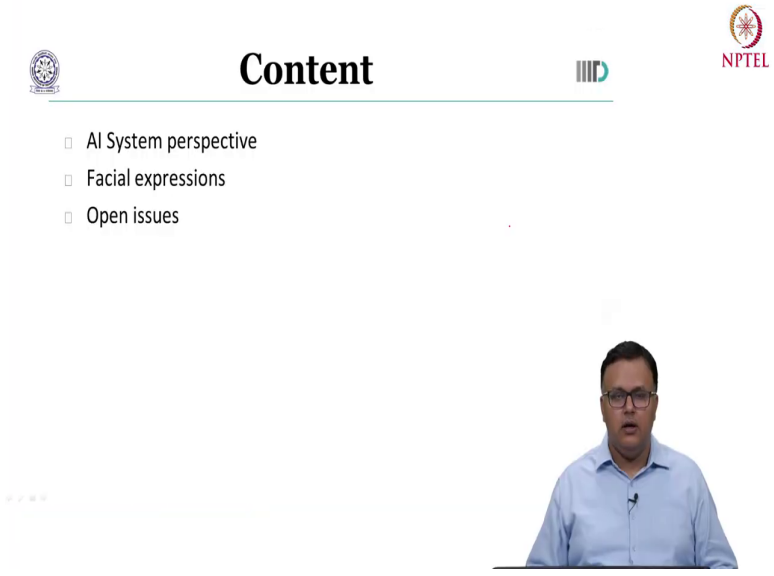
Affective Computing
Dr. Abhinav Dhall
Department of Computer Science and Engineering
Indian Institute of Technology, Ropar

Week - 12
Lecture - 02
Ethics in Affective Computing

Hello and [FL]. I am Dr. Abhinav Dhall from the Indian Institute of Technology Ropar. Friends, this is the second lecture for Ethics in affective Computing. This is week 12 of the affective Computing course. So, in the first lecture, we touched upon the concept of why its important for us to consider the ethical considerations, aspects when designing an emotional aware app. And then we discussed about the privacy concerns from the sensor and the software perspective.

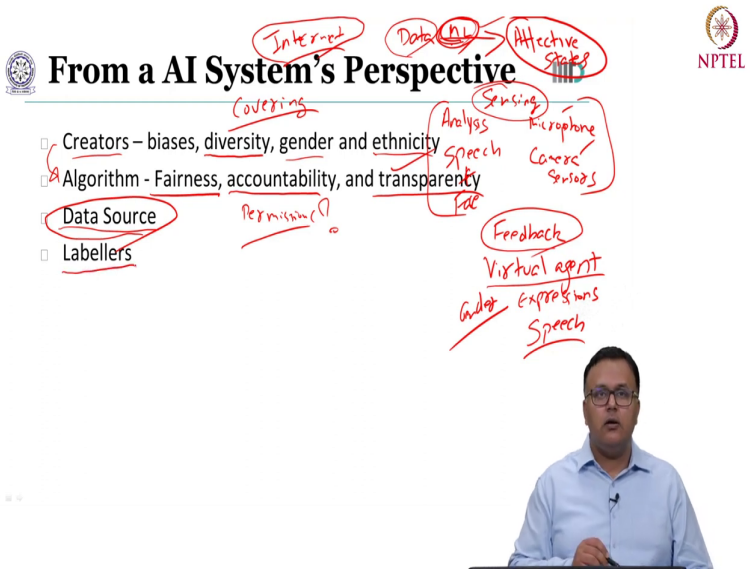
(Refer Slide Time: 01:05)



In today's lecture, we are going to move forward. And first, we are going to discuss about the ethical concerns from AI enabled system pipeline. So, we will are going to take an example of an emotionally aware AI machine. And we will see how the different stakeholders can add the prices and which can later lead to the ethical issues in the system when it is deployed for the user.

Then we are also going to talk about facial expressions in per se. So, how you know the ethical considerations need to take into the picture and what are the concerns when facial expressions are being analyzed. And then we will conclude by discussing the open issues.

(Refer Slide Time: 01:58)



Now, we are going to talk from a AI systems perspective ok. Let us say we are part of a large team, which is creating system for interacting with the user. Now, what the system does is, it analyzes the speech and face of the user.

So, you have the microphone and the camera sensors. Now, this is the sensing part. The feedback part after the affect has been sensed is in the form of a virtual agent, which of course, is showing facial expressions, which are response to the sensed affect and also is using a text to speech and varying there, ok.

