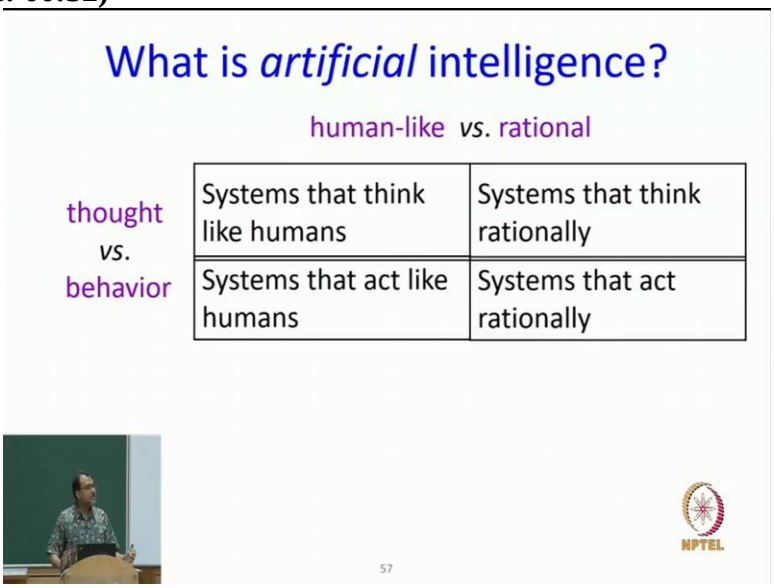
Artificial Intelligence
Prof. Mausam
Department of Computer Science and Engineering
Indian Institute of Technology Delhi

Lecture-7
Introduction: Definition of AI Thinking vs Acting Humanly vs Rationally, Part-7

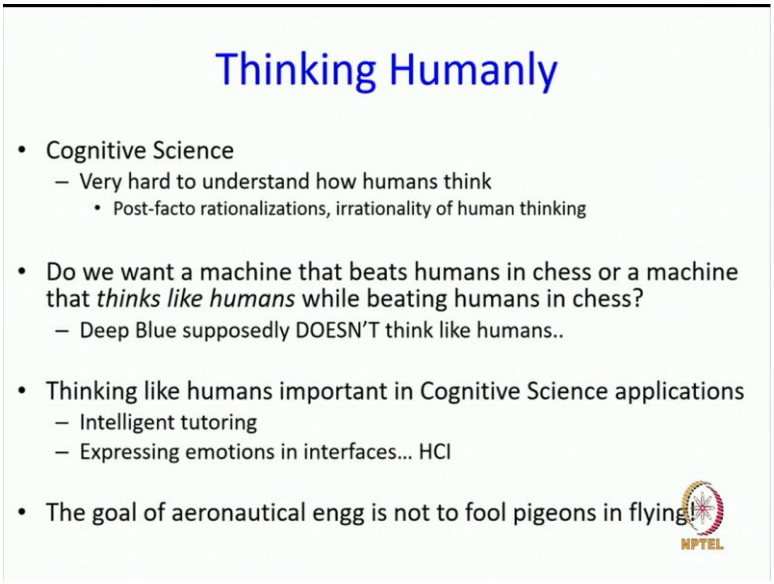
I would start with the disclaimer that whatever while I would critique these definitions I would also I would mention that many people still believe in these specific definitions and have their research trajectories in this specific, some aspects of it.

(Refer Slide Time: 00:32)



So, these are all good definitions, but for the from the philosophy point of view, one definition is now starting to win a little bit more and we will talk about.

(Refer Slide Time: 00:41)



So, let us talk about thinking humanly first with this had very few takers. It is also simpler to critique because if you ask Garry Kasparov, how do you think? He would not be able to explain that thinking we thinking is an innate process to us it comes naturally it is very hard for us to verbalize vocalize it put it down in words. So it is very hard to understand how humans think, more often than not humans do what is called post facto rationalization, like I believe that we just take decisions intuitively, often emotionally.

And then somebody says why did you do this? And then we just hook up reasons. And if we had done something else, and somebody had asked, Why did you do that, they would have still hooked up some reasons because in often in life, every choice has some pros and some cons.
But usually we are not doing a cost benefit analysis. Usually when we have a decision in front of us we do not. This is the benefit of this. This is the drawback of this.

