Hi friends, welcome to this week's module which is Emotional Empathy in Machines.

(Refer Slide Time: 00:31)

So, in this week we are going to talk about how can we create machines which are not only emotionally intelligent, but also which are empathetic. So, they have some sense of empathy. It sounds like sci-fi, but there are ways in which we can at least approximate it. So, here is an

outline for today's class. In today's class we are going to look at empathy, try to understand what empathy itself is and what do we mean by the empathetic agents.

We will next talk about how can we develop artificial empathy. Of course, in order to develop artificial empathy, we need to understand how empathy can be developed is developed naturally among humans. Next, we will talk about how can we evoke empathy using these techniques that we have learned from the natural empathy evolution in the humans.

Then of course, we will be looking at how this empathy can be used or has been used in the virtual and robotic agents. We will talk about how empathy is something that is not just an emotional state, it is beyond emotional state and how can it be evoked that. And we will finish the module with the evaluation of some performance metrics with that can be used to understand the naturalness or the empathetic of the (Refer Time: 01:50) empathetic interactions that are happening between the humans and the agents, perfect.
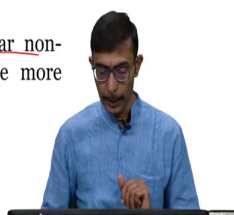
Empathy & Empathetic Agents

So, with that, let us dive in with the first topic, empathy and the empathic, empathetic agents.

Why Empathetic Agents?

- Empathy is the capacity to understand what other humans are experiencing.

- Presence of empathic responses by AI (machines, virtual agents and robots etc.) leads to better, more positive and appropriate interactions.

- **Assumption:** Humans prefer to interact with machines in the same way that they interact with other people.

- Attributing familiar humanlike qualities to a less familiar non-humanlike entity can serve to make the entity become more familiar, explainable, or predictable (Epley, N. et.al., 2007).

So, the very first thing that we want to ask ourselves is why empathetic agents and in order to answer that, we will need to let us revise first that what empathy is. So, empathy as we all know, empathy is the capacity of humans to understand what other humans are experiencing. And you may understand that this is different from sympathy.

So, this is basically an experience that you try to have what others are going through and then in turn this helps you to comfort them in a better way, right. That is what the empathy is all about. And of course, this when as a human also you may have seen that you want to we want to interact we prefer interaction with the humans, we want to be around with those friends and colleagues and family members those who are more empathetic to us rather than towards those who are not so empathetic.

The same idea is there behind them empathetic agents as well. And then underlying hypothesis is that if we can provide if the virtual agents let it be robots, machines, services any emotionally intelligent component, if it can have empathetic responses then that will lead to better, more positive and appropriate interactions with the humans. And this is what we want to enable.

And this is something that is already this is what happens when two humans which are empathetic to each other they interact, right. So, there is a better conversation interaction, there is a more positive interaction and it is also at the time it is appropriate when to laugh, when not to laugh for example, and then things like that. So, this is the same thing that we want the agents to have so that they can behave in the similar fashion to us and the entire experience can be more positive for the humans.

But it turns out that why giving this empathetic response also is going to make virtual agents more empathetic and in interact more interactive is because the underlying assumption that we have is we humans we always prefer to interact with the machines in the same way that we prefer to interact with the other people.

So, we want other people to be empathetic. Similarly, we would like to have machines to be empathetic. And this is something that is not a very new phenomena something that has been studied since long. So, the underlying idea is that we whatever entities that we are interacting with if we can attribute human like qualities to them.

If we can attribute human like qualities to those entities then that entity becomes less familiar to us even if it is a non-human like entity. And in turn if it becomes less familiar to us then you know we become more familiar and the entire experience around this entity become more explainable or predictable to us and hence we are very comfortable with that. So, give you an idea.
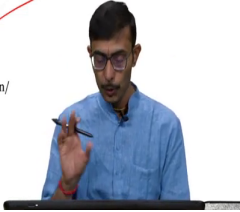
We can look at something that is known as anthropomorphic design some of you may have heard about it, but those who have not we are going to look at this. So, this underlying assumption that we are deriving that humans would like to interact with the machines that have human like properties is because we always have this tendency to feel better comfortable around the things that are that have human like features. And hence we know we develop a tendency to provide human like characteristics to non-life like artifacts. And this entire design is known as anthropomorphism.

