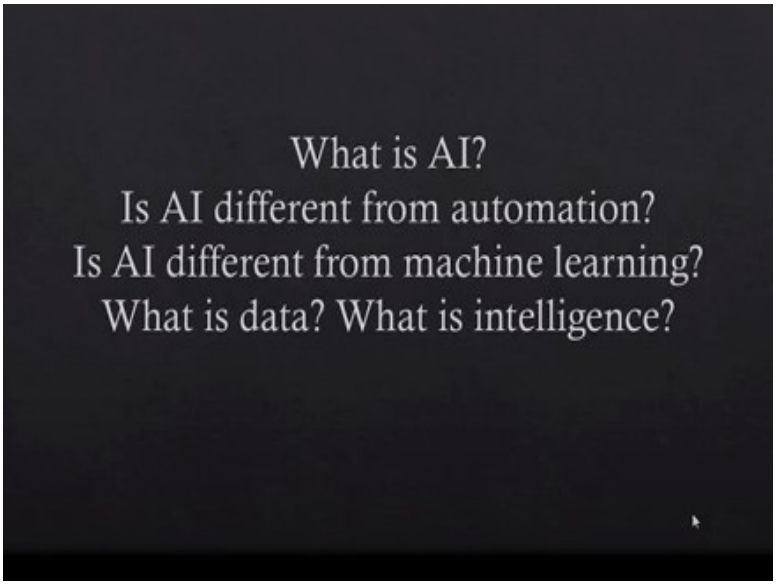
Towards an Ethical Digital Society: From Theory to Practice
Prof. Anupam Guha
Centre for Policy Studies, IITB
International Institute of Information Technology-Bangalore

Lecture - 07
AI and Ethics, Governance, Policy: Landscape and Roadmap

So the, I have two talks in a row. So it is going to be a bit unorthodox. The discussion sessions it would not be like a talk and then a discussion session, I would have stuff for you in the middle. And then in the end, there would be a large discussion. I am Anupam Guha, I am from IIT Bombay. I am a professor at the Center for Policy Studies there.

I work on artificial intelligence and policy. The topic is, AI and Ethics, Governance and Policy: Landscape and Roadmap. So, we will be covering ethics, governance and policy in that order.

(Refer Slide Time: 00:47)



So I mean, because this is, I was told that this is going to be a diverse audience. So may as well go over what is AI. And while you know there are a lot of answers out there, even within the community, even in the artificial intelligence community, what is AI is a contested question.

So you know in the philosophical realm, there is that distinction between artificial general and artificial narrow intelligence, right where artificial general intelligence is

supposed to be systems which can do things which humans do whereas artificial narrow intelligence is supposed to be systems, which solve specific problems albeit without a lot of human interference.

But in the engineering or even in the scientific research realm of AI, it further shrinks where artificial narrow intelligence gets into a bunch of techniques and the dominant most popular technique in research happens to be machine learning. And machine learning is nothing but a discipline wherein, given previous information, knowledge or intelligence, you make decisions for new knowledge or information or data points you are given.

And hence, for the purposes of this talk, when we talk about artificial intelligence, in general, we are talking about machine learning, that is not to say that there are not other important areas of AI out there in research, but machine learning is very clearly the dominant one. Hence, the third question is AI different from machine learning? Yes, in many ways, there are AIs, which are definitely not machine learning AIs.

But again, if you are talking from a policy perspective, you are going to see which of those things have the deepest impact in society right now from a political economic lens and that is machine learning. Is AI different from automation, because often the words are heard together AI, automation, especially in relation to things like for example, labor. So automation is a much wider term and has incorporated technologies for a while.

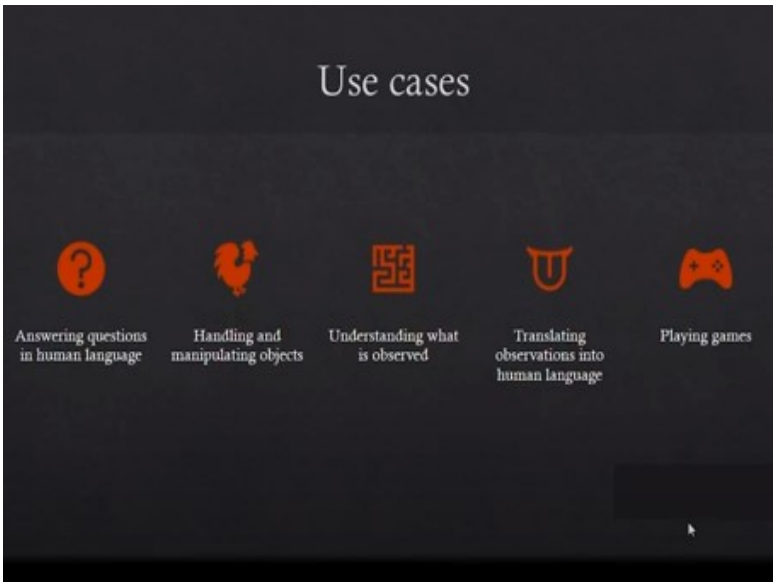
And often even with technologies, which are fairly elementary has had wide scale impacts, because automation is often heard in conjunction with like, for example, assembly line production. So you automate away certain processes or with for example, coding and development, where you automate away certain processes. Those processes may not have anything to do with AI.

But from an economics point of view, you like to club in AI and automation for certain purposes. And then what is data and what is intelligence? Like, we will not go into the precise definitions, because they are very involved, and way beyond the purposes of today's discussion. But intelligence is rarefied data. Intelligence is when

you have a lot of data and you sort of shrink it into something, which gives you the knowledge to perform certain actions.

That is intelligence. And intelligence hence has a value because you can perform certain actions using it.

(Refer Slide Time: 03:59)



So use cases of artificial intelligence, there are billions. But some examples, and I am talking like my background becomes a bit relevant here. Not only do I work in AI and policy, but my previous life, I was an AI scientist. So my PhD was in AI, not AI and policy. Because I worked for two years as an AI person.

So what I have noticed is that in policy discussions or in ethics or social science discussions, often the focus is on very specific usages of AI, which are often in the zeitgeist. But AI has a lot of uses. It is often interesting to know about them because they also have impacts. So for example, one of the biggest problems right now which is one of the focuses of language research is question answering.

What that means is that there is an entity. I ask it a question in a human language, in English, Hindi, Malayalam and it answers me back in a human language unstructured, without any forms. And question answering, as you could imagine, has a lot of usages. Everything from teleconferencing to human resources, there are parts of it, which might get automated away.

So it will have impact when it becomes a mature technology. And in many ways it is maturing very fast. My own doctoral work had a lot of question answering component. The second is, of course, handling and manipulating objects. When you think of artificial intelligence it is easy to forget that, you know once upon a time, it used to be synonymous with robotics.

And it in many ways, still one of the most important areas, though people have stopped to think about that in the policy sphere. Like robots have changed a lot when they came in the 80s to the Indian industry, they were these gigantic, clunky machines, which could do certain very specific things on the assembly line. And that was all they did. You had for example, car manufacturing robots.

But now you have robots, which you can teach to do various manipulation actions, manipulation actions in robotic linguists, the actions by our mandibles, our hands. And instead of doing one action again, and again, and again, you could theoretically teach a robot how to fold a shirt. And yes, it will fold that shirt much slower and in a much more clumsy fashion than you do.