So, let us say this is the system which we are going to create to understand the emotional state and then you have a conversation with the user. So, after the requirements having gathered by the team, let us look at once the requirements are understood by the team, the creators, what call could be the issues with respect to the privacy and the ethical concerns here.

First of all, the designers themselves may have certain biases based on their training right. You could consider an example, let us say all the developers are males and are from a particular ethnicity.

Now, they may have their own set of beliefs, which of course, is then going to effect the design of the system. Now, from the example which we have taken. So, let us say they may choose all the virtual agent avatars to be of a particular gender ok they may say, well, you know we only will have the avatars of a particular gender.

Now, friends, you can very well imagine the problem with this right. So, they are designing the system and they are influencing the system based on their own beliefs, which is due to the biases which they may have. There is a problem, maybe you know this hypothetical example, team which we have taken, there is lack of diversity on that team, which means there could be lack of different perspectives, which would go into the designing of this particular machine.

There is lack of gender diversity as well. And maybe it its just one ethnicity or two ethnicity is not, people from all spectrums are part of the team which is creating. So, there is a

limitation in the perspectives which are being you know taken into consideration when this affect sensing and reaction machine is being created.

So, there is a large possibility that there is a bias which comes from the creators themselves, because they have a certain notion of what could be the emotions and how a machine should react when the emotion of the user is sensed. The other is these creators would choose particular type of algorithms, right.

So, for the microphone and camera sensor data to be mapped from this data to affective state, as we have studied in the earlier lectures, we will have a certain machine learning algorithm there. Now, this algorithm itself, the nature of the algorithm may lead to some unfair behavior, some unexpected results with respect to the feedback.

Because, when it is sensing, maybe it is not unbiased to certain type of data because of how it has been designed by the creators. Further, it needs to be fair, right? Of course, you know, the you would not expect your machine to give incorrect emotion state results for a particular set of users belong to a certain you know corner of the society. And there needs to be accountability as well.

So, why does the machine think let us say the user is happy ok or the user is sad, because the feedback is also based on this. Further, there is transparency required. Why is your machine reaching at a certain conclusion about the affective state of the user, right? Now, if this is not taking into consideration, let us say the machine senses certain behavior, certain affective state of the in the user and the virtual agent starts to react accordingly, which in reality could be totally opposite of what the user might expect, right?

Then from the creators perspective, when they are designing the system, they would like to understand that how a particular machine learning algorithm is reaching at a particular state. So, the transparency aspect is extremely important. So, that you can understand, you know, you can interpret that why a particular affective state has been predicted.

The same goes from the algorithm perspective for the virtual agents feedback as well. Since you are expecting your virtual agent to show certain facial expressions to vary the speech as well. So, how is this mapping from the user's affective state to the synthesis of emotion is going to be done and then the algorithm needs to be having these particular attributes of fairness, accountability and transparency.

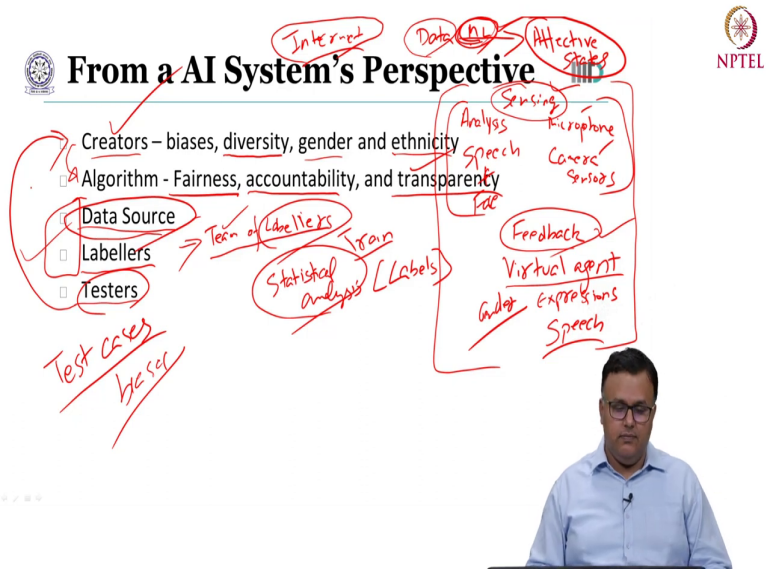
Now, another aspect is you have chosen your algorithm and you would now be collecting data, right? And friends, this is one particular stage at which large number of biases can be added to a system. For example, when you wanted to create this sensing part of the system, we wanted to train your machine learning algorithm, this data, let us say you were sourcing from the internet, ok then is this data which is being sourced from the internet, is that actually covering the spectrum of users who could be using the system, does it have enough representation from different age, genders and ethnicity.