So, anthropomorphism of course, is a Greek word which is made with two words anthropos and morphe; anthropos means human morphe means shape or form. So, basically anthropomorphism is all about having human like shape or form. But we will see that it is not only about the shape and the form it is also about the behaviour.

But let us see take an example here. So, the word anthropomorphism itself says that ok, its human like form or shape. So, but as I said it is not only about the form or the shape anymore, it is also about the way they interact with you. So, we would like them to interact as humans interact with us, right.

For example, looking with 2 eyes with 2 eyes rather than 4 eyes who is stopping you to create 4 eyes in the robots or in the agents or in the systems. For example, no one right, but then it turns out that the humans are always familiar humans would feel more more familiar and comfortable around the robots which have 2 eyes. Because of course, they are human like right, because of anthropomorphic nature of us.

But similarly, so this is behavior and this is also about the interaction and this is what we are talking about the empathic interaction. So, we would like to have the non-human like artifacts to interact also with us just the way humans interact with us, right. So, for example, imagine that you can have a robot which can communicate to you using your brain signals ok, that sounds resonating, but then at the same time this is not how humans communicate to other humans. Well, most of the humans will not do that some may claim, right.

So, the way humans communicate with other humans or interact with other humans is for example, maybe using voice of course, using gestures the way I am using and the way we use with each other things like that. Similarly, we would like to have the machines which can communicate to us using voice you see and communicate to us using gesture and so on so forth. Of course, facial expressions and on all that, right. And this is where we are talking about empathetic interaction as well.

So, you can see this image on the right hand side I hope this image is clearly visible to you. So, if you look at this image for example, you know this looks cute to us, this looks appealing to us or at least to the children and to many of us also. And this is the underlying you know idea why we like cartoons, why we like caricatures and all that. Because for example, here you can see this is the picture of a mouse, but the mouse is depicting human like characteristics.

What characteristics? Many different characteristics. For example, one of course, it is having glasses it is funny for a mouse to have glasses, but ok humans have glasses. So, it makes bit more funny and more human like. It ok, of course, it is reading of course, it has you know dress and then like humans of course, it is sleeping like humans you know on the pillow and things like that and it has socks like humans and all that, right.

So, basically all these are the human characteristics. So, if you see the form also ok, I mean it is trying to have glasses and trying to get the human like form, behavior it is reading ok, there is no interaction, but it can also interact you know maybe using voice and all that.
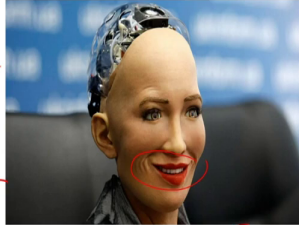
So, this is what is the anthropomorphism all about and because of the anthropomorphism we have a tendency to provide human like characteristics to the non-human like artifacts. And why we do that? We do that because it makes the entire interaction more comfortable for us and hence, we get a more positive experience for the humans. So, that is the underlying idea.

(Refer Slide Time: 09:05)

But then we have to look at this anthropomorphism with a bit of catch as well. So, it turns out that there are three things that we are talking about or mainly two things that we are talking about let say. One is the (Refer Time: 09:18) appearance, one is the one is the appearance which is the shape or the form, one is the function which is the behavior and of course, the function can include the behavior and the interaction both.

So, it turns out you know of course, you know it impacts the way we how we perceive it how we interact with it and in what sense we build the long-term relationships with it, right. And this is what this word is going to be very very important for the people who are in the industry because one problem is of the retention of the customer services and customer itself, right.

So, the customer maybe they will start using your product, your services, your agents, your robots and all that, but with the time they are going to lose interest unless and until they are

able to build a long-term relationship with it and long-term relationships are only going to be built not only, but can be built easily if you can have the other services they can provide human life experiences to them, ok.

And, but there is a catch here. So, the catch that I wanted to talk about is that you know the appearance and its capabilities or the appearance and the function of an agent. We can talk about an agent because it becomes easier to understand, but as I have said multiple times there is the same idea can be extended to the robots, services and things like that.

So, the appearance of a robot should also match its capabilities in the sense the form and the behavior they should be in sync and of course, in turn it is it has to match the user's expectations. So, just to make the thing a bit clear please look at a picture the picture that you have on your right side. So, the picture that you have is I am not sure how many of you know, but this is the Sophia humanoid robot, ok.

So, Sophia is supposed to be the most advanced as of now the most advanced humanoid robot. And if you look at the entire image of the Sophia robot it an anthropomorphic design, why it is an anthropomorphic design? It looks like human, right. And it is behaving like humanity, sort of in a smiling also like humans, to certain extent it has eyes like human, it has a form like human again and ears and all that.