But then it will not ask for minimum wage. And it will keep doing that boring action again and again and again, perfectly every time folding clothes. So that is one thing to think about from a political economic angle. Then understanding what is observed. You look at a thing and then you observe. Similarly, a large part of AI deals with looking at things.

You look at a fruit and tell it is an apple. You look at a face and recognize the person. Facial recognition, one of the largest topics in discussion, right? Then you look at an image. And instead of just understanding what that image is, you write an English paragraph about it. That is the captioning problem. It was solved in around 2012, 2013, one of the biggest shocks of my PhD life.

And one day, we got a call that you know what Li Fei Fei's lab has done it, restart all your research. And everybody in every American University had a sudden bout of anger towards Stanford and Google. But that, such is life. But that is one of the big

parts of AI, connecting. See, humans human intelligence primarily is rooted in two things, observing things, and talking about stuff.

There are species of talkers who like to talk. That is what we have been doing here, right. So if you can combine the two if you can combine looking and talking. That is like one of the holy grails of AI, can we talk about things we look at. And that is the use case. The last one of course very important, playing games. We have been winning at chess for a while.

But finally, we have won in Go, which is a much more complex game. Do you guys know what go is? It is a Chinese, Japanese, much more simpler to learn the rules than chess. And millions of times more harder than chess. It has only two kinds of gotees, the black and a white, no specific shapes. And you should you should look it up, go is very fascinating to play. Very easy to learn.

You would learn it in an hour, but you will spend your life playing it and you will still be an amateur and then a computer will come and beat you. So these are some of the use cases out of many. But each one of them, if you really think about it has a policy impact, has a political economic impact, even the playing game bit. So start to think how.

(Refer Slide Time: 09:05)

Now we are talking about AI policy and governance. What is India's outlook there? What are the documents reading, which you can understand what our administration

is thinking about AI? There are many. These are two more important ones. There is the artificial intelligence taskforce report, which came out around one and a half years back and then soon afterwards came out NITI Aayog's National Strategy for Artificial Intelligence.

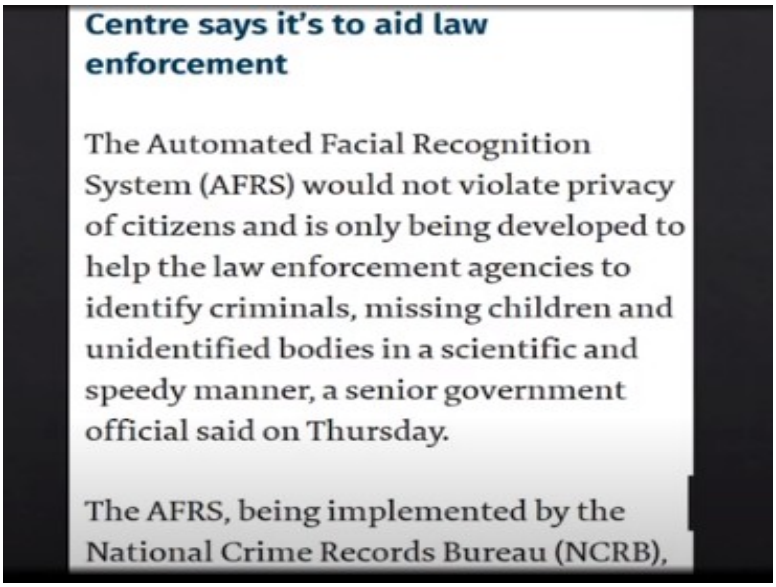
I would suggest that you guys go home download these papers, they're pretty big and clunky, read through them. Try to critique them try to see what are the strengths, weaknesses of these documents. They are easy to read. They are meant for like the meant for a general audience. And they have lots of flaws. You should try to pick what those flaws are. All right.

(Refer Slide Time: 09:56)

But one thing is to declare that here is what we as a country want to do. And what we as a country want to do is still extremely contested. Who the we are in we as a country is also extremely contested, always. But often what happens, especially in India, is that things get made first, policy comes later.

And this has been happening a lot like you know Punjab police was awarded in 2018 apparently some award for facial recognition, some system they are using there. They may have got this award, but later we got to know that the error rates were terrible in this particular system. And those reports are out there in the public. You can read those reports, something for your curiosity.

(Refer Slide Time: 10:46)



Now that now there are multiple states which have which most police are aggressively pursuing facial recognition tech. Hyderabad is one, northeast states are pursuing. It has been verbalized that there is going to be a national facial recognition thing. And well, should have written an Op-ed loudly yelling that why it is a bad idea. But I missed the deadline like every other academic does.

But I will still write it I guess. Now why is facial recognition tech a bad idea? Yes. Can I have a show of hands? I do not want to just drone on and on I need some thinking. Why is facial recognition a bad idea? Any answers? Bad answers are okay. But please it is a good one. Anybody else? Surveillance state, but surveillance state breaches privacy. Why is surveillance state bad? It does profile.

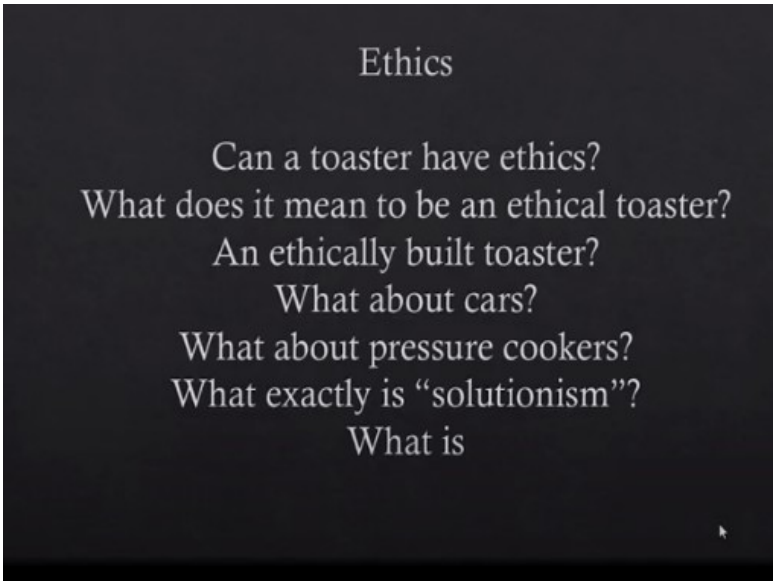
But there are some other reasons as well. Generally when we talk about facial recognition, the most popular discussion is around privacy. And I think it is an important one and it needs to be discussed. But there are a few more. For example, suppose that the state were to ensure, it is not ensuring anything right now. But suppose it were to ensure that all the facial recognition tech it does is for national interest.

And all the facial data it stores would be secured and surrounded by 11 feet walls, guarded with black cat commandos and your privacy would be extremely protected. And you know it, etc. I would still not like it. You know why? Well, one reason is that there is something in civilized countries known as presumption of innocence. What

that means is that I do not want the state to look at me if I have not done anything, right.

The facial recognition technology sort of inverts that logic. It needs to constantly look at you to make decisions, by definition. So you do not get forgotten. Somewhere you are there even if though your privacy ostensibly is not divulged or etc. And that is something I want to you, there are other points as well and we will return to facial recognition, it is a pretty big deal in AI policy. But I want you to chew on that for right now, okay.