And another aspect is, now this is again from the ethics perspective, what is the source of the data you are, let us say taking it from the internet to train your machine learning algorithm, but have the permissions be in seat from the data creators, right. So, that is another ethical concern which comes into the picture when you are designing this emotion aware AI system.

That the data which you are using to train who owns that data and also what rights has the curator of the data given to you to use that data in a particular sense. And then based on the discussion which we had in the last lecture, let us say these are face images, speech patterns which we are downloading from the internet, these belong to certain individuals.

So, while training our machine learning algorithm, right this algorithm, are we implicitly encoding the identity information about the subjects whose data is being used to train? So, this is an extremely important step friends that from where to use your data coming, who is the owner of the data, what is the type of processing which you are doing on the data and then the aspect of who is labelling? So, let us say we downloaded hundreds of images, hundred thousands of images from the internet and then we have a team of labellers.

(Refer Slide Time: 11:18)



Now, the same issues which we have discussed for the creators come for the labellers as well. When a label let us say looks at an image of a person and then decides if the person let us say is showing the fear related expression or hyper related expression, then that judgment is also based on how they conceive a fear expression could look like and which of course, that means, that their judgment could be affected by a lot of biases which could be there based on their earlier experiences and their training right.

So, once the data is sourced and the before the labelling starts, its extremely important to train the labellers to tell them what is expected, how the system would be behaving, what is the type of data which we have collected, when what were let us say any meta information about the data which has been collected.

And since we are having a team of labellers, in order to avoid the bias which can come from the labelling, we would be expecting that we are covering the different genders and ethnicity within the labeller team as well.

And also performing statistical analysis on the labels which are created by the labellers because ultimately these two things your data and labels are going to decide how well your machine learning algorithm is going to perform, which in turn is going to affect the quality of affect sensing in your machine, which will directly affect the quality of feedback by the virtual agent right.

So, from the designer perspective, these are extremely important points which need to be taken care such that your data is not biased, your data is clean, your data's usability parameters, its ownership and all is clear and the same goes for labelling as well. The labellers are aware of what they are supposed to label, there is a consensus among the labellers and the bias is avoided as much as possible.

And then a battery of statistical analysis test are performed on the labels to see the consistency for example, across the labels which are generated. Now, let us say the data is carefully curated, the algorithm is decided and the developers, they are training the algorithm, they train the algorithm they see that you know let us say the training error is within what they expected and they are able to validate on a validation set from the data itself.

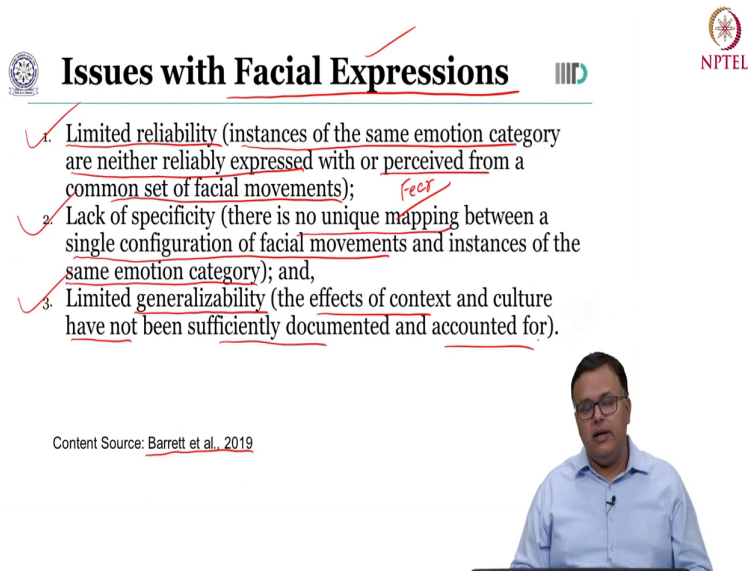
Now, before the system is actually sent out to customers, to users, there is a testing phase as well right. So, there is a group of people, they are testers, they are going to evaluate how this particular system is performing in different environments. Now, what that would mean is they would be writing test cases. Now, these test cases are based on the expected behavior when the system is given certain input. Now, friends here as well, there is a possibility of biases coming into the picture.