But then again what we want as I said, that we want the appearance and the capabilities to match each other and we will and when they do not match each other then what happens? Then the creepiness comes into the picture. So, then the entire experience becomes more uncomfortable rather than being comfortable, right.

Of course, I am not sure, but for example, for some it may be seen that ok, while it looks like human, but then at the same time if you look at the smile has some components which does not look like exactly like human. So, there is a mismatch between these two and maybe it is not going to give a very strong, sense of comfort that you wanted it to have, right.
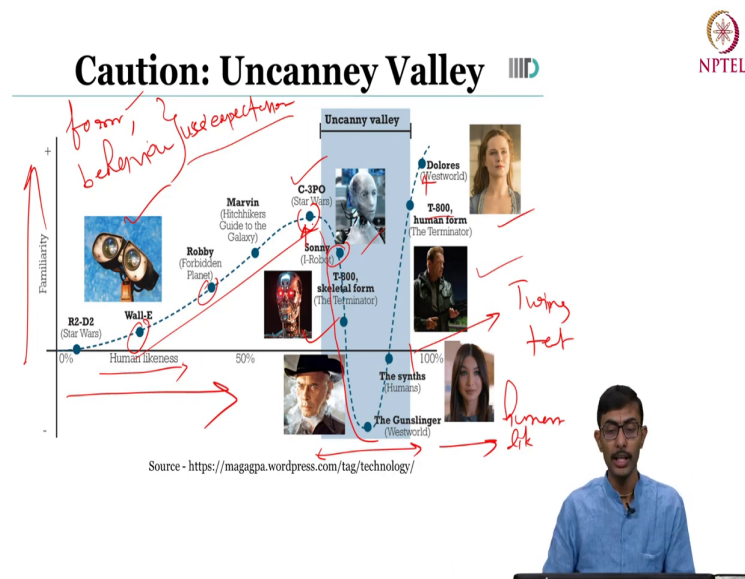
So, and it turns out this is a very well studied and established phenomenon which is known as Uncanny valley effect. So, the uncanny valley effect is what? Uncanny valley effect the term itself was coined by robotic system (Refer Time: 12:24) hero in 1970s which says that it is the dip in the emotional response of humans when there is a human like characteristic that is being achieved by the robots or by the agents, right.

So, for example, just look at this graph and it is going to become very clear. On the x axis what you have is the human likeness or human like characteristics ok, you are already have it is written here. So, human likeness or human like characteristics, right. So, basically as you go on increasing from left to right.

What you are seeing here that you have certain robotic characters, which are less human like such as for example, of course, R t D 2 you may recall very famous character from Star Wars

movie, of course, Wall-E you may recall and then the more you go on the right hand side then you come till for example, you know the since character in the humans (Refer Time: 13:21) series.

And then terminator in the T-800 and Dolores of course, in the Westworld right, they look very much like human. So, on the x axis you have humans or the robotic agents looking less human like from the like on the left hand side, but as you go on the right side looking more human like. Similarly, on the y axis what you have? You have the familiarity the sense of the comfort or the familiarity of the humans as rated by humans, right.

So, for example, what it is roughly trying to show you, the more the human likeness or the likelihood of looking like a human increases among the robotic characters the familiarity of these agents also increases with the humans and it makes sense, right. The more they are looking like human the more you are feeling familiar about with them and around them.

So, for example, if you look all recall R2-D2 or even for example, Wall-E character that is there in front of you I mean ok, it was so popular because at the same time it was not human, but at the same time it had some human like behavior and shape as well. So, for example, it had two cute eyes and you know it was able to navigate and things like that. So, not exactly like human, but it has some human like form and people loved it liked it a lot.

Similarly, was the case with the R2-D2 and so on so forth. And similarly, you know if you recall this again another very popular characters from the Star Wars movie which is C-3PO. So, C-3PO of course, also if you recall it was you know more or less like quite the it had like more human like characteristics and hence the familiarity or the popularity of this character was also quite high.

But suddenly what started happening now, you see then this is the dip that we are talking about, this entire reason is the dip that we are talking about. Now, suddenly in this area what we are looking at, that even though the likelihood of looking like the human or human likeness is increasing among the characters, there is a sudden dip in the familiarity.

