[DAISY THE GREAT](https://www.facebook.com/daisythegreatband/): 00:01 [Built My House On Hollow Ground](https://www.youtube.com/watch?v=z-1sC1lkmKw)

CRAIG: 01:50 Hi, this is [Craig Smith](https://www.nytimes.com/by/craig-s-smith) with a new podcast about artificial intelligence. As regular listeners know, besides understanding basic research in the field, I'm interested in understanding what's happening in AI research in various parts of the world. I've done episodes about [China](https://www.eye-on.ai/podcast-102), about [Russia](https://www.eye-on.ai/podcast-001), about [Europe](https://www.eye-on.ai/podcast-008) and [Africa](https://www.eye-on.ai/podcast-105).

CRAIG: 01:52 This week I look at India, which, with its massive population, fast growing economy, English language education, and large supply of brilliant researchers and engineers, one would expect to be competing with China and the U.S. for dominance in the space. However, it's not. Many of its best and brightest technologists live abroad where they populate elite research institutes and lead some of the world's [biggest tech companies](https://www.gadgetsnow.com/slideshows/meet-9-indian-origin-ceos-ruling-the-technology-industry/photolist/53708176.cms).

CRAIG: 01:52 To understand why and what's happening in India, I talked to [Partha Talukdar](http://talukdar.net/), a professor at the [Indian Institute of Science, Bangalore](https://www.iisc.ac.in/), where he's a founding director of the [Machine and Language Learning Lab](https://malllabiisc.github.io/). Partha, the son of bankers, grew up in the northeastern Indian state of [Assam](https://en.wikipedia.org/wiki/Assam), but did his undergraduate degree at the [Birla Institute of Technology and Science in Pilani](https://www.bits-pilani.ac.in/). He moved to the University of Pennsylvania to pursue his PhD and, together with [Tom Mitchell](http://www.cs.cmu.edu/~tom/), Partha was one of the architects of the [Never-Ending Language Learning](http://rtw.ml.cmu.edu/rtw/) system at [Carnegie Mellon](https://www.cmu.edu/), which has been poring through hundreds of millions of webpages nonstop since 2010, learning semantic relationships and accumulating knowledge. I'm delighted to have Partha on the show.

CRAIG: 01:55 I talk to a lot of people about China. Everyone is obsessed with China. India is the other massive population base.

PARTHA: 02:04 Sure.

CRAIG: 02:05 You go to conferences, you look at the CEOs of tech companies, there's an over-abundance of Indians. Why don't we hear about AI in India the way that we do about AI in China, or even Russia.

PARTHA: 02:20 Right, right. So, I mean there's a long history of AI in India, so the courses and the research, I think like the first courses in AI in India started like 1960s or so. And then there's a long history in doing research in artificial intelligence. So, I guess, kind of, like, things have been, kind of, like, going on, but with the resurgence and the scaling up in various parts of the world that we have seen is, kind of, like, still work in progress in India.

CRAIG: 02:48 Right.

PARTHA: 02:48 I mean of course you will see Indian researchers all over the world, kind of, like, contributing to AI research. I guess like your question is, kind of, like, AI research coming out of India.

CRAIG: 02:59 Yeah, yeah, exactly. Because a lot of the researchers are attached to international institutions, particularly American institutions.

PARTHA: 03:07 Right, right.

CRAIG: 03:08 So, one question is about the basic research. The other is the startup ecosystem in India is again, I know that there are things happening in India and certainly there's some very famous companies that have come out of India, but not to the extent that you hear about things out of China or the United States.

PARTHA: 03:29 Yes. I guess there are, kind of, like, the multiple aspects and dimensions that we can look at. So, in terms of just like academic research, so there are good pockets of very high-quality work that's being done in the country. But the volume of it, right, and the number of it is not, kind of, like, proportionate to the size of the country. So, we have, kind of, like, work to do in terms of scaling that up. Now in terms of the student population and academics, so there is, I mean India is a very young country in terms of population and there are like various statistics. About 50% of the population is less than 25 years of age now.

PARTHA: 04:07 And uh, the average is going to be also like, you know, like next few decades, going to be on the younger side. So, there is strong appetite among the student population to learn AI to, kind of, like, contribute and like become - tool themselves up in this particular area. So, uh, more than, kind of, like, a top down scenario, like in terms of planning in AI in the country, I would say kind of like, there's much more activity in the bottom up fashion, much like lots of other things in the country. So, there is a strong interest. So, in terms of like in finding data science talent, when I talk to industrial partners they look to India. Again, as you probably know like Bangalore is a strong computer science IT hole, which has been, kind of, like, in traditional areas but increasingly even to find data science talent, people are looking to Bangalore, kind of, like, as a go-to destination to find them.