(Refer Slide Time: 13:16)



We are I was brought here to talk about ethics. AI and its technology. I have been talking about ethics all this while. So let us get down to ethics. Tell me something. Can a toaster have ethics? Toaster, which toasts bread. Can a toaster have ethics? Anybody? Yes. No. Some interesting answers. Think of a toaster. You get in the morning. You are groggy, no coffee in the house. You drag yourself to eat a toast.

There is a toaster, it pops up a bread. Can that toaster be an ethical toaster? It cannot, right? Ethical toaster makes no sense. It is in an ethical state play. Can have an ethically made toaster perhaps, perhaps how that toaster was made? Perhaps it used, you know I do not know local labor, organic metal. I do not know what is ethical these days. Some of those things perhaps it could have used.

How about an ethically made ICBM? Local material, naturally sourced uranium. Ethical ICBM, anybody? Can you have an ethical ICBM? See, I am not trying to make fun of you. I am just trying to ask what is ethics? And the problem is that nobody agrees. Do you know how many ethics councils exist for AI ethics in the world? There are 70 major ones. They do not agree on anything.

Because I could tell you that, you know you think *bainadon* is unethical. And my framework of ethics is as valid as yours, because neither you have an army to back you up neither I have an army to backup. So okay, what about cars? Now here is, you know the car analogy has been thrown at me by so many Microsoft people that I may as well use it myself.

So I was at this conference, and there was a gentleman from Microsoft, I do not know why they like cars so much, perhaps because they are Americans who like their cars. But he was like, Anupam look, a car is a sophisticated machine. AI is a sophisticated machine. Can you prevent accidents just by regulating them? No, you need drivers to be ethical. I was like, sure, but that is not how it works right? You first have the laws then you.

They were like no, no Anupam, the laws are there just to have guardrails. What you need is ethics and fair cars and lack of pollution. And then I asked, like it might just struck him that the fact that in the world there are so many cars running around, and not as many trains running around, do you think it is a natural state of order or something that, you know God waved his hand and the world was full of cars and not trains.

No it was not. Do you know how it happened? There was murder involved. Like in America, there were lobbyist groups at that time. And anytime a suburban railway plan or a metro plan, there were some kooky plans like vacuum metro and all. Whenever those were given in New York, the lobby groups they would like blackmail, threaten, extort, sometimes murder them.

Car manufacturers in that day, were not like the **FL.** They had business sense. They got what they wanted. So they got a whole country where only cars were running

around, and you go to America and do grocery run without a car. And then what America does was everywhere, and it is a cool thing to do. Everybody was using cars.

What I am trying to say is that ethics sometimes is used to here is a particular other kind of power play, which comes from the structure of this. You can use ethics on the structure. But what if the structure is itself is the problem? What if you want trains and not cars because you know poor people want trains and you might want poor people to use transport more than people who privately owned cars.

But if you live in a world where everybody is driving cars tough luck getting your train, okay. What about pressure cooker? We have talked about toasters, we have talked about ICBMs. And we have talked about cars. Now we will go to pressure cooker. So I am going to repeat the story again. Sorry. And if I have bored you with this story, I am going to still bore you with it. Was it you who told me the story?

No, somebody, it is good. So here is the pressure cooker anthology of solutions. This whole room full of very intelligent people, some time warp thing happens. You know lightning crackles. And we are all magically transferred to 18th century Victorian England or some such place. Some terrible place. Streets are stinking. There is garbage all around. People are smelly. They are like where are we? Also there is a problem in this world.

A very obvious problem, which is that women do not have rights. Women are property. Half of you are property. Sorry. Blame the light. Now the men are very angry. They are like we need to do something. We need to be saviors. We need to save these poor people. Also these uncivilized wretches who live here. How do we save them? How do we save them? They do not know how to save them.

They want to do **(FL)** stuff, but they have forgotten they are Indians. They do not know how to do **(FL)** stuff. So then another wormhole opens in the atmosphere. An Elon Musk resplendent in glory comes down. And Elon Musk is like you poor fools. I will save you all. They are like sir, how will you defeat patriarchy sir? Elon Musk is like there is his magic wand and a pressure cooker.

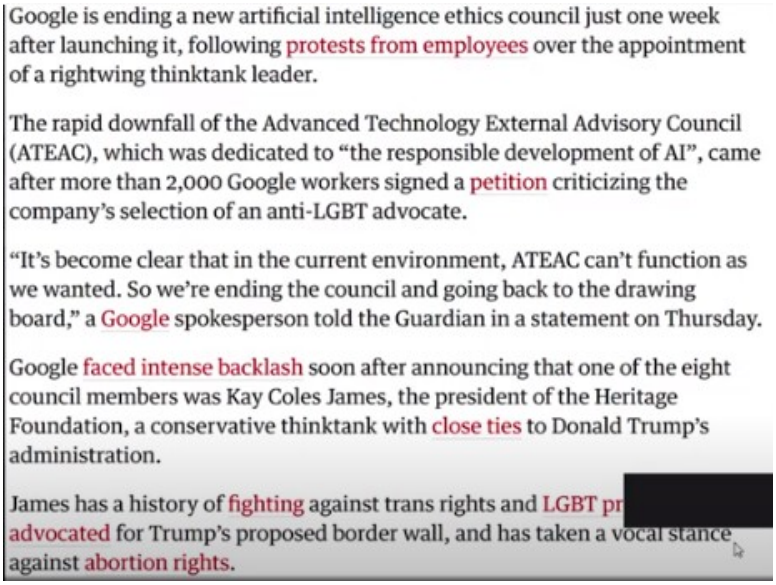
And he is like, well see the problem is that women do too much work in this civilization. This is called a pressure cooker. If you cook food with it, food will get ready in 40 minutes. See work has been lessened to a fraction. I have solved patriarchy. Everybody bow down before me, bye my job here is done. What is wrong with the story? The pressure cooker cannot solve patriarchy. The problem was not the lack of a pressure.

The problem was that the whole system was designed in a particular way. The problem with ethics is sometimes the same. And you are trying to play games over a system which is fundamentally wrong, who cannot ethic away the problem, because the problem is the whole structure itself. And the only way to change the structure is by coercion. And we will get onto that. So that is what is called solutionism.

And various kinds of solutionism is there. Thinking that technology will solve all your problem is tech solutionism. Thinking that a regulation will solve all your problem, or a law would solve all your problem is legal solutionism. People think like that I will make a law problem will go away. But you often forget that the whole system of law comes from some kind of coloniality power structure which people often forget.