So, of course, I mean we started we can talk about if you recall the I robot you know the Sonny character from the I robot this is the Sonny character from the I robot. Of course, you know with the T-800 skeletal form of the Terminator, similarly you know the synths as I said and so on so forth.

So, basically what is happening in among in the case of these all these characters the Gunslinger in the west world and all these characters, that they are looking they have started to look like more like humans. But at the same time, you know there is a mismatch between their form I mean I will note it down again so that you can recall it again. So, there is mismatch between their form and their behavior and more importantly the users expectations, what do we what do we mean by that?

So, if you look at the form, the form looks quite like human, ok. I mean for example, if you look at the Sonny robot itself it looks quite like human, but at the same time the behavior was not exactly like the humans, it has some superpowers also or maybe you know for example, there are certain things that was looking like creepy so, for example, all these things coming out of the head and things like that, right.

Similarly, if you look at for example, there a skeletal form of the Terminator I mean this particular character. It is looking very much like human, but at the same time it looks like a distorted human body I mean this does not look very good to us I mean this does not look very comfortable to us right, I mean this looks like I mean it is missing some limbs, some parts and this and that. So, it is like horrific it is not very so.

So, then of course, you know like we do not feel comfortable around this kind of designs. Because of course, one thing is the form is looking like human, but at the same time is not like humans and in the creepy itself. And then the form and the behavior maybe they both are also not matching you know ok, I mean if you are looking like human, if the robot is looking like human, it should also behave like human and it is really hard to behave like humans.

We have because of course, the naturalness that is there in the interaction, the voice and all that I mean it is not easy to plug in and of course, there is a lot of research that is going on this things. And more importantly there is a mismatch with the user expectations.

When you show a character which is looking like a human the users expectation is ok, it is going to behave like a human, but when it is looking like human and it is not behaving like a human, while behaving like human as I said is always challenging then there is a mismatch with the users expectations. And that is where you know the familiarity or the like popularity it starts dipping in and that is what is essentially, we can say is happening with the uncanny valley reason.

But then what happens suddenly you know then again, we see, but as we cross a particular threshold in a 90 percent or whatever you know like maybe something that is very very close to human. So, for example, this character of course, you know like played by a (Refer Time: 17:53) in this case and then of course, in Dolares Westworld, what is happening? I mean ok with this the likeliness has increased a lot, but at the same time the familiarity or the popularity also has increased a lot.

So, one simple thing ok, they are looking very much like humans and they are behaving very much like humans. So, is as simple as that, there is a very good match between the expectations of the users ok, if they are going to look like humans they are going to behave like humans and that is how they are behaving in this particular case.

And that is where you this is the point that you can also say that if you know or recall the turing test to certain extent this is where we can say that they have they may have passed the turing test also for the humans.

Because normal for a normal human it may be hard to differentiate ok, whether for example, this character is a human or this character for example, is a robotic agent because of course, they are looking very much like human and of course, they are behaving very much like human.

So, maybe they are passing the turing test also after this certain reason nevertheless I mean it is really hard to achieve you know this particular reason, but of course, I mean what you want to do is you not want to get your design trapped in the uncanny value effect, ok. So, so that is what is about the uncanny valley effect.

So, coming back to the summary of this thing what we wanted to understand? We wanted to understand why we want to have empathetic agents. We want to have empathetic agents because it makes us more it makes the entire interaction more positive as simple as that. And why it makes the entire interaction more positive? Because it gives us a sense of familiarity when the agents are interacting with us just like the humans are and this is what is known as the anthropomorphism or anthropomorphic design.

In the anthropomorphic design what we want to have? We want to have a human anthropomorphic design it allows or it is all about a tendency to attribute human like characteristics to non-life like artifacts, which may include their changing their shape to like humans, changing their behaviour like humans, changing their interaction like humans.

While we are doing it and that is underlying a hypothesis of the empathetic interaction, but while we are doing it we should be cautious about uncanny valley because it may happen that we are going to make their form like humans, but they are not able to achieve an interaction like humans and that is where there is going to be a dip in the emotional response or in their familiarity or in their likeness.

So, basically what we would like to have of course, this can be our holy grail in some this can be called as our holy grail which can be difficult to achieve you know like just human like form, human like behaviour, human like empathetic reaction, but maybe we can achieve till here and you know that can also be quite, good. Or for example, you know even here or even here or anywhere in this region, right, ok. So, with that is about the why empathy and what do we mean by the empathetic agents and why do we like them so much.

Development of artificial empathy

Now, let us try to understand the next module of the class.