Again, testers are also human being, they would design test cases, they will interact with this machine, they will test this machine and their response would be affected by their own prior

training, their own prior experiences. So, there also you know these biases can creep into the system because ultimately your testers are supposed to give feedback to the developers, to the designers so, that the machine can be improved.

Now, what this means is at every step of development of a system, there are these variables which can affect the privacy, the ethics, [FL] you know aspects of the machine and also the privacy of the user. Therefore, we have to be careful about the biases which are added when we are designing such a machine.

(Refer Slide Time: 16:12)



Now, we are going to talk about one modality here friends as well, ok. So, since we were talking about emotions, sensing and then response, let us look at the facial part of it, the facial expression. So, I am sharing this data from the seminal work by Barrett and others and I quote the authors here.

So, they say well there are three major issues with facial expressions when they are used as the window to emotion in the part of affect sensing. And this is going to affect the performance of the system and as we have seen in earlier discussions, the prediction of the system once that is complete, how that affective state information is going to be used is going to have a direct impact on the user, right.

Therefore, the fundamental step of the information which we are going to use to sense the affect that also has limitations right. And these needs to be discussed when you are creating an emotion aware AI system. Now, in the context of facial expressions, Barrett and all argued, well the first issue is the limited reliability, ok which means you know there are instances of same emotion category that are neither reliably expressed with or perceived from a common set of facial movements.

And a typical example of that is fear. Fear is a very complex emotion, the way in which users would be expressing that particular emotion varies quite a lot across different ethnicities, different cultures. And what; that means, is when you are using facial expressions, there is a possibility if we are going to interpret an emotion like this.

And we are making a generic system based on an image or a you know set of frames which we are capturing, there is a high possibility that for a certain section of the users from a certain ethnicity, the results might be actually incorrect right. So, when you are deciding a particular type of label which you are going to use to reflect the emotional state, you have to be careful of the aspect that particular emotional category has to be reliable in itself with respect to the users.

The background of the users who are going to use this particular application which you are creating to sense the affect. The second point from Barrett and all friends is the lack of specificity, which simply means, well there is no unique mapping between a single configuration of facial movements and instances of same emotion category, a very simple way to understand this is when let us say you have two people from two different cultures, they are smiling, they might be using the region around the lips quite similarly.

But you know the how they are using the eyebrows and all that could differ, the intensity of the expression that could differ, what; that means is, the muscles which are coming together to show that particular expression on the face, the different muscles could be activated for different users coming from different ethnicities right.

So, again this goes back to the discussion which we are having from the creator, designer perspective, one has to be observant of this fact as well. Then let us say if you are looking at the facial muscles in the form of your facial action units, then after your (Refer Time: 19:59) facial action units.

The low level information, how you are going to use for the higher level understanding of let us say the stress or the mental health or you know these kind of very serious applications, this may be very user specific, this may be very context specific, right because different facial muscles would be involved to express.

The third issue with facial expressions as mentioned by Barrett and others. Is the limited generalizability? Right. What it means the effect of context and culture, they have still not very sufficiently documented and accounted for right. The response of users in certain environment, again coming from different cultures might be a bit different, there may be a bit of variability of how let us say someone would smile or frown or show disgust expression right.

