

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

Week - 12
Lecture - 01
Ethics in Affective Computing

Hello, [FL]. I am Dr. Abhinav Dhall from the Indian Institute of Technology Ropar and friends; this is week 12 in the Affective Computing course series. So, till now in this course we have discussed several aspects about affect, how affect is perceived from a machine's perspective, how we can create a machine using different sensors which can sense the effect of a user or a group of users.

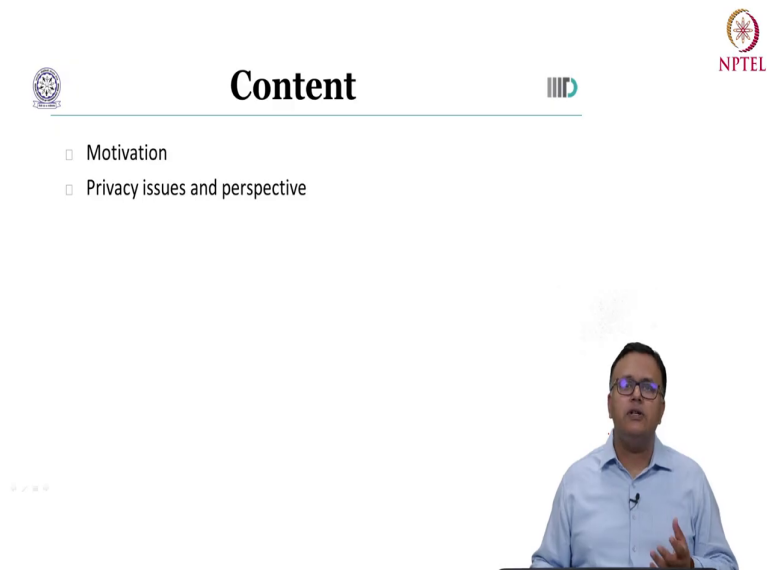
We have also seen how the different data sets can be created and what are the different applications, where affect can be used to have a more engaging, safe and productive experience for the user when the user is interacting with the machine. Now, what this also means is that affective computing slowly is becoming mainstream, so, for a more useful and productive experience of a user with artificial intelligence systems.

And with that be the very popular currently in 2023 the large language model kind of systems. It is important that after sensing the affective state of the user the machine also replies back in an appropriate affective way. Now, what it also means is as with any artificial engineering system a system which is trying to understand the user and then perform accordingly there are some issues, there are some challenges, which need to be taken into consideration before such systems can be deployed in the real world.

And these challenges and issues are essentially around the very important aspect of ethics. So, how are the ethical boundaries they been kept in mind when affective computing systems are being developed, are being deployed and what are the limits after which the use of such AI enabled systems can be counterproductive as well.

So, first we are going to discuss about the motivation why ethics is important. And we will see how let us say the concept of privacy that is being challenged that is sometimes effected when we are trying to understand the affect of a user.

(Refer Slide Time: 02:55)



The slide displays the following content:

- Motivation
- Privacy issues and perspective

Logos visible on the slide include the university emblem, NPTEL, and IITD.

A video inset shows a man in a light blue shirt speaking.

Later on, I am also going to share different perspectives with respect to sensors, with respect to the levels at which data is being processed, which is critical to the understanding of the ethical issues in affective computing.

(Refer Slide Time: 03:12)



Motivation



“The fictional message has been repeated in many forms and is serious: a computer that can express itself emotionally will someday act emotionally, and if it is capable of certain behaviors, then the consequences can be tragic”

- (Picard, 1997)



Now, this quote from Professor Picard and I will read out for you friends, “The fictional message has been repeated in many forms and is serious: a computer that can express itself emotionally will someday act emotionally, and if it is capable of certain behaviors, then the consequences can be tragic”.

What we mean here is after we have added the capability of understanding of affective state of the user into a machine and also have geared up the machine, we have taught the machine to react accordingly and have emotion.