PARTHA: 05:00 As you could imagine there is a crunch in talent in those areas. So, there is lots of bottom up interest and upscaling of people, both new students who are, kind of, like, graduating but also within the IT industry, like a large number of engineers who are, kind of, like, traditionally doing the regular software development and now they are interested in coming into AI and from that perspective there are lots of support and ecosystem has been, kind of, like, developing to upskill themselves. So, that's, kind of, like, the education and training part of it. Many of the major universities have also introduced a program specifically focused towards AI. For example, in my university in digital science from this year we are starting an M-tech especially focused on AI and there are a few other places like that as well.

PARTHA: 05:53 Now on the startup and the industry side of things, I mean there is also lots of like young companies and startups that are coming out, which are AI first, right? So, where, kind of, like, AI is not an afterthought but their main product itself is focused on AI and like heavily relies on that. Myself included. So, even though I have a faculty position, but also to translate some of our research, I also founded a company on my own - a startup called [Kenome](https://www.kenome.io/), which is trying to help enterprises make sense of unstructured data through [knowledge graphs](https://hackernoon.com/wtf-is-a-knowledge-graph-a16603a1a25f), things like that. So yeah, so there is lots of activities around there. I would say, kind of, like, in terms of the fresh startups and relatively large-scale startups which have been around for a few years and have been, kind of, like, collecting data by offering various services, they are also increasingly more interested in utilizing AI to improve their product. And there's lots of interest around that.

CRAIG: 06:46 Yeah, and we're talking about machine learning specifically.

PARTHA: 06:49 Yes, yes.

CRAIG: 06:50 Yeah. Are there, you know, we all know Baidu, Tencent, Alibaba, Amazon, Google, Uber, Apple, you know all of the big tech firms in the U.S. and China. Are there the AI research giants in the private sector in India and who are they?

PARTHA: 07:10 So, in the industry in AI is represented, industry in India is represented by these, kind of, like, large multinational organizations. Pretty much you can, kind of, like, name any company. They would have like a strong tech presence in India. So, apart from that there are, kind of, like, the regular Indian conglomerates and large businesses that have traditionally been around and have also some technical branches like say [Tata Consultancy Services](https://www.tcs.com/), [Infosys](https://www.infosys.com/). These are like large-scale IT services companies. And then you have the startups that have, kind of, like, are homegrown and competing with many of these multinationals that you mentioned. For example, in the e-commerce space we have [Flipkart](https://www.flipkart.com/) which was recently [acquired by Walmart](https://techcrunch.com/2018/08/20/walmart-flipkart-deal-done/). Then there are like other upcoming SaaS based startups like [FreshDesk](https://freshdesk.com/general/the-freshdesk-story-blog/) and food delivery services called [Swiggy](https://www.swiggy.com/). So, these are, I guess, kind of, like, in terms of size, still not as large as those, the Chinese counterparts that you are talking about, but they are, kind of, like, the dominant players in their respective areas. So, they have been through their services and offerings have been, kind of, like, collecting data because in order to do machine learning and AI you need, kind of, like, data as a first class object and now that they have been, kind of, like, running for a while collecting all this data there is, kind of, lots of interest and investment in utilizing that data to improve their services.

CRAIG: 08:33 Yeah, and that's something I wanted to ask you about because certainly I don't know if it's as much as China, but it's very well known, the massive data collection activities or enterprises or whatever the word is, innovations in India, the, the fingerprint, the biometrics and yeah, so there is tremendous data being collected that's held presumably by the government. Is that made available to private enterprise or is there a counterpart like Facebook or Amazon in the private sector that's also sweeping up huge amounts of data.

PARTHA: 09:10 So, the [Aadhaar](https://en.wikipedia.org/wiki/Aadhaar), the biometric system that is, kind of, like, maintained - developed by the government and it's maintained by the government for the identification and authentication purposes - those are I guess, kind of, like, available in some limited capacity to certain, kind of, like, sectors. For example, for banking, if you want to do some sort of authorization in some limited way you can get validation whether a particular person's identity is same as what they are saying that they are, uh, whatever they are kind of proposing their identity to be. So, they can validate in those limited contexts. But there are, kind of, like, lots of discussions and litigations also going on in terms of like how to utilize this data set. And, of course, privacy remains a very strong concern on that. Yeah.