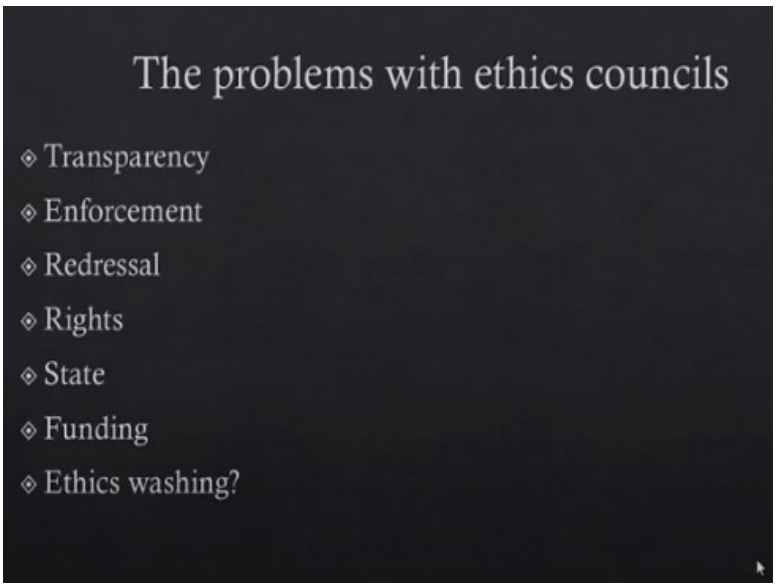
Policy solutionism also exists. Policy people like to do that all the time. We will make a policy, problem solved. But the problem was not from a lack of policy, it often came from socioeconomic structures. I hope that answers what is what is the problem of ethics we are trying to deal.

(Refer Slide Time: 21:10)



Now I was telling you about ethics councils, right that there are these big powerful companies who have these ethics councils on AI. Google was forced to close it recently. It turned out one of the members was like a Nazi, or something. I am sorry if I am like, defaming him. He said that I do not know. Didn't like LGBT few people perhaps, or something? I do not know. But they have shut down their whole ethics council. What is the problem?

(Refer Slide Time: 21:41)



Okay, let us get down to business. What is the problem with ethics council? First of all, who made them my boss? What is, why do they have the power to define how AI should exist? What goes on inside that? There is absolutely no transparency in how they are formed and why they work. What do they do? Second thing. If they do things I do not like, how do I force them not to. coercion? You cannot have civilization

without coercion my friends. And ethics councils are the biggest example of that. How do you force them to do anything?

They are not beholden to you. You can keep shouting outside their office, it does not matter. Third one, redressal. Suppose an AI system is made which, I do not know, performs some dystopian shit on you. Who would you go and complain to? Google's HR department? Fourth one, rights. And this is the most interesting. See the problem with ethics is that it very cleverly bypasses the question of rights.

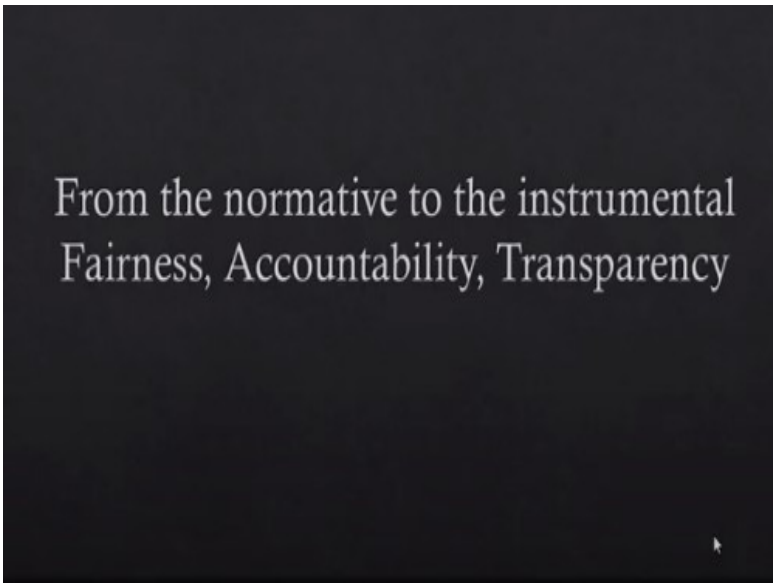
Because ethics does not need to answer questions about human rights. A system can exist and by its very existence, erase rights. Ethics is about how that system is made. It is not about the normative impact of that system. Fourth, no what fifth, state. Well, ethics councils are not connected with the state.

State cannot do anything here or has the excuse not to do anything rather because often the state is in bed but would not like to seem that it is in bed and it is a complicated relationship. Then funding. Who funds ethics councils? Private companies. Why would they do things which cut their bottom line again? Last one, ethics washing. Have you heard that of pinkwashing? Any of you? Okay, good.

If you have not heard that of pinkwashing, pinkwashing refers to that, you know you have these countries which are like colonial or imperialist are very bad countries, which do a lot of brutal stuff. But then they say, our army has minority representation. Yes, Israel is excellent at this, but this is all getting recorded. So let us chill. So as you know tomorrow I can like, you know run deeply brutal in our state and then upset.

But my deeply brutal police enforcers, half of them are women. I am woke. This is called pinkwashing, using social justice to bullshit away our structural problems. Ethics washing is the same. It is using ethics to divert attention from the fact that it does not answer the question of wealth. It does not answer the question of regulation. It does not answer the question of accountability. It does not answer the question of enforcement. And ethics washing is has become like the de facto way of dealing with AI problems.

(Refer Slide Time: 24:35)



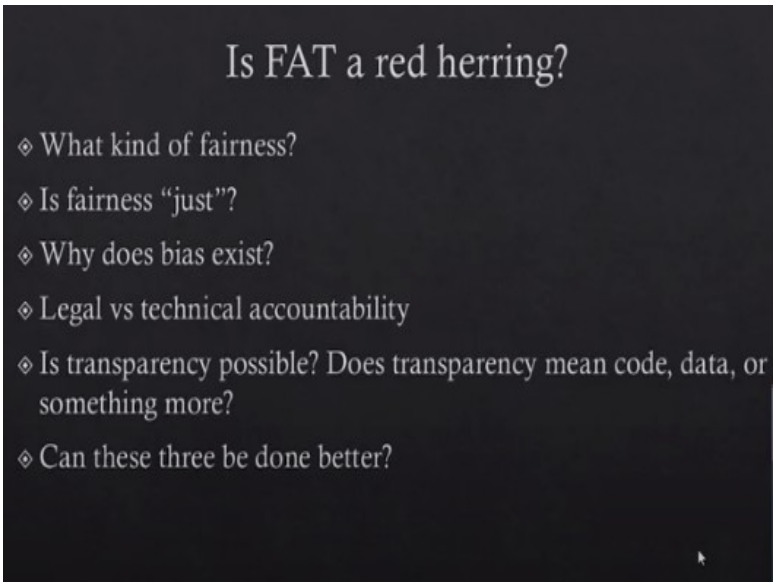
So we go from the normative to the instrumental, yes. So most big companies, yes they have them and some of them are trans company like EU has done lot of these and sometimes civil society activists, make them and try to force companies to act according to them. They failed miserably. That is the thing. So normative to instrumental. What I mean is that from what to do to how to do it.

So people started to, so actually it was a one after another, there were two parallel ways of AI governance, which were being thought of. One was with ethics. The other was with something called FAT; fairness, accountability, transparency. This was a instrumental way of governing AI, not look at the end to end thing, but rather how AI works inside. The first one was fairness, the idea that a piece of machine learning would not bias itself towards some or against some demographic.

The other was accountability that if an AI system produces a result, it should you know you should know why it fails. If it fails, then who is accountable, who is responsible, if it fails, etc. The third one was transparency, if an AI system gives a reason, why did it give that? Could you go inside that system and tell why it gave that reason. Now the problem, I mean, fairness, accountability, transparency, I would say are much more objective than something called ethics.