So, once the machine will have an emotional persona and the actions of the machine are also affected by its own emotional state, which is driven by the external entities which includes

the user where the machine is then what effect can all of this have on to the user and the task which the machine is performing.

Therefore, the emotional identity of the machine and the emotional response of the machine that needs to be very carefully curated; such that, the task for which the machine has been created that is not hampered and we are also taking into consideration the aspects about the user, the interest of the user, the safety of the user and so forth.

(Refer Slide Time: 04:46)

The slide features the title "Affect Sensing Concerns" and three main points: "Affect sensing systems encode a designer's ethical and moral decisions: which emotions will be recognized," "who can access recognition results," and "what use is made of recognized emotions." A fourth point states "Users want feedback and control over such ethical choices." Handwritten red annotations include "VAD" with an arrow pointing to the first point, "Apps / cloud" and "App -> Stress" with arrows, "Speech" and "microphone" with arrows, "Designer" with an arrow pointing to the first point, "Spoken music" in a circle, and "GAM" and "aim" in a box. The NPTEL logo is in the top right. A presenter is visible in the bottom right corner.

Affect Sensing Concerns

Affect sensing systems encode a designer's ethical and moral decisions: which emotions will be recognized,
who can access recognition results,
what use is made of recognized emotions.
Users want feedback and control over such ethical choices.

Source: Reynolds and Picard, Affective Sensors, Privacy, and Ethical Contracts, Extended Abstracts CHI 2004
Slide Source: Prof. Beste Filiz Yuksel

Now, first we are going to talk about affective sensing right understanding of the emotional state through different sensors. And friends you recall we talked about how we can use a camera sensor for understanding of the facial expressions, body language of a user. Then we can use the microphone to understand the speech signal and that gives us very vital cues about the emotional state and then we have text and physiological sensors.

Now, from the perspective of an affect sensing system during its design the designers ethical and moral decisions they are also embedded into that system. Now, a very fundamental question is you are designing a machine, which is detecting the emotional state of the user. Now, how would you decide which emotions the machine should be trained to recognize. Now, let us say the example is you want to create an app which is analyzing the speech patterns of the user when the user is talking on the phone.

So, it is not actually looking at the semantic content of what the user is speaking, but essentially the pitch and fundamental frequency and so forth. So, as to look at things such as stress ok; so, the example is you want to create an app which is going to detect stress based on the speech, which is captured from the microphone in a mobile of a user.

Now, if you want to understand stress how are you going to define objectively the criteria and then what all emotions would you like to measure would those be the categorical emotions or you would like to go on to the valence arousal dominance continuous emotion scale.

So, that you can map the speech pattern on to the state and then detect if the user is stressed or not. And further if this app is created and is installed on a user's phone how much is the user's understanding of what all the app is capable of is it just through the contract which the user has agreed upon.

Or let us say there is a training phase as well wherein the user is made aware of the capabilities of this app and where we say well. The app is going to measure the emotions so, as to tell if during a certain point during the day when the user is using the mobile phone if he or she is stressed or not, ok.

Now, once this machine is able to assess the stress versus non stress state, which all entities within the mobile phone and outside the mobile phone are able to access that information. So, are there other apps which are installed on the phone are they being given access to the predictions, which the stress app is making are there any outside servers any cloud based installation where the user's stress versus non stress state is being communicated to.

Because if you are going to store the emotional state of the user either on the device then the method, which is used to safely encrypt that particular data so that any malicious user cannot read the private information of the user that would be required. Now, if you are sending the emotional state data to a cloud based system then one what is the safety mechanism there.

How the user's privacy is going to be maintained? Then other important questions are going to come into the picture. How long this data is going to be stored on the machine? And who all at the cloud end at the servers end is able to access this machine and then access this data do they have the rights to access this kind of emotional state data of the user?