I do not want to go to I want to if I go into medicine, then I will be helping people. If I go into engineering, I will be making a lot of money. If I can into medicine I will be on call for one and a half days at a stage if I go into engineering I can be working from my bed if I especially computer engineering. If I this, so then you say for me, making a difference to lives of people is more important.

So I will go to medicine or luxury and money is more important. So I will go to engineering. So, this is a caricature, please do not take me seriously of course. All walks of life extremely important, but this is how we make decisions. This is how we do not make decisions. More often than not, we do not do a cost benefit analysis and even if we do a cost benefit analysis, which is maybe a very, very fraction of percentages of us.

How do we say that making money is, we know, worth 10 points, and making a difference to people is worth 5 points and luxuries. We do not know what is the value of each of these things. It is an innate process of one thing feels more than bought into us one thing feels less important to us. So eventually, we do not make decisions like this. We make decisions by just intuitively feeling like I will go to engineering because I love games.

And I get to play more games, because I will be on the computer all the time, or something like that. I will go do engineering because all my family is set of doctors and I do not want to become like them come on, things like that. We make decisions very intuitively and emotionally. And then we sort of justify it to people. We do not want the AI systems to do that, of course. And moreover, we already discussed that humans are not always intelligent, we commit suicide we jump off the cliffs.

You know, if you start taking looking at these various type of crazy videos where people are just by definition, doing crazy things, we realized the whole average intelligence of mankind may not be very high, and we do not want to really take AI systems and put that put it there, but we do not want to have AI systems having breakups and, you know, crying through the night and in the not functioning in the morning and so on.

So, let humans do that kind of. Moreover, even from the philosophy of the field, it is not important how deep blue thinks. It is important that or how Garry Kasparov thinks it is important that the deep blue defeats Garry Kasparov. See, that is the demonstration. Whether you defeat Garry Kasparov by thinking like Garry Kasparov or thinking in a genuine way, it is all kosher. See, moreover, and we will see this again and again and again.

Humans have some hardware, actually, you can think of humans as intelligent beings. And we have our own hardware. We have hardware of the brain in this hardware of the brain, we have some neuron connections. And we have some specific electrical signals that are going in, that are placed in a certain way in the brain and have been trained in a certain way as you were going and through our genes.

The machine has not had the exact same training process, the machine is a different machine, we may have a supercomputer where the neurons or the processes are arranged in a different way, we can get an approximation of the brain in the machine, but we do not really know exactly how our brain is structured, we only have an approximate idea, etc. So because of hardware is different, it is very hard for us to claim that we can build a machine that can think like humans, and in fact, it may not lead to success.

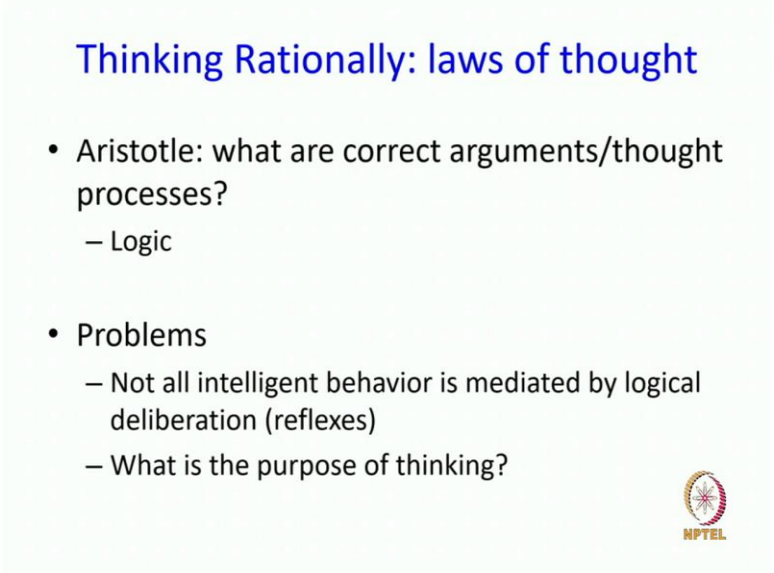
We have massive parallelism; the machines do not have that much massive parallelism in general, etcetera. Of course, we have big quantum computers things change, but for now, that is the case. So therefore, a said look, thinking like humans is a reasonable endeavor. You can spend a lot of energy and thinking about doing psychology experiments of behavior, that experiment and taking insights from there and a can use it.