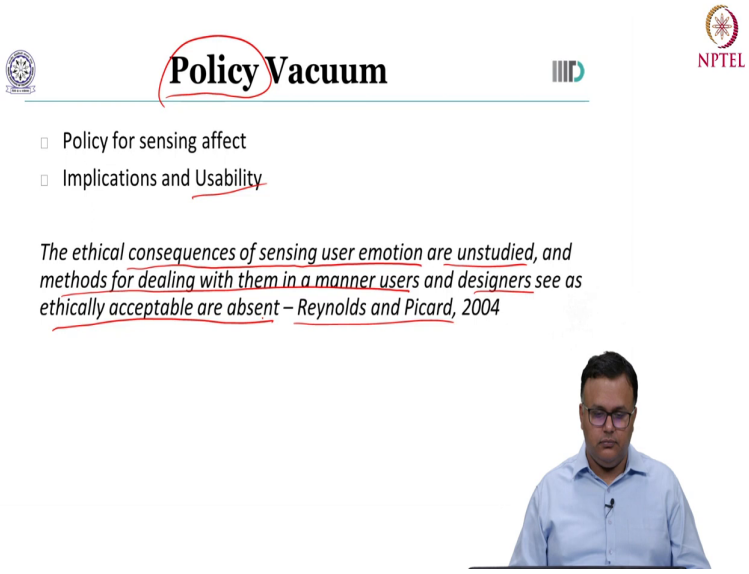
So, its non trivial to generalize and we have discussed this friends when we were talking about how affect is sensed for facial expressions. There this is actually a challenge when it comes to generalizability of machine learning systems for understanding the emotional state of the user. Now, the same goes back to the ethical concerns as well. If the creator of the system assumes that the system would generalize based on the data which they are using to train the machine learning system to sense the affect.

It is possible that if a particular ethnicity, culture or a certain spectrum in the age range, those users have been missed from the training data. They may get biased results when those

particular users are in the test phase, right because of the lack of generalizability of facial expressions.

And friends, you can also link this to the speech patterns as well. Similar issues, there is limited reliability of how people would be expressing when they are speaking. Same lack of specificity with respect to let us say the behavior in terms of the words which people will be using to express the same emotion. And then of course, there is limited generalizability as well when it comes to the speech patterns right.

(Refer Slide Time: 22:18)



Now, let us change the gears a bit and we are going to talk about the policy aspect right. So, if we are going to use this emotion sensing machines, certainly there needs to be a policy, there needs to be guidelines with respect to where the machine can be used, where it cannot be

used, on whom it can be used, for whom it can be used, who can create it, who can validate it right.

And the same goes back to not just emotion of AI systems, but AI systems is generalized as well right. So, since this area is. So, new, affective computing is not a very old area. There is still a gap, there is still lack of policy which needs to be formed right. And certainly, this is going to take into consideration that what are the implications, what are the pitfalls, what are the positive use cases of when the particular emotion, recognition and emotional feedback systems are created.

And again, you know who are able to use it? And I (Refer Time: 23:25) Reynolds and Picard here. So, the ethical consequences of sensing user emotion are unstudied and methods for dealing with them in a manner users and designers see as ethically absent acceptable are absent right. So, there needs to be a very careful deliberation with policy makers.

So, as to see that the way in which the creators invisionize the use of emotion aware machines is that actually how the users would be using it one and how it is safe for the users to actually be interacting with these kind of emotion aware machines. So, certainly policy needs to be taken care of.

(Refer Slide Time: 24:14)

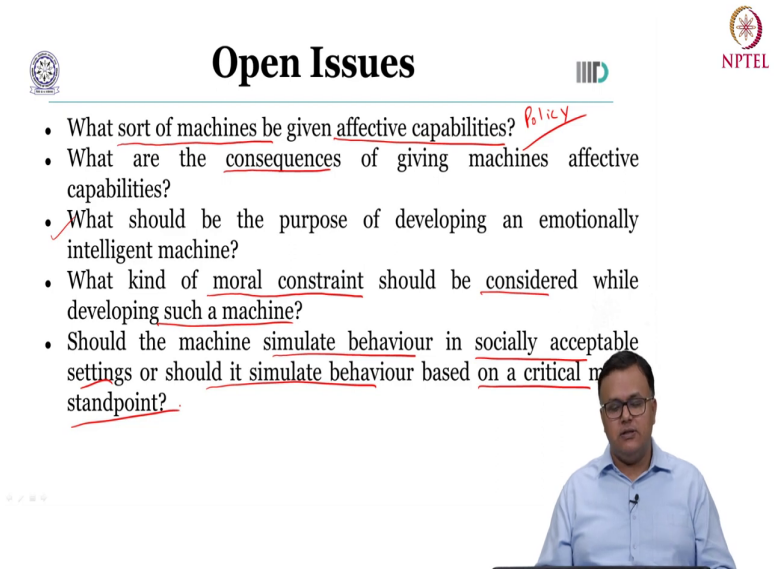
Now, again from the work of Reynolds and Picard, there is another interesting aspect that is about the user agreements and contracts. What that simply means is, let us say that you have a affect sensing app which is either implicitly getting the inputs from the user based on sensor data or explicitly it is making clear to the user that you know, well, high stress was detected. This is what could be done, a feedback mechanism is there as well.