Further once this is solved you know where the data is stored, and what are the security mechanisms around the data? The next question which arises friends is the machine has identified that let us say the user is feeling stressed ok, this is what the speech patterns are telling.

Now, what will the machine do with that, what would be the patterns which would be analyzed? That is first and once those patterns tell the app that the user is stressed, what is the use case, what is the application of that? And who all are going to use that information is it just for the mobile phones user that they would let us say be given a feedback, they would be given an alarm that you know you can look at the longitudinal data throughout one day.

And see how your stress was varying based on the speech analysis or this is going to be shared outside with let us say a clinician as well. Now, since this data is critical this is very much personal to the user. So, after let us say the clinician accesses it is the data being still stored at (Refer Time: 11:03) end is it at the clouds end? So, these are the kind of very important questions which are coming to the picture.

Now, let us say once this is also solved. So, we now we know from a designer's perspective that after the app is going to sense the effective state of the user after we have discussed and decided where, that information would be stored. What would be, let us say the security

parameters around that information then we discuss what is the use how are we going to use this information, which we have gathered from the app.



The next question would be what would be the way in which the feedbacks will be given? Ok. So, one example I have already shared with you that the machine can show let us say a graph ok. So, it could say for example, from 6 am to 9 pm this is how the stress you know kind of varied upon, ok. Are there other ways as well could this is quite explicit way are there implicit ways of conveying this to the user?

What would be the right time at which this information should be conveyed to the user? And if at all if the user is very stressed should this information be conveyed to the user that you are very stressed or should there be a feedback, which could explicitly help the user let us say in calming down. An example of this could be some suggestions regarding soothing music and so forth.


Now, from the user perspective the user would like to have control over these feedbacks as well. So, as now an app designer you will have to give these options to the user. So, that the user is aware about the whole gamut of how this data is going to be used and what all effect it could have on the user and let us say if there are some pitfalls as well.

So, these ethical choices these start at the designer's end. So, the team of researchers who let us say are gathering some requirement and then would be designing a system, which would be sensing the affect of a user.



(Refer Slide Time: 13:37)



Ethical Considerations



- ✓ **Privacy:** Emotions, perhaps more so than thoughts, are ultimately personal and private (Picard, 2003).
 - Affect aware interface, invasion of privacy?
- ✓ **Emotional Dependency:** Could having users routinely use these moral agents create codependence?
- ✓ **Emotional Manipulation:** Is it ethical for computers to detect, recognize, and then attempt to modify certain behaviors?
 - "motivated by good will" (Picard and Klein, 2002, p. 16).
- ✓ **Building Relationships:** At what point will a person begin to value affective technology and its well-being over that of another human being?



Now, when we are talking about affect sensing - these are the important ethical considerations friends, which we need to take into perspective. The first is privacy of the user. Now, in the context of emotions, emotions according to Professor Picard are perhaps more so than thoughts, since these are our internal thoughts these are our internal stage, these are personal and private, ok.

Now, should a particular machine be allowed to access this very private information and what would be the particular use case within which it needs to be allowed, right? Not every app should be allowed to sense the user's affect because not every app could be designed could be capable of the repercussions of the future steps after which the affect has been predicted because it is a very personal information. So, personal information needs to be very carefully taken into consideration.

Now, another interesting aspect is. So, you designed this app which is understanding the user's emotion and then let us say it is reacting by changing the color schema the font of the user interface of the app. So, now, you could say that the interface of your app is also emotion aware right it. The app is sensed the emotion of the user and then the response is also in the form of the visual changes which are there in the interface.

Now, since the app has sensed and the app is now reacting accordingly then is this also invasion of privacy of the user? Because the user has very personal thoughts and emotion is very personal to a particular person. So, that is from the designer's perspective it has to be extremely carefully taken into consideration that emotion unlike simple attributes, which are typically you know looked at a phone for example, how many steps one has walked, right. Unlike this emotions are far more complex far more personal.