We do not mind using it, you can do a lot of work on, you know, understanding brain and how its connections are set up, great, do all that work, we can take it eventually if they succeed. These are all worthy endeavors. And especially the field of cognitive science actually studies how people behave, and various models for how people's act but or think, but eventually, for AI, that is not the definition. Intelligence is more than humans and thinking is not sufficient.

And thinking like humans and replicating all their behavior is not ideal. I should point out that thinking like humans can be important in some applications. For example, if I am building an intelligent tutoring application, then understanding how humans think and how they learn is extremely important to give the right advice, or if I want to have like an elderly care robot, which is going around and in an elderly care home, you know, helping the elderly people take medicines on time, you know, be their buddy and so on.

So forth, then you need any system which can understand all people's emotions, which can also, you know, express emotional intelligence. And therefore, understanding human behavior will continue to be important in various applications, but it is not necessarily the definition of here.

(Refer Slide Time: 07:14)



Now, let us come to thinking rationally, which is a really good definition, in my opinion. And it comes from, you know, the old, you know, thousands of years history of what is the basis of a correct argument or a thought process. And this, all this endeavor started in a way back with Aristotle and earlier, where, you know, they created this whole logical framework, so, you know, what is the right reason? Or what is the right reasoning or what arguments can be proven in a given context and so on, so forth.

A lot of people believe this is a good definition of more than you would be and lots of people spend their lives work on studying logic and improving logic. And it is a good idea except that there are 2 problems which are commonly suggested. One is that not all intelligent behavior is mediated by a logical deliberation. So as a simple example, let us suppose I by mistake touch a hot stove. What do I do? In a reflex I get my hand out, quickly; I do not want to be born.

Would you call this intelligent behavior? In the given context, getting my hand back from a hot stove would be considered intelligent behavior. Now, how did I decide this? It is actually really important to think about this. I did not do a logical deliberation I took a reflex action. But if I had

done logical deliberation, it would have gone something like this. I have just touched the stove. The stove is extremely hot.

If I keep touching it, slowly my skin is going to born. And that will be very hard for me to integrate your papers. Therefore in order to, for me to allow myself to grade your papers, I should remove my hand from the stove. By then it will be too late. The process of deliberation is quite accurate, quite appropriate. But the fact that we may not have time to do the deliberation, or a deliberation may not be needed in this case is actually quite important.

If we just remember the rule when touch a hot stove event as something hard remove your hand is going to do the work and will exhibit intelligent behavior. Therefore, it was felt that while logical deliberation is a cornerstone, an important process for most intelligent behaviors that we show, it is not for all, there are some which may simply be reflexive, which may not need deliberation. A similar example holds when we see a big truck coming in, and we are in the middle of the road and it is coming fast, we just quickly run away.

Believe it or not, this kind of reflexive behavior is extremely important in robotics. Can you think of one place where something like reflexive should be extremely should be used? One behavior one specific example of a behavior in robotics. Suppose you have a robot moving on the floor. There is one place specifically where the board is asking Think, just act quickly. Can you guess? The someone comes in front, what do you do? Stop or go back a little bit so that you avoid the obstacle. This is called obstacle avoidance.

What is your name? Kishore. So Kishore says, obstacle avoidance is one such thing. We need not be deliberating where imagine that if the machine fails, it is going to collide with something or it has collided with something quickly go back. That is a behavior that is very reflexive. That is a behavior which is part of most robotic systems. And that is the behavior which you will consider intelligent in the process, but it is not necessarily deliberative.

There is one other one other point here, which you may which you will also agree hopefully, so I will explain this to the job. This joke is not due to me it is. It is one of the top few jokes a few

years ago and I borrowed it from another professor, but so nice joke makes the point so Holmes and Watson Sherlock Holmes and if you know the joke, please do not ruin it for everybody else. So Sherlock Holmes and Watson go camping. They set up their camp and in a night, you know, just a beautiful night.