What they found is that the users trust are more comfortable with the app when there are explicit agreements and contracts made right. So, the user when they are aware of what the app is capable of doing, how the data is going to be used, where it is going to be stored.

So, then they are more comfortable in letting the user sense their affect simply because a contract has been done between the app and the user and the user is very much aware about what the app is doing and will be doing with the data based on the sensed affect.

And what it means from the designer perspective again friends is that we need clear documentation, we need clear contracts. So, that the user is clear is about the capability of the app, what is going to be recorded, how is it going to be recorded, what is going to be analyzed, who has access to that data, right. And that way this is a mechanism to build the trust between the app and the user.

(Refer Slide Time: 25:53)



Now, we have reached to the final part of this lecture and we are going to talk about the open issues, issues which are still in the form of challenges and opportunities these exist for the

researchers and designers in affective computing. First, we still need to agree upon what sort of machines should be given affective capabilities.

Now, an example of that is should your ATM machine which you are going to use to dispense money, should that also be aware of the emotion, if there is a camera should it be sensing that when user came to dispense money, how are they feeling.

Or a machine in the form of a computer which is doing an online interaction with the user in let us say an interview based scenario, should that be allowed to understand the affective state of the user and then use that information for assessing, for giving feedback to the interviewer. So, this actually goes back to the policy aspect that which particular use cases actually require the understanding of affect and which use cases if are enabled with affect sensing might let us say be harmful to the user.

Then the other is you let us say we agree upon a particular type of machine, particular use case to have the affect sensing capability, then we need to know the consequences of giving machine this capability as well right. I have given you example of emotional code dependency and let us say a machine of user getting too dependent on the you machine to get their feedback about their emotional state.

And then of course, also you know the non-trivial aspects of a users affective state information being shared with other users or let us say clinicians in some use cases right. So, what are the consequences of it right.

So, we need to be very clear about that before a system is deployed. Further this comes from a more philosophical angle that what should be the purpose of developing an emotionally intelligent machine and friends this goes back to the first lecture in this affective computing course series right.

When we quoted a Professor Marvin Minsky right, he said well you know the question is not if intelligent machines can have emotions, but if machines can be intelligent enough without emotions right. So, the purpose essentially here is that we would like to enable these

machines with affect sensing capability in different ways in different in limited capacity or you know in a full-blown capacity.

So, that the ultimate aim is to serve the user have more engagement and more productivity for the user. Further there needs to be discussion on what are the moral constraints which need to be considered while developing a machine right. So, this example of a machine which is sensing the emotion during an interview right there are these moral constraints.

Perhaps maybe the interviewee is just stressed during that interview, but otherwise you know they are a person who does not gets stressed in a work environment right. So, will it be morally correct for the machine to sense that the person was stressed during certain phases of the interview and share that with the interviewer and maybe this could actually not you know work in the favor of the interviewee.

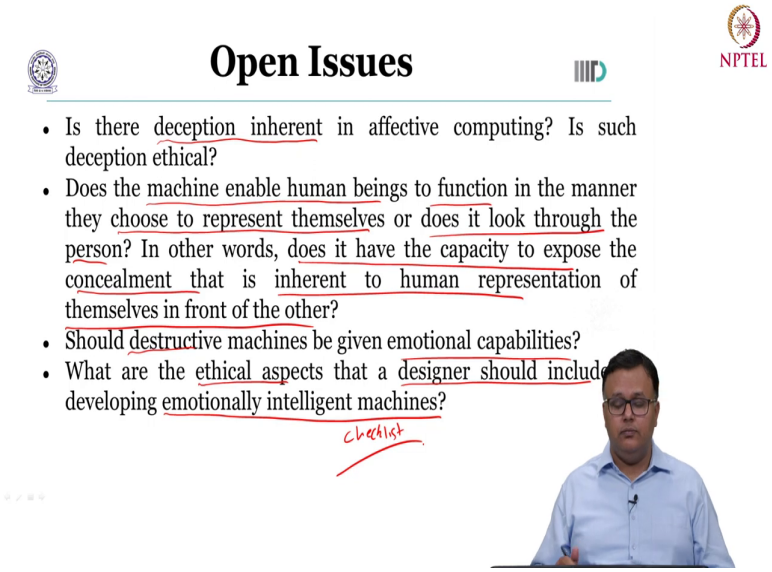
Further, should the machines simulate behavior in socially acceptable settings or should it simulate behavior based on a critical moral standpoint, right. So, once the feedback is happening once the machine has sensed that the user is having a certain emotional state and then let us say there is a virtual agent which is giving a feedback is this machine supposed to give a feedback as if it is a machine or let us say is a companion, right.