The other important aspect, which needs to be considered from the ethics perspective for affect-sensing friends, is the emotional dependency. Now, the question is as follows with respect to emotional dependency. If a user is using an app which is sensing the effect of the user and then giving feedback to the user very frequently now the user might become dependent on to the app as well.

Now, this one way dependence right that could actually be unhealthy in some cases. Let me put this in simple perspective here. Let us say the user gets too dependent on an app for the feedback about their emotional state and they are believing the emotional state result which the app is telling them. So; that means, their future action their immediate future action might be influenced by the emotional state which has been communicated to them by this app right.

So, they will get dependent on the inputs from this particular app and that is a very fine line friends after which this dependency can become unhealthy. So, again from the designer's perspective the designer who is creating an app this is an important ethical consideration that the user should not become excessively dependent on to the app, ok.

You could also consider this particular point from the perspective of a virtual agent. Let us say the user is interacting with a virtual agent a 2D or a 3D character and the character is also able to sense the emotional state. And then the emotional response of the character and the personality of the character is tuned with respect to the emotional state of the user.

So, if with a longer use the user develops a kind of relationship with the agent the virtual agent then the user will become quite dependent on the virtual agent. So, you know these kind of remote cases also need to be considered. Now, friends the third point with respect to the ethical consideration in creating these affect sensing systems is emotional manipulation.

So, after the machine has detected the emotional state. And let us say you know the example which we were discussing about the user being detected as feeling stressed and then the app suggesting some music to soothe down you know some soothing music some relaxing music and so forth.

Then is not this essentially an attempt to modify the certain behavior of the user? So, in this pursue of giving suggestions to the user the app could eventually lead to certain behavioral changes in the user as well with longer you know continuous use of the app. So, is that actually healthy for the user could this create a longer term problems for the user. So, this needs to be taken into consideration while creating the systems for affect sensing.

Now, the another aspect is you could say that you know the system is motivated by goodwill because the user was stressed and the machine wanted to help the user. So, there is a good intent, but with long longer and very frequent use this may become counterproductive for the user as well.

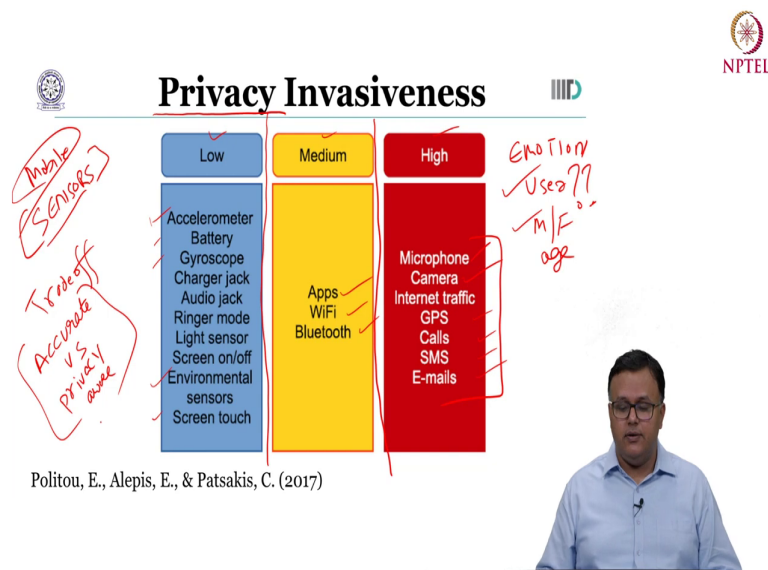
Now, the fourth aspect in the ethics consideration is building relationships. Now, in the very near future a person who is using these affect sensing apps very frequently. At what point will that particular user begin to value affective computing technology and its well-being over another human being?

So, consider the hypothetical scenario a person is using affect sensing app very frequently and the affect sensing app is telling it you know this is your emotional state based on let us say the valence arousal kind of emotional continuous emotion representation.