But there is a big problem with them. And that problem is that as an AI guy, as a person who worked with AI, I can tell you, each of them fails in a special way, and not in a sort of a fails in a structural way, but rather fails in a technological way.

(Refer Slide Time: 26:23)



So first of all, let us talk about fairness. You cannot remove bias. You can try to, you could try to make better data sets. But then most of the important data sets come from very big collections of data obtained from society. And unfortunately, society is bigoted. You cannot clean that data. You could try to, but it would not ever be complete, like everybody in the AI realm knows this, that you cannot achieve good fairness.

And in trying to achieve fairness, you would compromise on accuracy. The machine will stop working if you do not feed it real world data, right. But real world data comes from us and we are racist, sexist classist people casteists as well in India. So either you sacrifice accuracy or you okay, then there is a third problem with this.

Fairness often is in the cost of justice, like you may want some systems to be a bit unfair towards like you know if a system, for example, is having like insurance schemes, you may want it to be a bit biased towards the economically weaker sections of society. Engineering that is not trivial as well. So fairness is not that easy to achieve first of all. It has problems. And sometimes there may be competing kinds of fairness.

Do you want this demographic to be preferred or do you want that demographic to prefer. Who measures that? Who gives somebody that authority to decide that we go away from real data and then tweak this system to work in a particular way. This is a

deeply contested field bias, fairness and bias. The second is accountability. Accountability is again interesting.

And you have talked about that as well in accountability, that if a system fails, who is accountable? And sometimes, you know systems fail. AI systems or other technological systems, does not matter how robust they are, they will fail. And you will not be able to get anybody who is accountable, because they are made by multiple scientists and engineers. And it would be almost hard to you know pinpoint.

And it would be unfair to pinpoint some human and say that you are accountable. So there is that problem of legal versus technical accountability. That you know sometimes these might differ. Suppose a self-driving car crashes. Now that statement itself has a lot of presuppositions. Why is there a self-driving car again? Who made self-driving cars possible or allowed or you know on roads.

That was a political decision that was not a technological decision, the existence of self-driving cars. But if you have one, it will someday kill somebody and has already. Self-driving cars have killed people. At that point, who exactly is responsible? The CEO of the company who made this self-driving car or what and then legal accountability is different. It is like insurance. Who will pay?

By when would insurance get paid, when you will not get paid? Now the thing I am trying to point out here is that the existence of self-driving car is itself something you should have contested. So the structures on which these automated systems are often built is where the area of policy contestation lies. That is what I would say. Once these systems get ossified, it becomes harder.

The last one, transparency. Is transparency possible? The answer is no, rather it is not. You could say that, you know if a system gave a result, I should know why that result was given. But as a person who has built a lot of deep learning models, let me tell you sometimes the creators of deep learning stuff, do not know why a thing works. They just know it does. If it works, it works. If it does not work, it does not work.

And you know I had the same conversation in Berlin, and one of the people asked, why cannot we, can we not ask the company to show their source code for example? Yes, you could. And yes, they would show the source code. But showing the source code does not equate transparency. What I mean is, is that there is no malice involved here. The source code creates a structure, that structure gets trained with data.

And after lot of training, it gives a certain result. The mathematics that are happening inside that structure is not something that the source code reflects. That mathematics gets created when the source code when the training happens. And that mathematics is beyond the person who coded it. It is beyond any math, mathematician on Earth. It is beyond you definitely.

And you are not even a engineer, you are a bureaucrat probably at that point. So transparency is not happening, not in the way you imagine it. But another way of transparency, all of these things could be done, but not in the way people think. So for example, accountability. Stop throwing away accountability to machines. Like if a problem happens, make certain that companies have humans who are accountable, who were there to oversight the machines.

I will give a very simple example. A trolling moderation on YouTube comments, for example. Now you know how it is done, right? First, there is a machine learning thing, which flags what it thinks are objectionable comments. Then a human looks at it, and if they find it objectionable, they remove it. But it does, like go to YouTube, you will see so many alt-right videos, communal videos, hate videos, etc.

And you will be like, why it is not working. And some reports have come out that the humans who are in charge of looking over the flagged videos, they are underpaid, under trained, deeply working in precarious conditions, and they have watched so much shocking, hateful content, that it is not good for their mental health, because they keep watching that kind of stuff again and again.

Now whose fault is that? That fault is of the company. Fetishizing artificial intelligence will not solve your problem, you have to make people who make those systems, especially the company CEOs accountable. So you would say that, look I

understand that technical accountability is not possible. Hence, you have a human in the loop. Otherwise, you do not make this system.

If you are making this system and you are taking commercial responsibility for it, you need to have enough human being in the loop to make certain the problem does not happen. By the way, this very argument, cut short the automated car thing, if you like we will take it to its logical conclusion. Similarly, transparency. You cannot have technical transparency. And I have just explained why.

But you could have other kinds of transparency. You could have which part of the company gave funding for what kind of research. You could have why a certain artifact got researched and built in the first place. You could have data on who researched it. There are various kinds of transparencies, which are as important as the transparency you think about that why did this decision happen?

It was not just the AI artifact, which made that decision, a series of economic decisions led to a creation of an AI artifact, which made that decision. And that transparency is something you should as a policy person perceive, okay.

(Refer Slide Time: 33:42)

Function creep, idiot god, artificial scapegoat, and I am not abusing anybody. Function creep. Happens a lot. What this means is that you make an AI system, and people start using it in context, it is absolutely not meant for and it starts to fail

miserably. So for example, we keep talking about facial recognition tech, right? Not meant to work for context, they are not trained.

Similarly, natural language processing, certain systems not made to work in context, they are not trained. This happens a lot with AI because most of machine learning is trained on data. And no data is universal. All data come from a context. Outside that context, the system would fail.

But unfortunately, humans who use these systems do not realize that they use it in contexts they are not meant to be used for and end up committing massive crimes on human beings, by way of negligence, by way of bureaucratic debts, by way of stuff which happens which we have seen. So one of the things you should always demand as a policy person is for these companies to show their datasets.

Because you need to understand the context of that dataset. Not to remove bias from it because like we have discussed that cannot be completely done. But the context is something very important. Second thing, idiot god. So idiot god refers to this idea that you have these systems, which on one hand are very powerful, which can deeply impact your life, which can completely spoil your life.

But on the other hand, they are completely stupid, like facial recognition, right? It will always be non-deterministic, which means it will always do an error. But if it, facial recognition is being used for the carceral state, if you are using it for policing purposes, are you willing to play dice with somebody's liberty? You should not. It is against the constitution, right.

The constitution is either valid for everybody or it is not, right? But we have police already doing it. Why? Why, by the way? Because we do not have a law. Hey, this is interesting thing. Whenever we start to advocate regulation, people immediately say, oh state, state is scary, stay away, do not make laws. But by not making laws, often people enable states to do far more worse, because laws clearly lay down the no go zones.

Regulations are the only important tool in your arsenal for governing AI, not ethics councils. That is my line and I stick with it. Third one artificial scapegoat. Yes, please. Well, you look for where they were collected from. Those datasets were probably collected from certain demographics, right. So suppose it is medical data, and you have a system trained on that data set, it will not work for some demographic from which it was not collected.