CRAIG: 09:58 Yeah. The government, you were saying, it's bottom-up. China is decidedly top-down, although also bottom-up, but, but it's got a very focused top down approach. The U.S. has a top-down approach in some sectors largely coming out of the defense sector through DARPA and things like that. But there isn't the kind of national strategy that exists in China. Is there in India that kind of a national strategy, top-down, to drive ML development?

PARTHA: 10:33 Right. So, there was a policy document, or a suggestion, a draft was put up last year by what is called [NITI Aayog](https://en.wikipedia.org/wiki/NITI_Aayog) as the [AI policy for India](https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf), so it's called like AI for all. They have identified certain sectors to, kind of, like, focus on that. Also focus more both on the basic research and, kind of, like, the translational aspects of that. So yeah, so that just came out last year and I guess, kind of, like, various policy positions that were being made in that particular document are currently being worked on to implement those in terms of ...

CRAIG: 11:06 Yeah, in China the government, I mean is still a very much command and control economy, the government has tremendous financial resources so they spend like crazy when they focus on something like this. I mean they built the [world's largest high-speed rail network](https://www.guinnessworldrecords.com/world-records/largest-high-speed-rail-network) in record time. So, there's that money that goes down, but it also triggers regional governments and municipal governments to come up with their own programs and they start spending. So, there's a tremendous flow of capital through the government into research and development and the startup ecosystem. Is there money flowing down from the top like that in India?

PARTHA: 11:49 Right. So, I mean traditionally funding in India or at least like you know the well-recognized universities are, kind of, like, well funded by the government and increasingly there is funding from the private sector as well. So, funding for, kind of, like, regular research has not been that much of an issue. And also, the students are covered in terms of fellowships and the tuition by the government. But I guess like in terms of scaling it up significantly, right, we, kind of, like, more resources and investments are necessary. I'm hoping that some of these, kind of, like, policy and works in progress, will, kind of, like, contribute towards that. I mean there is an effort to have like a nationwide upgradation of the infrastructure for computing to have, kind of, like, a [grid style super-computing or distributed infrastructure](https://en.wikipedia.org/wiki/National_Supercomputing_Mission) so that like computing is not the bottleneck in terms of whoever wants to do innovation in this space.

PARTHA: 12:45 But I think we need more researchers and more people already skilled in this area to kind of like, you know, find their base in the country and then, kind of, like, form groups locally and then propagate that. So, while there has been most of the students in India, kind of, like, go abroad to, kind of, like, for their higher studies, although in the recent years there has been, kind of, like, a trickling back effect. I guess like that we need to accelerate to, kind of, like, create opportunities for people who are abroad who are, kind of, like, already experts in this area to come back at least partially or fully and then like, you know, contribute towards the ecosystem in the country.

CRAIG: 13:23 And you're an example of that.

PARTHA: 13:25 I am an example of that, but I kind of came back out of my own, kind of, like, personal reasons, but I think we should also create professionally attractive opportunities for people who belong to this category. From what I have heard China has done quite aggressive and attractive positions of this kind for people to come back. And I think we need to have more along those lines to create those kinds of opportunities.

MUSIC: 13:25

CRAIG: 13:55 I mean one of the things that's drawn people back to China, uh, besides the national policy is a booming economy. Does that play a factor in India?

PARTHA: 14:06 I think that has been one of the positive factors in terms of the [booming economy](https://www.cnbc.com/2019/04/18/how-indias-economy-went-from-weak-to-fastest-growing-in-the-world.html). And So, there is, kind of, like, lots of opportunities to both from, kind of, like, a scientific perspective but also in terms of business opportunities, right? Because India maybe has a little bit of a different setup in terms of its diversity. So, there is like huge linguistic cultural diversity, but the problems in terms of education, healthcare because it's, kind of, like, still a developing country. So, all of those things are also a major, massive problems, right? So, one of the big opportunities there is how we can say use AI to, kind of, like, leapfrog from the traditional solutions towards like, you know, making say a healthcare more affordable to people in rural areas.

PARTHA: 14:51 So, who cannot say take a day or multiple days off from their work to go to the city to, kind of, like, get them there themselves tested or like get themselves treated. So, how can say AI like you know, do like remote health monitoring so that like at least for the basic screenings, right, they don't have to miss their work and like at least for major ones they can come there. So similarly, for education, finding quality teachers to educate this massive population is not going to be available, is not available now and it's not going to change anytime in the near future. Right? So, having a purely manual approach is not going to work. So again, can, can we use AI like and technology because the need is immediate and we cannot wait a decade or two to actually find a solution for that. So that way I think there is massive opportunities to, kind of, like, have an impact using machine learning AI in these areas.