And the user gets so, used to this that it is now dependent to get the feedback from the app for his or her own emotional state rather than you know talking to or discussing any problems or you know getting feedback from the human beings around that particular user.

So, the effect of this particular user's relationship with his or her human beings around his friends and relatives around can in a very remote scenario these be affected if a particular user gets too dependent on getting emotional feedback from a machine and rather than not you know consulting with fellow human beings. So, that is a very important aspect, which needs to be taken into consideration.

(Refer Slide Time: 21:55)



Now, from ethics we are guys going to discuss privacy, ok. The approach which we are going to take is as follows. So, we will say well you know your mobile phone is essentially a combination of a large number of sensors, right. So, you have a large number of sensors and these are collecting information in different formats at different frequency, which can be used to understand the affective state of the user.

Now, we are going to talk about how these different sensors are capable if not handled very carefully by an app designer creator in evading the privacy of the user. So, these sensors friends, these are divided into three very broad categories, ok. So, these are based on their capability in privacy invasiveness intensity. So, you know if we have low, medium and high.

On the lower end of the spectrum where the chances of the invasiveness or loss of privacy of the user are relatively lower are sensor such as your accelerometer, your battery, gyroscope,

charger and so forth right. Now, let us take an example ok. So, you could use the accelerometer to understand the physical movement of the user, right. The aspects about the movement of the user when they are using the phone and this could be linked to how much active they are and that can be used to understand the affective state.

Now, this could be a very loose correlation, but what we are actually getting is just the affective state we do not get the information, which is about the user who is the user. And other attributes such as you know age, gender and so forth about the user. So, the personal information is still intact let us say when using an accelerometer.

Same goes for things such as you know your screen touch sensor and your environment sensor. So, the pattern in which the user is interacting with the screen. So, that tactile feedback that pattern can be used to map to certain emotional state of the user, but still that is it is not actually (Refer Time: 24:41) you know who is the user and so, forth.

Now, moving forward the set of sensors, which have a bit higher you know medium intensity for invasiveness in privacy of the user from a app perspective are the apps themselves. Some apps which are you know looking at the browsing pattern of the user, ok, and then things such as Wi-Fi and Bluetooth sensors.

So, where is the user right? So, based on the Wi-Fi router network to which a user is connected that gives very wider information about the user. Same is with the Bluetooth sensor as well being able to identify if there are other devices around the user also gives you information about the user right, is the user alone is the user you know in a group, where let us say there are other mobile phones which also have Bluetooth sensors around. So, you know one could create metadata out of it.

Now, from affect sensing the most critical set of sensors with respect to the privacy concern are these ones, ok. So, from the microphone you can understand the speech if you can understand the speech not only you can understand the emotion, but using speech one could identify the user as well.

Essentially who is the user is this user a male or a female and then things such as based on the speech signal what could be the rough age group, right. So, now along with the emotional state you are making the user very personal information accessible interpretable.

So; that means, if you are going to use the microphone in your app in your software you have to make sure that the user identity information that is essentially removed in the beginning of analysis of the data, which is being captured. Now, same goes for the camera as well friends you are able to record the face if you are able to record the face you know who is the user you again have these attributes like male, female, age and so forth right.

So, along the emotion if you are going to use the camera and you know since we are talking about mobile phone as an example you could have multiple cameras the front camera and the back camera. Again, you know from the privacy perspective that is tricky, right. Front tells you about the face back camera could tell you about where the person is right and who is the person with.


So, even though you know the intent of the app would only be looking at the emotions through the camera, but if not carefully you know created with respect to hiding the identity of the user and if the security information is not very well maintained this app could leak information could be hacked and so forth.


Now, again guys there are similar GPS where the person is call records again very private information SMS and emails tells you a lot more about of course, the affective state of the user. How the user is you know drafting emails, which tells us about the emotional state and so, forth, but in the same pursue of course, the email or the SMS could contain very personal very private information, which could be identifying the user as well.