So, there is a moral dilemma there is it supposed to give the factually correct information or is it supposed to actually modify that information a bit or you know improvise on that particular information. So, that the emotional feedback is also appropriate with respect to the user. And you know this actually brings to a memory. So, for example, some of you may have seen this popular Hollywood movie called Interstellar right in that the user adjusts the funny quotient of the robo right.

They he gives feedback to the robo ok I want your responses to be funny of level 0.6, 0.8 because if its in a critical environment and you are expecting the factually supportive feedback from the machine an emotional improvision by the machine can lead to dilution of what was in originally expected from that particular machine, right. So, this is a very serious

point which needs to be taken into consideration with respect to how the feedback is being designed to for the affect aware machines.

(Refer Slide Time: 31:48)



Now, moving further in the direction of open issues the question is the there deception inherent in affective computing right. So, if the machine is understanding the affect state of the user and wants let us say to help the user, but in this pursue of helping the user let us say it is not showing it all the calendar bookings which the user has done.

On one hand the intent is to help the user let us say to the user to not get too stressed, but on the other hand this is deceptive you are actually withholding some information and highlighting another information. And that is a moral dilemma there right. So, how ethical it will be for the machine right.

And then then and then of course, this also means that the design of affect aware machines is should not only be limited to computer scientist, it should also take into consideration the feedback from different stakeholders coming not only from anthropology, but from the legal aspect from the you know the medical the psychologist aspect and so forth.

Now, further friends does a machine enable human beings to function in a manner which they choose to represent themselves or does it look through the person in other easier words does it have the capacity to expose the concealment that is inherent to human representation of themselves in front of the others.

Now, let us say you are in a formal environment you come across a person, a colleague and you know you have a very formal discussion with them. And during this formal discussion some facts are implicitly concealed its not that you know the intent is wrong its just that its a very formal discussion its for example, need to know basis, right.

Now, the same is applicable to machines as well now perhaps maybe a user is trying to conceal some facts is trying to hide their emotion due to any reason which he or she feels is you know right. But then what if the machine actually is able to sense that even though the user is relaxed and composed, but internally they are not because you know you are doing a multi-model analysis of the data maybe you are using physiological senses as well right.

So, is that right or wrong maybe the because of the context and the use case where that particular system is being used the machine is not required to do that access you know it its a formal communication. So, maybe something needs to be concealed again this means now if you are trying to understand the emotional state of the user.

And we expect the machine to be companion in achieving a certain task then it has to also understand that concealment for example, does not really mean that you know someone is trying to hide something for nefarious purposes it is simply based on the use case and the environment.

Further should destructive machines be given emotional capabilities let us say there is weapon machine you know machine which is capable of doing harm to the user or to the environment should they have emotional capabilities. For example, you know these anti missile based systems are should they be affected by the emotional state of the user should they be affected by how let us say the sentiment about a certain adversary is in the news or in general right.

So, that is a very complex question that is a very difficult question to answer, but till our understanding grows of how an in AI machine. And of course, then you know the next of the emotional emotion aware AI machine should be behaving should be reacting we need to have a simpler understanding of why you would need let us say emotional capability in a destructive machine maybe this particular use case does not require that or maybe it does.