PARTHA: 15:45 And I think like the policymakers on the government also sees this as an opportunity, but of course more work needs to happen there.

CRAIG: 15:53 Yeah, I mean you mentioned the cultural and linguistic diversity. Has that been an impediment to date in developing India as an AI machine learning powerhouse?

PARTHA: 16:05 Right. So, in terms of training and higher education, most of it is still done in English. So, from a tech training perspective, probably that hasn't been as big of a problem. But what I was saying earlier was, kind of, like, the economic and societal impact opportunity that's there using AI. But at the same time from a research perspective and technology development, having this diversity creates both maybe like opportunities and challenges, right? Because if you have like a deep learning speech recognition model, right? That requires 10,000 hours of speech data to be like that to train your model on.

PARTHA: 16:41 But in India they are like 22 major languages, 22 official languages and 200 major languages and like 1,500 plus other languages, right? So, you're not going to have as much, you know, a high quality, large volume data to train all of these models on, right? So that also throws up new research opportunities, right? So how do you kind of say build better transfer learning model to do this speech recognition better language processing better in all of these languages, right? So now say if you take the example of e-commerce, right? So, I gave example of FlipKart and all. So traditionally, I mean we are given, kind of, like, the business opportunity they started with like the major cities to, kind of, like, serve the customer in those areas. But now they are moving to tier two, tier three cities where English is not the predominant language, right? So, people want to converse and place orders and like discuss using their local languages. So, from a business perspective also the importance of these languages are becoming more important. So, in addition to while the government has been supporting technology development in these local languages over the years, but now I also see increasing interest from the, from the industry to actually support research and develop technologies because there is a business need.

CRAIG: 18:01 Yeah. I had an [interesting episode](https://www.eye-on.ai/podcast-009) with a fellow named Liang Huang from Baidu research about simultaneous translation work that he's doing and as well with Misha Bilenko at Yandex about translation in Russia and for both those countries, opening the English language Internet to their citizenry is very important because so much information exists in English and English is, unlike India, is not a primary language in their countries. It seems that that would be a fantastic opportunity in India in a different way. I mean, you know, rather than opening to the English language world, just interconnecting the languages inside India. So, in Russia, it's Yandex is doing that. In China, it's Baidu and Tencent and Alibaba are the primary forces funding that kind of research in obviously in the U.S. we know who the players are, who would be, in the private sector, who would be funding that kind of research?

PARTHA: 19:05 Right. So, in, I guess, kind of, like, the in case of Russia you have like Russian. In case of China you have like Chinese and Cantonese. So, like a very small number of major languages. So, I guess that helps in, kind of, like, focusing the resources and developing things quickly. Right? In India that problem is a little bit more complicated because the space is much more fragmented in terms of languages, right? While like, say there are like a few major languages like Hindi, Bangla, Tamil and all, but then there's a long tail of, uh, like a large number of languages. I mean, even my mother tongue [Assamese](https://en.wikipedia.org/wiki/Assamese_language), right? So, which is, kind of, like, maybe come towards like the middle of the language pack that has like 26 million plus speakers. Right? So that, kind of, like, divides the focus, right?

PARTHA: 19:52 So in terms of like what do you go after? Right? Um, so in terms of, I mean I, kind of, like, see, I guess like e-commerce players, kind of, like, doing, kind of, like, investing more resources in developing these local language technologies with proliferation of like personal assistance where it's much natural for people to, kind of, like, speak in their local language and get information needs satisfied. So, I see like, you know, these large players because mobile is like taking over, right? And then having these, kind of, like, personal assistants on mobile devices is a natural choice and that's where, kind of, like, things are evolving. So, I see also these large companies, right? So, like an Apple, Google who have like these personal assistants, Amazon with their Alexa, to cater to this larger population base., kind of, like, investing in these types of ...

CRAIG: 20:41 Right. There isn't a homegrown personal assistant yet that's dominating the market. I mean Alice in Russia? Yeah.

PARTHA: 20:48 Yeah. So not, kind of, like, at a well-recognized level yet. Yeah, yeah.