So, what it essentially means is from a affective computing enabled app or software developers designers perspective you have to be very careful in what all sensors are you going to use to un sense the affect of the user. And in that very pursue be mindful of the vulnerability the user may have due to the app using these particular sensors.

And of course, you know from the app designer's perspective this would mean that you may have to do a trade off between things such as how accurate your affect sensing is versus how much privacy aware your app is. So, this is a very important trade off. Sometimes you may make the app a little less accurate just because that may enable you to preserve the privacy of the user, which is extremely important.


(Refer Slide Time: 29:17)





Stage	Sensing	Inferring Context	Predicting Context	Intelligent Actioning
Description	Collect smartphone sensor data.	Extract features from raw data. Machine learning connects features with higher-level concepts.	Build models of future events and predicted user behaviour.	Construct a decision framework based on past, current, and future events.
Challenges	<ul style="list-style-type: none"> Adaptive sensing Energy-efficient sampling Data storage 	<ul style="list-style-type: none"> Features and classifier selection Scalable machine learning Balance between processing on a phone and on a cloud 	<ul style="list-style-type: none"> Short- vs. long-term predictions—different forecasting horizons for different purposes Incorporate data from multiple users, multiple views 	<ul style="list-style-type: none"> Learn from mistakes: reinforcement learning for improved decision making Curiosity vs. accuracy: a value of a decision depends on how reliable and how proactive it is.
Example	 <p>Monitor user's voice as the day progresses. Regulate sampling rate according to resource levels and events observed.</p>	 <p>Process user's voice; create a Gaussian Mixture Model to identify user's voice and measure the stress level.</p>	 <p>Use personalised history of behaviour to predict a health hazard—a high stress level due to a busy workday.</p>	 <p>Reschedule user's meetings and their locations to reduce the future level of stress.</p>

Pejovic, V., & Musolesi, M. (2015)



Now, moving on in to similar direction we discussed about the different sensors, right So, from the same example which is analyzing the speech pattern of the user. Let us look at the perspective of the privacy from an affective computing apps angle from the different stages at which the information is analyzed and what could be the challenges which are possible in these very states, ok.

So, again you have your mobile phone there is an app which is analyzing the speech which uses the microphone, ok. And the software is understanding the affective state of the user. So, we are going to use the same example. So, the four stages here friends, are sensing, understanding, inferring the context. Once you have understood the context then predicting the context let us say the future context future action. And then intelligent actioning the feedback and what could the machine do once it has understood the affect, ok.

Now, you understand sensing would mean friends as we have talked about earlier. You are correcting the data from sensor in that example from the microphone. What are the challenges? Well, you know this could be the quality of data which is being collected from microphone could be dependent on the sensor itself onto the software, which is being attached to the sensor and also where is the machine being used.

Now, the machine which is mobile phone in this case could be let us say recording the speech samples of the user at a certain frequency. So, as to look at how the emotional state is varying for the user throughout the day, ok. Once this sensing has been done. So, we have collected the raw data comes the second stage right, inferring context.

So, now this is your machine learning part we are extracting features. So, in the speech it could be looking at your fundamental frequency and then extracting the MFCC or you know the representation learning pre-trained based features. So, once the feature has been extracted, we are going to then use machine learning to predict, ok.

For example, you could use a Gaussian mixture model or you could use you know a support vector machine neural network and so forth, which is measuring the stress of the user. So, we have now inferred the state of the user. The next is we have been recording the user and then we have predicted the stress versus non-stress.

If we have sufficient data it allows us to do predicting context that is we can build models for future events we can predict how the user could behave at this particular day time based on

the data collected from the n earlier day's, right. Now, in this there are some very important questions with respect to privacy.