They sleep at and then suddenly homes wakes up Watson and Watson wakes up, you know, from sleep at home says, What is the look up? What do you see? Watson says, I see starlit sky. Home says so what does that tell you? Or Watson says, there are stars, which tells me that there is a huge galaxy, there are many galaxies and there is a moon which tells me that the sun gets the lights get reflected and we can see light even in the night. And you know, it also tells me that you know, I am very insignificant, it is part of the bigger.

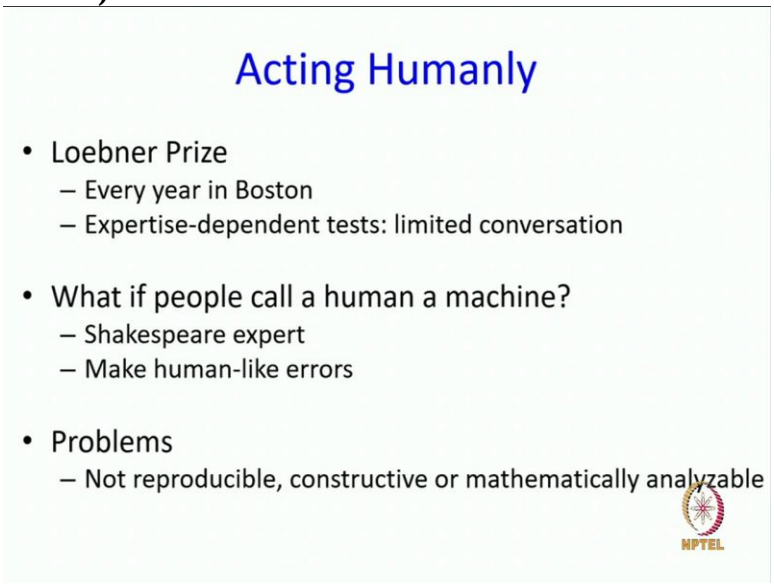
We are part of a huge galaxy this this this idea that he keeps on going and making all these scientific assertions and home says shelob says, what is on you fool? It tells us that somebody has stolen our camp, our tent. That is why we are able to you know, look up and see the sky sir? Nice a joke of course. But the question that this particular Joker is was Watson being incorrect? No, he was making all conclusions based on the observations and so on in terms of logic, he was accurate.

But at that point in time, you and I would agree that making the assertion that somebody has stolen our tent is more important than making the assertion that you know, there are so many stars in the sky in this galaxy and so on so forth. And so begs the question, is there a purpose to thinking? I remember, you know, I had a collaborator one of the shots best minds I have ever seen. If you ask him one line question, he will send three page responses.

If you ask him to make progress in the research, not so much, he will deliberate a lot. He will think a lot and then he will think about his problem and then somebody else will ask him a question he will think about their problem and somebody will ask him a question he will think about their problem. He will give them a lot of insights, but will never be focused enough to think about his problem and make loggers, the sharpest minds, but could not remain gold directed, very hard to get him to finish his PhD.

So you often ask, should the fact that I can think Is that sufficient for intelligence? Or should I be focusing my thinking in a certain direction? Should it be purposeful and more than more than so I just feel that thinking should be purposeful it should be directed towards a goal. And thinking rationally does not naturally express it, although in some ways it could later.

(Refer Slide Time: 15:13)



Last but not the least, we have acting humanly. In acting humanly, the idea is let us mimic human behavior. And obviously, I think the main argument against it is that humans should not be your ending point for intelligence. It could be a temporary milestone in the middle intermediate milestone. But once you have achieved that milestone, you should keep going forward. And now there is a funny story about the Turing Test which was mentioned by (())(15:48). So a Turing Test is now operationalized.

In this loebner price, which is a an event which happens every year in Boston, where the Turing Test is specifically the You know, people present their software's and humans are brought, and judges are blood. And the judge asked questions from the screen and the human or the machine response. And the goal is that someday, a machine will be able to defeat a human into a defeat the judge of confuse the judge in making them believe that they are not a machine. And the loebner price would be one itself; I think 100,000 dollar prize or something.