CRAIG: 20:51 Yeah. Are any of the big us multinational tech players focusing on the long tail

PARTHA: 20:58 At least in the, I guess like the head and the torso of that language back. Yeah, so I do see, I mean at least for a first Lexi top five top 10 languages, Microsoft has been investing in developing local language technologies. I hear also like Apple, Amazon, Google, kind of, like, investing in. Google has this like [next billion user initiative](https://www.blog.google/technology/next-billion-users/), which is to, kind of, like, take it to the next set of population which have all of these like additional ...

CRAIG: 21:25 Yeah. Yeah. I just had an [interview with Yann Lecun](https://www.eye-on.ai/ai-articles/2019/6/19/interview-yann-lecun) and he was talking about this problem for Facebook because there are all these language pairs that are not parallel. The data sets are not,

PARTHA: 21:36 Not parallel, Yes. Or comparable.

CRAIG: 21:39 So it's an issue they're struggling with. How do you translate between these languages when you don't have a dataset where things are already lined up.

PARTHA: 21:49 In India, traditionally people have collected these kinds of datasets. For example, like the parliamentary proceedings are parallelly translated to most of the major languages, but all of the data is in silos. And kind of like, not accessible for research, but there are like ongoing efforts to organize all of that data and make it more publicly available.

CRAIG: 22:11 That’s very interesting because again, the U.S. struggles or is beginning to struggle with data access because of privacy laws. And I can see China has kind of free rein because there's this industrial government fusion and not that there are no privacy laws in China, but it's much easier to access government collected data. So that would be a constraint in India.

PARTHA: 22:37 Right? So definitely the privacy is definitely of paramount importance. But that datasets that I was talking about so, these are, kind of, like, public data already, right? So, like parliamentary proceedings and all that you are already, kind of, like, translating and I guess like just now just like storing it and archiving it, but that's, kind of, like, all in like government silos and like you know, not readily accessible. So, the ongoing effort is to kind of surface all of those things, which is public data. Nothing personally identifiable.

CRAIG: 23:08 The reason it's siloed is simply outdated technology or policy or?

PARTHA: 23:13 It's, kind of, like, also the need. So where, kind of, like, people, where whether people were enough people were asking and demanding for this type of data, which clearly is changing from what was before. Right? Yeah. Yeah.

MUSIC: 23:13

CRAIG: 23:35 Another area that China is outstripping, the U.S. is in autonomous vehicles. Again, because policy there is less constrained by the concerns of the private sector or the citizenry. I mean the government sets a policy or backs pilot programs for example, that the U.S. private sector has difficulty getting support for. How is the autonomous vehicle space in India?

PARTHA: 24:05 Right. So yeah, I mean there is kind of like, I mean I don't know whether you have been to India, but our roads tend to be much more chaotic.

CRAIG: 24:13 Of course. Yeah. Yeah.

PARTHA: 24:15 So, so that poses additional challenges for autonomous vehicles. But there is, kind of, like, ongoing effort. So, there is even like multiple startups in Bangalore. So there's a company called [Ati Motors](https://www.atimotors.com/) that actually is trying to develop autonomous trucks, which in certain like restricted environments maybe in like mining and all could be kind of like, you know, used that where it might be dangerous for humans to work in. There have been also like dataset creation.

PARTHA: 24:40 So in [Intel and IIT Hyderabad collaborated](https://www.intel.ai/iiit-hyderabad-and-intel-release-worlds-first-dataset-for-driving-in-india/) to create like an autonomous vehicle, kind of, like, datasets suited towards like the Indian conditions. So, I think there are, kind of, like, ongoing efforts.

CRAIG: 24:53 And the biometric space, facial recognition is that being implemented at a government level the way it is in the US and China and is the research advancing in India?

PARTHA: 25:06 So that data is not available for research and it's, kind of, like, purely used for like a delivery of a service and in a very restrictive manner. Also, the biometrics that's being collected is the iris scan and the fingerprints but not necessarily like the facial recognition. So that, kind of, like, I guess like restricts its use from let's say large scale surveillance and things like that. So yes, I would say while that data set is quite large, but still right now it's being used in a very restrictive use case.

CRAIG: 25:38 And that's a policy issue.

PARTHA: 25:39 That's the policy. Yes.

CRAIG: 25:41 And the same with health data. China has, although the health data is maybe not as high quality as it is in the U.S., it's much more readily available to researchers than it is currently in the U.S. how does that ...