So suppose you know you collect and this kind of changes happen even like locally. Suppose you collect medical data from a certain age group, it will not work for some other age group. So these kind of things are very important, because they often work in these crucial system like healthcare AI is pretty delicate, right? You do not want to misuse it. So that sort of stuff. Artificial scapegoat.

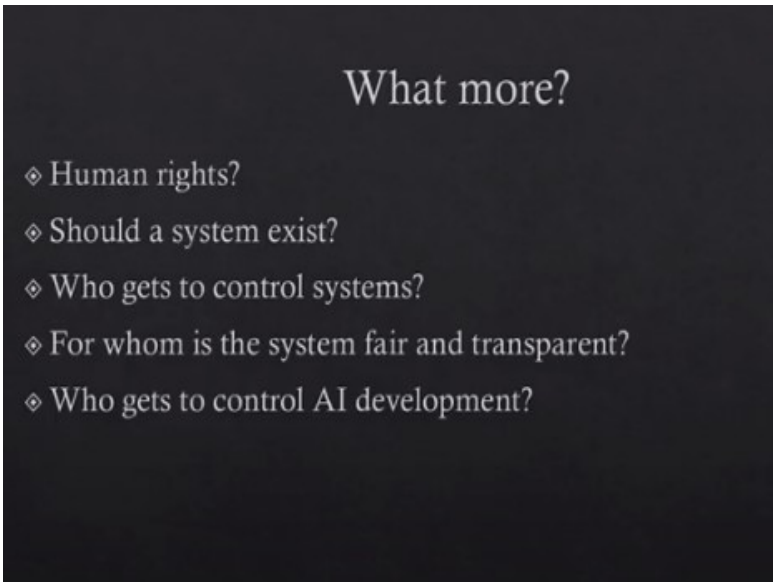
This is I actually wrote an Op-ed in it and I would advise you guys to read it for fun. The argument I made in that is that recently what has been seen across jurisdictions, across countries is that officials and corporate people, they have started to scapegoat AI. Essentially here is how it happened. Some police wala somewhere has a facial recognition system. Naturally, it will give errors.

Then they would arrest somebody who do human rights violations on them. Then when the press comes, it is not my fault, the AI told me to do it. That thing said that to arrest this person. We have seen this so much in India, where the technology is blamed, but the misuse of the technology by the officer is not held into account. This is artificial scapegoat. It is not artificial intelligence.

Like you have to keep this in mind that decision to use a particular tool is always a human decision. Suppose some machine is not working, and you decide not to give food to a particular person who dies of starvation, you made that decision. The machine might have been making an error and it was your job to have done due diligence to find out what was happening.

People often do not, but that does not make an inanimate thing have responsibility for moral actions. Political and moral actions are done by humans not by machines, this is something we have to internalize.

(Refer Slide Time: 38:36)



It may have been biased in the first place, the point is you have to you have to accept that there will be bias. And you may bias it later as well. And hence you have to be careful, like most machine learning systems are biased. Most machine learning systems will do, all machine learning systems are non-deterministic. And hence most of them would do some error at critical points where you may need it for life saving or whatever purpose.

Am I saying do not use machine learning at all in such cases? No, I am not saying that. I am saying that use it with enough political sense. Use it with enough human responsibility. You cannot assume that not just machine learning. You cannot assume any machine on earth would always work. But we mean to do with other machines, right by the use of this thing called technical and common sense.

Why cannot it be with AI as well? And we can encode those things into regulation. We could do a lot but let us move on because time is passing. Human rights. Like I mentioned before, ethics, do not account for rights. And rights are by the way, something very useful. Human rights are not some vague concept. They are very well documented. We have papers and reams of paper discussing human rights.

Unlike ethics, people generally agree on what rights are over jurisdictions. So human rights could be one of the extra frameworks you use to bolster ethics with, and could be also one of the extra frameworks you use to sort of make no go zones for your

FAT, your fairness, accountability, transparency, what systems it should not be made. This argument has been best articulated by one of my colleagues, Vidushi Marda.

She wrote a paper for our organization, Article 19, I would advise you to read that paper where she advocates human rights as a additional measure over ethics and FAT. Then, but then I am not satisfied with human rights. Because I could, theoretically, you could have a system and you could prove that it is ethical. And you could also prove that it is not violating human rights, and it still should not exist.

I will give you an example. Suppose there is a system, which makes you, you know which completely gives you the political leanings for some party, and allows you to accumulate extreme political power without violating privacy, without breaching anybody's human rights without killing somebody. But you would still get a lot of political power using that system.

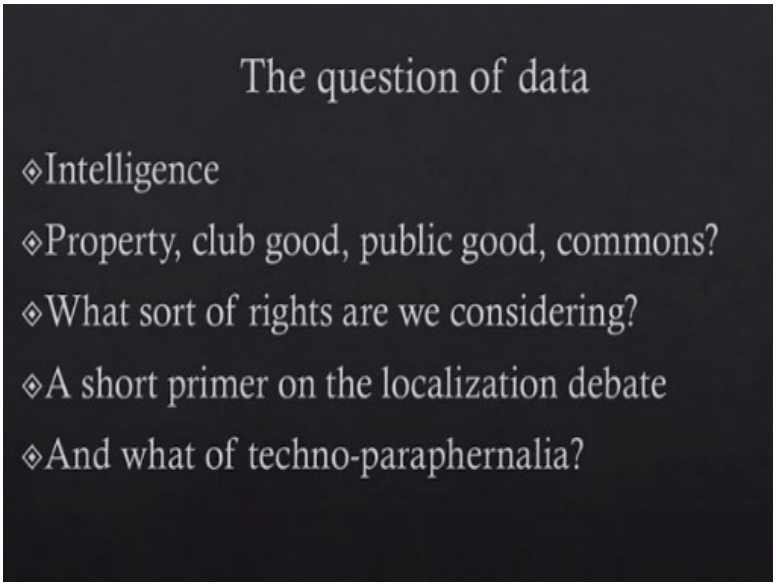
That system is subverting your democracy, it should not exist. It is not violating human rights, but I do not care. There are other things I care about as well. So the point I am trying to make is that certain systems should not exist, not because they individually harm somebody, but because they might collectively harm society.

And because AI systems are built from intelligence obtained from society, they are built from collective intelligence that build from data, and societies should have a stake in deciding what AI systems get built within that society. And who gets to control systems? Or other for whom are the systems fair and transparent? Is it for the company CEO? Is it for the government bureaucrats?

I would say that, if a community is generating the intelligence, which makes the existence of a particular system possible, then the community also needs to have the oversight, democratic oversight over that system. It is a fairly unheard of like, this is me going too far like this is the line I have had, but like this is me saying that AI should be regulated essentially by communities in which those AI systems operate.

Because certain AI systems could have effects which go till that level, and we are talking very simple thing. And the last one who gets to control AI development. So here is the thing.

(Refer Slide Time: 42:51)



The thing we have avoided talking about till now is that AI systems, they are essentially intelligence systems, because they are intelligent systems, they essentially can make you do things, certain things very efficiently, logistics wise, trade wise, etc. Using those things, you could make a lot of wealth. Now how did that wealth happen?

You made a system which was created in research, which was publicly funded comes out of universities, but somehow magically gets privately appropriated under IP laws. And IP laws are a whole tangled web and I will not go into them, but they are, it is not the way AI should be developed. And that data those systems learn also come from human communities, often for free.