And they have specific verticals like they will do something on a specific kind of dialogue system or a specific kind of theme and sorts of so these are expert eyes dependent tests. So, very

interesting things happen. So first of all, once there was a Shakespeare team and so these judges asked questions about Shakespeare and the machine and human was supposed to respond and now they were able to bring in one particular human.

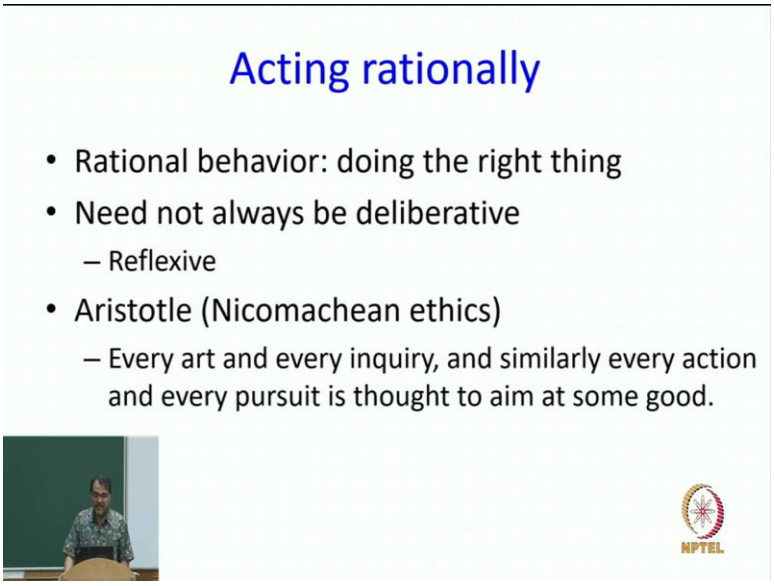
Who knew everything about Shakespeare knew every little detail and all the trivia and random stuff about Shakespeare, all his novels, all his work. And so whatever question this expert would ask, this judge would ask, the human would be able to respond to all of it. So the judges thought that there cannot be any human who would remember so much thing about Shakespeare, this thing has to be a machine. So the machine did not defeat human.

The machine did not succeed in winning Turing Test, the human failed in the Turing Test. The human made people believe that it a machine he is a machine because he remembered all the details which we thought could not be a machine. Now is that the write demonstration that we are looking for. It is said that if you really want to win during tests, make spelling mistakes. Because humans make a lot of spelling mistakes.

If you never make a spelling mistake, judges will feel that, you must be a machine, you are not making any spelling mistakes whatsoever. We do not want to build systems that make human life just so that we can win to the test. That should not be the demonstration of a right. And there are so it is a slippery slope. It says it is very difficult to reproduce. With the specific experts, and with the specific humans, a machine may win with a specific experts or humans it may not.

It is not a constructive test. It is not a mathematically analyzable test. This kind of an evaluation is good for news stories, but it is not necessarily good for making progress in the field. Therefore, it was it has been over time agreed upon by more people that an appropriate definition is to act rationally.

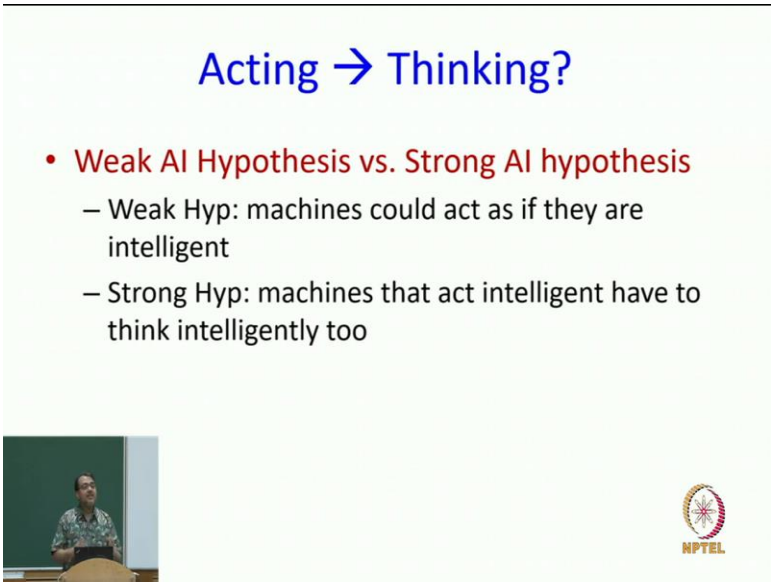
(Refer Slide Time: 19:12)