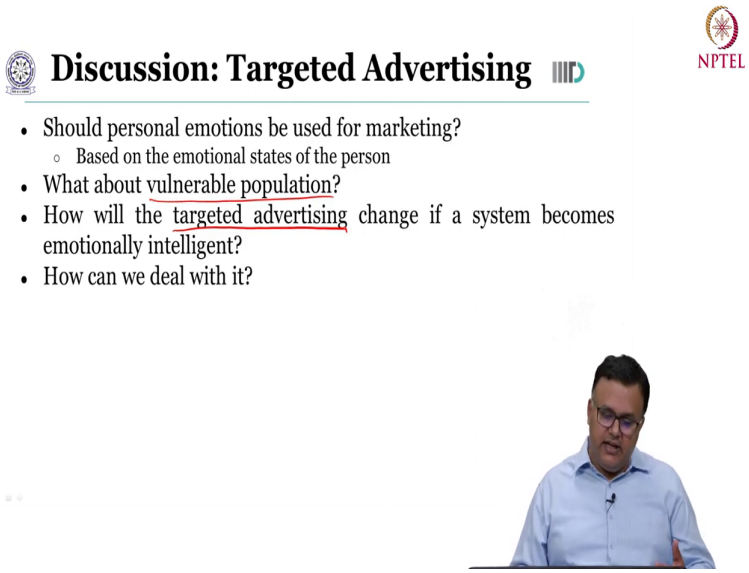
So, let us say if a machine which is capable of doing certain harm is able to sense that the user is too stressed and since the user is too stressed the user may make an incorrect decision which the user may regret later. So, what does the machine do? Does not really take the order does not really do the task which is the machine is being asked to perform.

So, there is no right or wrong answer at this point and which simply means that we as designers of this emotion enabled AI machines need to consult the outside world outside of computer science to understand the different challenges and aspects which could lead could affect the society and also how the psychology studies have been done to understand human behavior.

And what could be the response of a machine which let us say is trying to mimic or assist human. Further friends now what are the ethical aspects that a designer should include in developing emotionally intelligent machines, right. So, what is the checklist? You have already seen the discussion you know we have already discussed this when you are creating this emotion aware AI machine what could be the different aspects with respect to biases and data aspects coming into the picture right.

So, can we create this you know checklist which the designer could follow and then be observant of the ethical aspects which need to be taken into consideration? So, the sooner there is a guideline I think the better it is because we are seeing a phenomenal progress in AI enabled machines and certainly with these AI enabled systems, machines, apps being made more deployable being more accessible to the users it is important that we understand the human nature of the user as well the emotional quotient of the user as well.

(Refer Slide Time: 38:56)



Now, let us take an example friends these open issues with respect to targeting advertising. Now, let us say you are able to understand the emotional state of the user should these advertising advertisement which are shown to the user on a social media platform should they be based on the emotional state of the user.

A trivial example is the user is let us say feeling sad should the user be shown example advertisements of a chocolates, right? So, that is actually a very a moral dilemma right there right maybe solve the short term problem, but could lead to a longer term problem with respect to dependency of the user on these products which are being shown to the user in this targeted advertisements.

Now, should this user anyhow allow or the software which is being used by the user to browse let us say the browser should it be allowed to actually look at the emotional state of the user? And then what about the vulnerable population? Let us say you know there is this teenage teenager in some part of the world who is not feeling.

So, confident about themselves and then you know there are these targeted advertisements which let us say talk about a certain type of wardrobe a certain type of food which can you know enhance that particular teenager. So, that there are these very vulnerable populations right someone let us say who is sick and is searching for a certain answers and then you know through cookies system analyzes the let us say the emotional state.

And then shows them these advertisements which could be you know something which the user might not directly be looking for, but can strongly influence the user. Now, the other question is if let us say these questions are answered right. So, the limits of targeting advertising with respect to emotions are answered.

What would be the change in advertising world itself once these are the system becomes emotionally intelligent right? Currently the advertisements which are shown on social media platforms are based on the different attributes such as age gender browsing patterns so forth.

If you add emotion to it as well then it is going to have a massive impact on targeting advertisement right. So, before such a deployment is done its quintessential its super important that the affects are studied not only from the user perspective, but from the advertisers perspective as well and then you know how would we deal with these kind of

issues certainly again this goes back to the point which I was making earlier friends which is about the policy currently in the affective computing the policy part is missing, right.

We see that generally for AI systems different nations are and different consortiums are coming up with these policies these blueprints of where AI systems can be used how AI systems are going to affect the jobs and the user the user reaction the user knowledge and so forth. Same needs to be done for emotion and then when we talk about this use case of targeting advertising the same discussion will come into the picture.

So, friends with this we have reached towards the end of the second lecture of ethics in affective computing. We discussed about the effect of biases on privacy of the user which is based on the different stages of the development of an emotional aware AI system.

We discussed about how the facial expressions and speech patterns have limitations and that effects the privacy of the user. And later we discussed on the current open issues in with respect to the creation, the usability and deployment of emotion aware AI systems which will have massive effect on the user.

Thank you.