PARTHA: 25:56 Yeah, so in India that's still again very fragmented. So like use of electronic health records is still in its infancy. I would say like maybe in a few major hospitals in the big cities have started adopting EHRs over the last few years. But, kind of, like, the vast majority of the country is still based on paper. There is no, kind of, like, centralized are even digitized versions of those records. There is an effort called like an [India health stack](https://niti.gov.in/writereaddata/files/document_publication/NHS-Strategy-and-Approach-Document-for-consultation.pdf). So, there is something called like an [IndiaStack](https://indiastack.org/) that's being, kind of, like, done for various things.

PARTHA: 26:30 For example, for payments, there is like a unified, uh, payment processing network on which the like various startups can like, you know, or various companies can offer services. So, I guess like one at the government level that I have seen in India is that they are, kind of, like, looking at creating frameworks over which a lot of additional services and offerings could be done. For example, the payment network that I mentioned, it's called [UPI](https://en.wikipedia.org/wiki/Unified_Payments_Interface). That's one example. Similarly, there is a part of the IndiaStack called India health stack, right? So, which will basically allow easier digitization and transmission of information across like, you know, various whoever wants to participate in that network. So that's still in development. So that's not deployed yet. Maybe in some limited capacities. But I guess we first need to get to a digitized state of those information before those could be utilized for research.

CRAIG: 27:31 Are all of these constraints one of the reasons, I mean beyond economic incentives, are these constraints one of the reasons why so many of the top Indian researchers choose to build their careers in the U.S.?

PARTHA: 27:45 Historically I guess that was very much the case because there wasn't like enough ecosystem and opportunities back home to consume their expertise. Right. So, where they could actually like enough flourish and do interesting work. But I think that is changing, you know with development of I guess like not penetration of technology, the support from the government and like at the local ecosystem overall advancing in multiple ways. I think there are lots of interesting opportunities.

CRAIG: 28:13 Yeah.

MUSIC: 28:13

CRAIG: 28:16 Looking probably 30 years ago or 25 years ago everyone was talking about the BRICs, you know Brazil, Russia, India, China. I spent most of the intervening years in China and clearly all of those expectations have become true for China, not for Brazil, Russia and India. Is there a concern in India as I have heard that there is in Europe, certainly there is in Africa, there is to some degree in Russia that they are falling behind the U.S. and China in artificial intelligence or machine learning development?

PARTHA: 28:50 So, I mean I think there is clear recognition that we need to, kind of, like, do more and that we can do more. Right. So, I think there is, kind of, like, a realization of that. Right. And definitely kind of like, given the size and opportunities that are there in front of the country, the level it is at right now, that definitely is suboptimal, right. And I think everyone recognizes that like, we need to do more and various efforts towards that direction is, kind of, like, ongoing.

MUSIC: 28:50

CRAIG: 29:27 Yeah. Let's talk a little bit about your research. You've worked on never ending learning. Which I find fascinating. Have you carried that research back to India?

PARTHA: 29:37 Yes. Yeah, so I started working on never ending learning when I joined Tom's group back in Carnegie Mellon in 2011. So, I was there from 2011 to 2014. then I moved back to India in 2014 to start my faculty position at the institute of science. And we have continued doing that line of research, which I think is, kind of, like, the next frontier because the way learning is practice right now, machine learning is practice right now that's very myopic and you try to learn like, you know, one particular task or one function in isolation and you kind of like, you know, if you're fortunate you get lots of data for that particular problem and you try to learn there, right? So, if you move to something else, you, kind of, like, abandon that completely. And then like start from scratch. So that for certain high value problem that might be okay.

PARTHA: 30:24 But that's very different from how you and I learn, right? So, our learning is not just focused on a few minutes or few hours of training. So, it starts right in the womb and continues till we die. And whatever we have learned in the past, we utilize all of that knowledge to learn going forward. So those are, kind of, like, the goals of never-ending learning. So how we can learn over extended periods of time and utilize whatever we've learned in the past to learn things in the future better. So, the never-ending language learning. So that's NELL with one more L is one instantiation of that never-ending learning, kind of, like, the philosophy of learning. So, once I moved back to India, we collaborate quite a lot with industry and one of the things that came up again and again was like you know whether they could build a knowledge model which is domain specific, right?

PARTHA: 31:14 So let's say you have a data from automobile industry or from like biomedical domain, how you could quickly build a model for that particular domain. The reason for doing that is we want to make machine learning more knowledgeable, right? So, we want to build the models, not just based on the training data that we have for that particular specific task, but how we can sense the world better and utilize all of that external knowledge to make this particular learning task faster. And the hope is that if you're able to exploit those external knowledge, then maybe you can cut down on the amount of training that will be necessary to learn this particular task better because the hope is if you are more knowledgeable you can become better learners quickly. Right?