Now, when you say act rationally well it need not be deliberative, you need not have a logical process at the end, you may have a reflexive process you may have any process that leads to the right action. We are with that it also is comfortable with this idea that every art every inquiry and similarly every action should be in pursuit of something good. And that you will see we will get to when we define what is rationality.

So, now, there is one problem however, with this definition and the definition the problem is we have said that AI is nothing but acting rationally. But now we have a problem what is the problem? What is acting rationally what is rationally what is rational? What is the definition of rationality? We have not solved that particular problem yet

(Refer Slide Time: 19:59)



Before I go there, and the next few minutes will be spent on defining what is rationality? I would say that there is also one particular important question where there are two sides within AI when it is called the week AI hypothesis one is called the strong AI hypothesis. The week AI hypothesis says that machines may not have a good thinking procedure, but they can still act intelligently. It is saying that you can build AI systems.

Which internally are not doing the not doing the right logical processes, not having the right reasons for the right action, but they can still exhibit the right behavior. Whereas strong hypothesis says that machines in order to act intelligent have to think intelligently to. So in other words, thinking rationally is a sub part of acting rationally. Now, people disagree on what is the right answer here. Here is a very famous Chinese room example.

This happens in the context of language, but I think it makes the point still suppose I want to convert English to Chinese. And now we can say that if I am able to translate from English to Chinese, I will be able to claim that I know both languages. However, there is a room and inside the room there is a big fat book. A huge book where somebody has enumerated all possible sentences in English, let us say up to a very large length.

Which is unknown to the expert? And for that somebody has already written down all the Chinese translations and some, some third per

son has access to this book. So whenever a query comes, this is my English sentence translated, the human goes it opens the page which has that particular sentence looks at the Chinese just copies the characters and sends it out. Would you say that such a human which is reading from the book knows both English and Chinese?

We would not say that. But would it be able to fool someone in believing that it he knows what English and Chinese he or she? Yes. So this is sort of like the week AI hypothesis. So it is saying that you can have a system which can act engine intelligently, without thinking intelligently without knowing the thing that is necessary for the intelligence specifically if you have a large set of rules, which somebody wrote for you.

So I remember I said it 40% of AI startups do not really have AI in it. A lot of people think of AI as follows. Somebody, a domain expert looks at some data, looks at past behavior, write down rules. Rules could be if your income if you have transactions we are then 1 lakh rupees, you should file a tax return. And if you do not file tax return you will be in you will be found guilty, a new question will be raised or if you have home in greater collage then you should file tax returns or whatever,

They just write these rules. And then they say it is an intelligent system, it can automatically predict which people might be doing tax fraud would you AI say these handwritten rules to be an AI system? This is equal to this Chinese book example Chinese room example. We would not and a lot of people who say who claim they are doing a startup actually writing these kinds of rules at the back end.

Then you can start to put in some intelligence one step at a time you can say I do not know whether 1 lakh should be the threshold or 10 lakhs should be the threshold of 1 crore should be the threshold So let me have the AI system figure it out. I do not know where the greater collage is the right thing, or GK 2, to Gk 1, we should differentiate or what, let me have the AI system, figure it out.

And if the AI system automatically figures out the parameters or the values that lead us to this particular intelligence, then we can start to say that there is some automated intelligence that is coming out. And so then that is where they start becoming AI systems. So anyway, so the point is that people disagree on whether acting rationally is the right definition. Also, not because they say that you can act rationally without really thinking intelligently, and that is not great. But you know we have to come up with some definition.

So we come up with the acting rationally definition. So now I am out of time, but what I am going to do in the next class, and that will be 15 20 minutes of the next class, is that I will talk about agents and what is rationality? And that would sum up the definition of AI for us, and then you will start talking about our first algorithm. First competition modern. So real technical work starts in the next class. Thank you.