You know you are clicking away on Facebook, you are giving your data, the Uber driver is driving the car, they are giving their data, everybody is just giving their data, but that data has valuable intelligence, which is creating wealth for those who monopolize the structures, which are extracting that intelligence from that data. And that, in my opinion, is deeply iniquitous and deeply unethical.

So I do not separate this individual ethics from the ethics of the society and AI gets built in. And it is not some ephemeral moralism I am saying. That kind of inequity has

political and rights based impacts which far exceed that which comes from essentially a very vulgar look at rights. Your very democracy, perhaps and other politic like we will touch some points on media and power and things like that.

But if not have, essentially AI as property because right now that is what we have and we have somehow community data coalescing into private property and I am saying that that is unjust. So what could it be if not property? There are actually you know in Berlin in the Internet Governance Forum there were certain ideas which were floated. So one was AI as club good. Do you guys know what a club good is?

Suppose you suppose you are rich and you belong to a club, go and play golf. I do not know what rich people do but imagine, right. Now the club is private. It is not public. Yes, it is a the club good is like commons but not common. It is like a trust. Yeah. So you could hold data as a club good. You could, now some people are like club good, it does not go far enough.

Ultimately, National sovereignty etc. Evgeny Morozov makes that line. And he is like it should be a public good. Data which is not private and not contestable and not rivalrous, that should belong to your whole country. That is Morozov's line. And then people may go further, like Ostrom, Ostrom, had this theory of Commons. And she is like that, no, the state cannot be trusted with powerful things.

But communities have shown that they can self-regulate, and hence the data of a community should be controlled directly by that community, the means of production should be owned by those who work there. If you are creating the data, you should have direct control over it. Hence, community based data ownership or a real common like forests and water.

I think water will not be a common at the rate we are going, but let us hope that they remain. And here my question is, I am not settled on what sort of commons it should be. I incline towards it should be some kind of commons. But my question is more fundamental, that do we hierarchize rights? Like everybody talks about privacy, but nobody talks about inequity.

And when it comes to AI, inequity is the most glaring problem. There are job displacements happening while we speak. measurably, right, year after year. And no new jobs are going to happen. All that optimistic techno optimistic nonsense is like I have the numbers, I am not going to argue about those things. The point is that we are forming monopolies, which are going to be oligopolies, rather sorry for that, which are going to be much more brutal than anything we have seen because they would be intelligence oligopolies.

They would know all that is to be known. And that is something that we cannot allow. So I would say that we start to seriously think about the rights to live, which is the fundamental of, most fundamental of rights, right? The right to exist. And you cannot exist in a world which is permanently careering towards oligopoly. Now there is an answer to it, which is a very, which is a debate I would rather not go into because I find all sides in that debate problematic. But it is the data localization debate, some of you may have heard that. So the here is how the debate goes and no, it is not my position. Yes, but it is like it has conclusions which are different. So there is the localization debate. It says that, okay, Professor Guha, you are right. Facebook, Google bad.

So what do we do like, okay, here is what you do. Take all the data, and force all these companies to share their data, then you make a national commons. And then you use that commons to do development, okay? And my problem is the same problem I have with all social democratic processes, that it does not really change the ownership game. The ownership is still not with the people.

Not a big one, but a small one, big fish, small fish, etc. I am like till a data commons are completely on board I am like, let us make a data commons. But then I am like, let us control it democratically, at which people contemplate, democracy, ew, inefficient. I am like I do not care about efficiency. I would rather have direct control, direct public control and oversight over how that data is used.

And you should read on the localization debate, let us not like seriously not go into that. But there is another problem which as an AI person, I often think when I hear these data debates, because I have sat in a lot of them, everybody is talking about data

commons this that. They are making these bad comparisons to data is oil, data is wealth, etc.

And despite it being a bad comparison, there is an interesting angle with data is oil, which I would like, as a provocation. Suppose data is oil, it is not but suppose. There is a huge pond of oil I have made, I call it my oil commons. I have dug a hole. I have taken everybody's oil by force at the point and made all the private Google walas give their oil and poured all the oil in that hole.

And I am like, everybody, it is your oil. It is the public's oil. Anybody who wants to use the oil can use the oil, okay. Everybody runs towards that oil, all the common people, Google, Facebook, everybody runs towards that oil. But the thing is, only a few of them have oil refineries. I mean you cannot make an oil refinery. But Google already has an oil refinery.

Does not matter that you localized all their oil and put them in a giant big hole, they still have the refineries. And in data language, techno-paraphernalia, data centers, algorithmic coders, companies. So my question often is that when we sort of diverge that data debate from the AI debate, we forget that the value of the data is not in the data itself. It is from the power to extract intelligence from that data. And that part, yes.

(Refer Slide Time: 51:17)

Now coming, like did you get that point I was trying to make? And hence if you are talking about regularizations and you are talking about laws, and you are talking about commons, try to seriously think of public owned automated systems. And I am going to throw it out there. BBC exists. It is a public owned media system. And it is very well run. Of course, BBC has tons of fault.

And what they are doing right now with they are propaganda is terrible, but there I just made a partisan comment, but still, like compared to our guys like who honk on TV all day, the point is that you really cannot think of a data commons without a intelligence commons, without a algorithmic commons, without our refinery to get any benefit out of your oil.

And whenever you will hear that debate on how that refinery, how that public refinery gets controlled, there are multiple models, but everybody would want to say, you know let us give it to the "local players" they will do it efficiently. To which I would be like, but why not why not do it democratically. To which people would say, democracy is always inefficient.

And we can again go back to the efficiency argument. And you know what, machine learning is by the way very efficient. But like Vidushi, or Jyoti, or one of my many of my colleagues would say, they articulate this in a very crisp manner. That it is very efficient if you want to repeat the past efficiently. We have all your machine learning. We have all the structural injustices in your civilization.

Machine learning, it just repeats very efficient. But if you want to change the past, if you want to make a better world and that is where your policy brains need to start working. You do not want to be efficient. You want to be slow, you want to be just, you want to look at all possible angles, you need to be historically materialist, you need to be dialectic. I am sorry to use all these social science words, but that is how you need to do it.

(Refer Slide Time: 53:32)

By the way, we have been talking of oil extraction, this, that, but we have not really talked about the ethics of extraction, the whole thing of AI and ethics. This is another big part of ethics debate, which so ethics does not talk about labor, and I have tried to cover that with inequity and all. But there is another thing which ethics does not touch, which is extraction. I will tell you what that means.

By the way, I have another talk which covers this in much greater detail, which will happen after the break. So we will get to that in detail. But here is the thing. All your laptops, smartphones, etc. in this room, all of them are using lithium batteries. Are you aware of how lithium is extracted, like the working conditions of those mines? You should read that later, okay?

Similarly, germanium silicon, yes. And there are these special economic zones in the world where these things happen. It is extreme level of brutality. Making the techno-paraphernalia has a real extractive cost on the planet, which for some reason, nobody talks about when people are talking of ethics. I think lives of human beings are the most ethical thing to consider.

But as we have proven with workers, nobody cares about that. Because that work is alienated away from you. You do not see that damage. I am not saying buy ethical phones, there is no such thing as ethical consumption. I am saying that as a civilization, we need to perhaps, you know really clamp down on this efficiency, this production thing. We still need to be efficient and we still need to produce.