PARTHA: 31:58 So that way if you want to build like a learning agent for a particular domain, so having that domain's knowledge is important. So, one question is how do you encode all of that knowledge? Right? So, one possibility could be that like you know, for our communications already we are generating all of these data through documents to inform each other, you know, humans, whether we can read all of that data, right, automatically and build these kind of domain specific knowledge basis. So that's something we have been looking at. So maybe starting with the generic broad coverage knowledge base, how we can quickly deep dive into a particular domain and capture that particular domain's knowledge better.

CRAIG: 32:39 Yeah. I want to do a separate episode on NELL, but is it one system, one block of code, you know, a million lines long or something. I mean, what is NELL as an object?

PARTHA: 32:52 Right. It has evolved over time. I mean at the end of the day it still is a research system where like multiple PhD thesis have come out and different students and researchers have like contributed towards development.

PARTHA: 33:07 So there are like different, if you look at the architecture diagram there are like different blocks in it's working. And those different blocks are developed by different subgroups and different sets of researchers. They all interact together. Yes, yes. So, at the end of the day it's one single system with multiple components inside it.

CRAIG: 33:26 And have you ported that over to India or do you communicate with it remotely or how do you work with it?

PARTHA: 33:33 Yeah, I mean we have an instance of that and we have tried to use that for specific domains to instantiate that for specific domains. But we also have an effort on our own called INELL, which stands for India centric never-ending language learning where we are trying to build a knowledge graph that's, kind of, like, rich with Indian entities and relationships. So, you can think of India as a domain on its own.

PARTHA: 33:55 So maybe a little larger and diffuse domain. But the idea is to, kind of, like, capture knowledge about that geography and the entities and objects that are active in that particular space.

CRAIG: 34:06 That's fascinating. And the idea is this network of blocks will continue to grow and will continue to accumulate

PARTHA: 34:17 knowledge

CRAIG: 34:18 data and knowledge.

PARTHA: 34:19 That's right. Yes. Yeah. These agents are still not talking with each other. I mean there are, these blocks are within one particular agent, but as we were talking in the tutorial, how do you build like a cooperative community of these agents, which are, say, independently working in their own environments, building knowledge and then maybe like, you know, sharing notes with each other to improve each other's performance. So that's I think is, kind of, like, an open question, how to do that. But that's a very interesting one.

CRAIG: 34:46 Yeah. And for this work in India, do you have adequate computing power?

PARTHA: 34:52 Right. So, I guess with like in a cloud computing and also like computing has become democratized, right? And you have easy access as long as the resources to access those are there, right?

CRAIG: 35:03 Yeah. I mean it's expensive.

PARTHA: 35:04 They can be expensive. Yeah. So that way there are, kind of, like, grants. So, we build our own infrastructure at the university level, but then there also like grants from various companies like it's like cloud credits to be utilized.

CRAIG: 35:16 So going back to kind of the overview, is compute a constraint in India?

PARTHA: 35:22 I would say, kind of, like, in the well-funded groups, it's not a constraint as much, but definitely it's expensive. Right? So, you have to find a way to kind of like, you know, support that either through funding or cloud credits. But this kind of, the super-computing mission that I was talking about is trying to create a fabric, computing fabric inside the country where it will be much more accessible in maybe a government supported way so that you don't have to rely on other expensive options.

CRAIG: 35:52 Can you give us a vision of where you want INELL to go in five years? What it'll be able to do?

PARTHA: 36:00 Right. I guess like we would, kind of, like, want to go towards a situation where we are able to quickly capture the knowledge in a particular, at least in some specific restricted way about knowledge about the particular domain fairly quickly and then utilize that knowledge to improve learning agents. Right? So maybe to cut down on their amount of training that's necessary or improve their accuracy at similar comparable amounts of training data. I envision a situation where like humans and the machine, right? So, there's this, kind of, like, learning agents are able to converse with each other and are able to, kind of, like, exploit and build on each other's strengths. For example, right now the way machine learning is practice and like we provide supervision is very limited in terms of the training data that we provide, right? So, which is, kind of, like, label instances.