If the user state is going to be predicted for a future event, how much the prior information is being used is that a short term information or longer term information, what is the context in which that information is being used? And then again, the points which I was discussing with you with respect to who is going to get access to this future prediction.

So, you could use this prediction of stress level to do some type of feedback some warning to the user as well right. Now, that would be our intelligent actioning. Perhaps when the user looks at the calendar and sees n events which are planned the higher precedence event or the events, which are just going to be happening in the very next few hours those could be shown first. So, that you know the user does not get sense that you know they have let us say a very very busy day ahead, right.

But in this very aspect the machine might anticipate that an event is far x is more important than event y and then hence the visualization is showing that more, but perhaps it is not the that particular way. May be one of the event is just a one-off event and that is important for the user, right. The machine could then actually end up being trying to regulate the behavior.

Further it is also question of curiosity versus accuracy, right? As we have discussed with the sense with the sensors perspective as well. What is the value of a decision, right? What is the cost if the prediction is wrong? And that again links back to the privacy concerns of the user with respect to how the app data is going to be used and who accesses and so forth.

(Refer Slide Time: 35:16)



Existing Approaches: Insufficiency

- Existing approaches which can help with the problem of privacy (e.g., cryptology, privacy-preserving data mining, anonymisation), are often insufficient both in terms of technological and in operational issues (A. Kapadia, 2009).
- There is no such thing like a foolproof anonymisation since almost all information can be defined as “personal” when combined with enough other relevant data (L. Sweeney et. al., 2013).



Now, there is a problem of insufficiency with respect to the current privacy approaches. So, current approaches they use things such as cryptography and then there is privacy preserving data mining approaches then there are approaches for anonymisation of the data, that is, removing identity information.

It has been observed that these often are insufficient by themselves or in combination when they are deployed right. So, even if the app predicted that the user is stressed and that longitudinal data is being stored in an encrypted way on the mobile phone, on the user's phone there are still possibilities that data could be hacked that could be you know decrypted.

What this also means is then after let us say the user has been given feedback about how they were feeling how the machine sensed their affective state is it really important to store

longitudinal data? Is it important to store very fine grain data? Because you know there are these privacy concerns.

So, again this goes back to the designer that if you are storing the information which has been predicted by a machine learning model do you actually store the exact predictions or some interpretations, which could be used by the user in longer term, but do not really effect or evade the privacy of the user as much as if the predictions themselves were stored and unfortunately there was an unaccredited access to that data.

Now, with respect to anonymisation, right, which is removing the identity information it is it is fairly complex to anonymize data because it is non-trivial to have a full proof anonymisation method. It has been shown in recent research works in signal processing and machine learning that how for example, the pattern of key press on a keyboard or on a mobile phone can be used to roughly identify who the user is.

Further things such as using the Wi-Fi signal using the radio you know the radio signal in turn can be used to look at how people are let us say walking or are doing certain activity in a room, right. So, one can deconstruct the identity information or some vital information about the user from the sensor data even though a designer may think that the sensor data is not recording the identity information the personal information of the user.

But with these newer mappings which are being enabled by the deep learning based systems one could make sense of the private information, which let us say is captured from the Wi-Fi signals or from the accelerometer plus gyroscope kind of data, right.

So, what this means again this is the whole discussion goes back to the designer themselves that they need to be very careful observant of this fact that even though they are trying to anonymize the data, which let us say is going to be stored or which is going to be analyzed at a cloud server end to predict about the user's affective state.

There are possibilities that the identity or a pseudo identity or the smaller set of attributes, which are still extractable from that data which could identify the user and hence affect the

privacy of the user. So, friends in this first part of ethics in affective computing we discussed aspects of why you know it is important to consider the different gamut of ethics.

How the sensors can be used to understand the personal data which from a designer perspective is extremely important to know. So, that they do not affect the user they design the machine such that the privacy of the user the ethical concerns which we have with respect to the affect sensing and affective feedback they are taken into consideration.

Thank you.