And we need to create more powerful AI so that everybody could be fed well and have wealth and all that. But the real costs of that in terms, not just of metals, but energy, climate change, etc., we will talk in the next talk in detail about these things. That is something. So extractions, it is not just your labor being extracted, it is also actual extraction happening from the ground.

(Refer Slide Time: 55:34)

I am trying to subvert the whole topic. Like the topic was AI and ethics. And I essentially have a contra view that instead of it being AI ethics, it should be AI and governance. Essentially, I am trying to say that most of the things we use ethics as the framework is actually, national and transnational governance, because ethics by itself often is either painfully limited or it is actually harmful.

Because it avoids the lens on certain problems, which have a far greater HR impact than things which ethics is concerned about, which is building ethical systems, because those techno-social systems are not isolation. They are not produced in isolation, but are produced in context of everything else.

So when right now if you ask that the companies have ethics boards, and when you ask them, what do you mean by ethical systems, they often mean systems built in an ethical way, rather than building systems whose impact is ethical. I am critiquing the framework of ethics because I mean, I could critique that. It is board of say the 70 different ethics paradigms that exist.

But my problem is that as a policy person, it does not mean anything to me. Like if I were to really say that, how should, you know in the context of India, how should we develop AI, I keep coming back down to governance tools, I keep coming back to the fact that we need hard lines on what can be done and what cannot be done. Ethics boards have unfortunately, at least to me, proven to be remarkably ineffective.

I am sure, I might just critique them and leave the framework of ethics. But I think that would be an intellectually lazy thing to do. I am trying to demonstrate that and by the way, my next talk on a materialist conception of ethics tries to actually provide an alternative framework of ethics, which is away from this sort of consequentialist sort of ethics we are using with AI right now.

And is a sort of different kind of, it even tries to redefine certain concepts of what it means to be ethical. I think that is the primary problem with ethics, right, that we need to first agree on that what framework of ethics are we using, and by the way, there is no such agreement within the field. Even within European and American ethics councils, there is not much agreement on what goes in.

There are some literature, which are investigating into the contradictions. And my honest line is that instead of like trying to make sense of all of them, why do not I just make a line of what I want AI systems to do and make laws around them. Because from a policy point of view, that just looks more like policy needs to be predictable first of all. It needs to be doable and predictable, and it needs to work.

And I think we really need regulation, frankly, at the end of the day. But I was getting to the last slide before we show you that little movie on code standards and policies. A bit of a real politic real world thing here. Even with ethics, or even with whatever little policy tools we have right now, the tragic fact is that with AI, often what happens is that an AI system gets built first by some company.

Then that company pedals that system to some larger company, to some international conglomerate, to some government agency. That system gets used. It becomes a

standard. And once that standard has been built policies made to back fit the existence of that standard and make it legitimate. That is often how things happen.

So often you have standards being made by essentially, bodies who are not liable to anybody who are not government bodies, not private bodies, but somehow they get to define standards and standards get made. And then apparently it is all consensual. Everybody agrees that that should be the standard. And I do not know how this consensus gets formed. But then those bodies of people then make the laws.

And those laws have no oversight and they get passed. Often I would use the word bulldozed and that is how the system is. So one of my, like inclinations is that if we are having this conversation on the policy around techno-social systems, we need to work in a way that policy is abreast, if not before standard making by essentially bodies whom you have no oversight over.

(Refer Slide Time: 1:00:24)

All right. That is my mailing address.

(Refer Slide Time: 1:00:25)

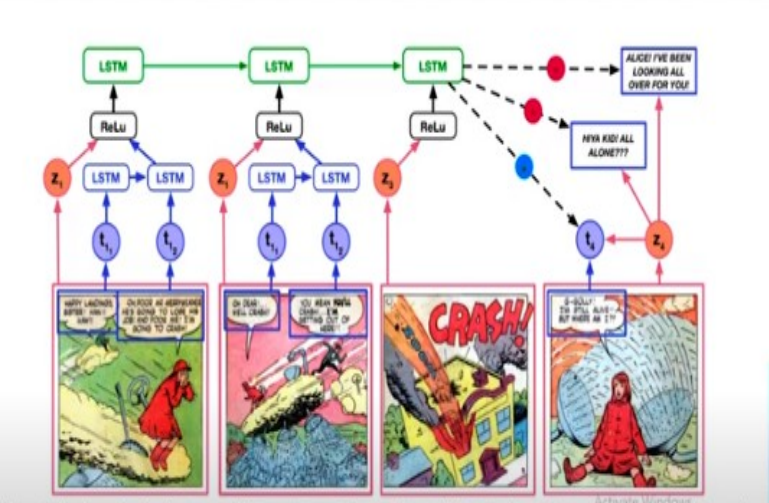


Figure 5. The **image-text** architecture applied to an instance of the *text cloze* task. Pretrained image features are combined with learned text features in a hierarchical LSTM architecture to form a context representation, which is then used to score text candidates.

But before we move on, now see it was a sort of a morbid presentation, where I sort of, did not seem to like a lot of things. But I wanted to show you one AI usage. And I wanted you to think how, you know this could be used in ways which would have impact on human beings. It is a very obscure piece of AI, which I once had a hand in making. We collected around two and a half million comic strips.

And all the images in them and all the text in them was annotated. And we created a machine learning tool, which given a series of images with the dialogues in them, predicts what the last image says. You know you have this person who is flying on this cloud like thing, and she says, oh, I am going to crash or something, and then she crashes.

And then there is a crash and says crash, like you know what was the Chacha Chaudhary comics where you had a punch and you had dishum written over it. And then she crashes. And if, and the machine can predict the last box that what would she say? And she would say, oh, golly, I am still alive or something like that. Now looks very fun, right? But then something like this could theoretically be part of systems which can have deep socioeconomic impacts.

Think about it, how. Like, really think about it. I will give you a hint, banks. Banks, banking. Think about how could something like this be used for banking? You know how, like, what is one of the primary areas banks are trying to cut costs on, especially American banks? So there is a lot of humans who do non-banking stuff in banks;

tellers, people who talk to people, people, you know who have these complicated scripts, what would the person say.

And then I would say that, that sort of stuff. All of that is going to go away. And this is an industry worth hundreds of attach the zeros and think about it. Should talk, you should delve deep into how much automation you just want to happen in the banking sector. And this paper is very interesting. If people are interested in such cartoonish usages of AI, please read, it is on my webpage.

I do not know why we collected 2 billion comics, or 2 million I have forgotten how many but I guess I just like comics. Now I will show you a movie. And that movie has nothing to do with AI and ethics. It is a very speculative movie by Satyajit Ray. Credits for finding that movie goes to Bidisha, professor who pointed out to me and she said you might find it interesting. We have like 20 minutes left.

And because I have another session right after this so we will have a large discussion like a very structured sort of a discussion on various things. Because like in this one I have tried to say I do not like the current modes of ethics. So in the next one, I am trying to you know not be such a negative person and try to propose an alternative way of looking at ethics. And after that we will have a very detailed sort of discussion.

So in this one we can use the 20 minutes to watch a movie. It is not a realistic description of AI; a Satyajit Ray movie, it is all very imaginative. But still has some good questions, philosophical questions you might want to. **(Video Starts: 1:03:55) (Video Ends: 1:25:26)**.