PARTHA: 36:56 So we are looking at how we can expand that bandwidth and provide like no more supervision that's expressed through natural language. So that like, you know, it's not restricted to only the experts to train these agents, but like, you know, a common person up to this, kind of, like, a conversational interface, will be able to train these kind of agents. So, we want the agents to be, learning agents to be more knowledgeable so that they can have intelligent conversations with humans and learn in that process. Right? So that's, kind of, like, the overall vision where we want to quickly bootstrap learners in a particular domain, utilizing various forms of weak signals that are, kind of, like, already available.

CRAIG: 37:37 And the hope would be that that would reach the long tail at some point.

PARTHA: 37:41 So yes. So, my research has been, kind of, like, how we can learn from weak forms of supervision, right? So that's, kind of, like, maybe already is naturally occurring that may not have been prepared. Keeping this task that I'm trying to learn in mind, but how we could utilize all of those things to learn this particular task better, right? So, all of those are, kind of, like, geared towards and relevant for how we can take it to the long tail because the idea is to bootstrap learners and these, kind of, like, an agents quickly in situations where you don't have too much supervision, right? So, which doesn't exist and may not be practical to create all of those large datasets in the short span of time in which we want to build these agents. So, leveraging whatever weak signals you have becomes important, which could be not only from texts but also maybe from other modalities.

PARTHA: 38:32 So yeah, so all of these things are like not very relevant towards how we can adapt these learning agents toward less supervised settings. In terms of summary for India. So, I guess kind of like, you know, there is like optimism, there are like lots of opportunities in front of us, which is not borrowed opportunities, right? So, these are actually important opportunities to solve problems inside the country. Right? So, we can like use AI as a tool, AI and technology more broadly to, kind of, like, leapfrog and provide those services and, kind of, like, think differently than what way it has been approached so far. We certainly need to, kind of, like, do more than what has been going on. Right? We haven't, again, it has been, kind of, like, more organic growth so far. There are like initial attempts in, kind of, like, having policies around them. I guess like things - in places like India have, kind of, like, developed that way. Right? So, where there is, kind of, like, momentum and then, kind of, like, government comes in rather than doing it the other way around. So, I think that involvement is, kind of, like, starting to come in, but I would like to see more like actual policies and funds being allocated to these kinds of problems and projects and like to make sure that those are executed.

CRAIG: 39:52 That raises another question that I didn't ask. In China and in the US national security or the military or the defense departments or however you want to label them, have been a tremendous driver of that kind of research because there are clear national security implications and applications. Does that exist in India as well?

PARTHA: 40:16 Right, so India, like the defense departments have their own research labs, so most of their research they have, kind of, like, done within their own labs. Right. I would say there hasn't been traditionally this, kind of, like, strong tie up between the military and government, like the DARPA model, kind of, like, doesn't exist as efficiently as it does in the US and only, kind of, like, I would say like in restricted amounts of research that they were not doing in the labs, now maybe because of both the opportunities and the amount of talent that's, kind of, like, available in house. People are, I think like now looking more broadly to tap into the talent base. Yeah.

CRAIG: 40:57 I would think though that the military in India would have difficulty attracting the kind of talent they need to do very significant things.

PARTHA: 41:06 So that's actually quite interesting because I mean like you may have heard of like [ISRO](https://www.isro.gov.in/) which is like the Indian space research organization. I mean they have been able to do like very, very impressive things in terms of like, you know, having an [orbiter to moon](https://en.wikipedia.org/wiki/Chandrayaan-1) right like NASA took to Mars in like first attempt with like all homegrown talent. Right. So yeah, I mean it's, I guess like attracting like top tier talent is like a challenge everywhere and like defense establishments are not an exception to that.

CRAIG: 41:37 Yeah. Yeah. Okay. Great. Well I really appreciate it and I'll be watching your work.

CRAIG: 41:43 Oh, thank you. Yeah.

CRAIG: 41:45 Eager to see how INELL develops and India in general and

PARTHA: 41:49 Sure, sure, sure. Yeah, yeah.

CRAIG: 41:51 Okay, great. That's it for this week's podcast. I want to thank Partha for his time. For those of you who want to go into greater depth about the things we talked about today, you can find a transcript of this show in the program notes. If you don't see it, visit [eye-on.ai](https://www.eye-on.ai/) and you'll find a link for it there. We've been getting good feedback from listeners and I hope to hear from more of you. I wanted to thank [Tom Koppel](https://www.linkedin.com/in/tom-koppel-38862b1/) this week for his valuable advice. Let us know whether you find the podcast interesting or useful and whether you have any suggestions about how we can improve.

CRAIG: 42:32 The singularity may not be near, but AI is about to change your world. So best, pay attention.