions for you.

So how are we doing on time?

OK.

So that's OK.

A lot of questions, so I'd love to.

Is there a mic that we can?

Thank you.

My name is Karthik.

I work for IT service industry.

so you're saying that you're working on LLM sorry fine-tuned LLM on top of llama my basic question fundamental question is we don't have a foundational model for India most of the models are basically using English or those kind of things for example even Andrew was talking about

tokenizers and things like that.

So are you working on anything like that or you want to use mostly the existing models and run on top of it?

Good question.

You asked a cherry question for us.

No, I think the interesting thing is that if you look at, and we have actually a blog on this, on our website, I think one of the things that we actually built

a customized tokenizer which actually fundamentally changes the cost of some of these generations in India languages and I think that we are not just we are actually we are leveraging the existing training but we are doing what's known as continual pre-training which actually but having said that you know I think that once we have to figure out where is the data to train an extremely large model from scratch and some of those things are things which will happen

over time.

But I think that yes, I think that we will be doing various kinds of things.

But the interesting thing is that if I want to change the accessibility problem with an existing open source model, how do I do that?

And that's the problem that we think we have solved.

And it's going to be the heart of this.

It's extremely well explained in the blog, even I could understand.

Hi, I'm Prashanth.

I work for a fintech company.

My question is like, unlike China, we never had a consumer facing application coming out from India and in web one, web two, crypto and on.

Why do you think it would be different this time in like AI?

Because will the DPI and other things will serve the same purpose what the Great Firewall did in China?

Or do you think like, because AI is a strategic sector,

no outside country can work in NASA projects, maybe all government content will go to them.

What is at least the mode here for any income?

So, I think the question is,

I don't know the answer to these questions right I mean I think that it's difficult to predict but I do believe and as I am repeating that the combinatorial effect of reusing Gen AI at a large scale in addition in along with the DPI work that we have done in India

will have people and I and I think that in the end it is the intent is that people need to be able to use it and they will vote by things that are useful for them and if that doesn't happen you're right that that I think that we have to figure out what is the mechanism of delivery of apps right I mean how where do Indians consume

I am so sorry, but we are out of time.

Vivek will be outside.

So he will be able to answer the question.

We have time for one last question.

Can I just take one last?

Yeah.

Thank you.

Thank you.

I am Manish Kothari.

I am from ISBR Business School.

Good that I got a chance to ask you this question.

During lunchtime, there were a few of our educationists whom we were talking about.

There was one from school and we are from the MBA institutions.

We were thinking of these present generations.

How do we get them into what you are doing?

There is one thing that they have been regularly that the concentrations that they are working on but artificial intelligence and getting into this, getting them into their academics and making them a part of it is very important including the trainers who train them.

Making them future ready into what you are doing is amazing and the speed that with which is growing.

It is calling for a lot of training that needs to be done.

Can you from your angle through some light on how we could make them future ready?

How these people who are who are management graduates and from schools who are coming out?

How do we get into this part of technology that you spoke about?

This is really a challenge because I think everyone will need to understand at some level what this technology does and I think that we have to rethink how we get everyone into this and that this kind of education has to be at many different levels, right?

There are from a core set of having people who are extremely good

And there you don't need as many, but then there are vast numbers of people who can actually leverage these tools.

By the way, the most important thing about, and maybe that's part of what makes an LLM interesting, is that how you use it, your mileage varies by that.

And to understand how to actually leverage this in an interesting way is something that we have to widely teach many, many people.

And because asking the things in the right way and having the right kind of applications will make a difference in how people get used to this.

Awesome.

Thank you.

Thank you very much Vivek.

Very good luck to Sarvam and good luck to India.

I think it's going to be a lot right on your shoulders.

Thanks, Vala.

Thank you Mr. Raghavan.