P01

Face-to-Face, Duration: 17:52 minutes

[00:21] Thanks for your participation.

[00:22] Your welcome.

[00:23] My name is Mario Neururer, I am a Master's Student at the MCI in Innsbruck and writing the thesis about social intelligence of autonomous conversational agents. Your participant is much appreciated. Just before we start the interview. I'd like to reassure you that as a participant in this project you have several rights. First, your participation in this interview is entirely voluntary. You are free to refuse to any answer, to answer to any question at any time. You are free to withdraw from the interview at any time. You are free to refuse to the audio recording of this interview. So the interview will be audio recorded and is strictly confidential and will be anonymized, anonymized during transcription. Excerpts of the interview may be part of the final master's thesis, but under no circumstances will you name or identifying characteristics be included. I would be grateful if you would agree verbally to show that I have read it contents.

[01:24] Yes, thank you.

[01:27] Would you like some of the results of the research project at the end.

[01:33] Yes, sound very interesting.

[01:38] So, just to go over some questions. I would like you to give me some demographics, name, age, your residence, highest education and specialization and your current occupation.

[01:49] Mhm, my name is [redacted]. I am thirty years now and I studied computer science at the Johannes Kepler University in Linz. I finished there as a [redacted], Master of Science and I am in the business now for round four to five years. Mostly specialized in big data and data science, with a big focus on machine learning in the last years.

[02:19] Ok. So relating to the final questions, do you have existing knowledge about messaging services?

[02:29] Ahm messaging services in what sense?

[02:33] Like What's app, like short messages.

[02:33] Yeah. From a technical view or from

[02:35] Just from a user's view.

[02:36] Yes.

[02:39] Existing knowledge about artificial intelligence or intelligent systems?

[02:21] Yes.

[02:41] Yes. Ahm, previous encounter or test of messaging platforms or services.

[02:46] Ahm, I'd say yes.

[02:49] Ok, and previous encounters or test of intelligent autonomous interfaces?

[02:55] Do you have an example for that.

[03:00] Ahm, an example would be a messaging agent or an agent.

[03:05] Google now for example?

[03:09] For example.

[03:10] Mhm.

[03:12] So, what's your favorite AI or AI-related movie?

[03:15] Movie, Ex Machina was pretty cool in these recent years. Yeah, I'd say that the one like the most.

[03:32] How would you imagine some kind of a Turing-proof conversation with an agent.

[03:39] You mean, what would I ask it.

[03:42] Yeah. What would you ask, how would want it to behave, how would it answer your questions or queries?

[03:50] Mhm. That's interesting. The thing I think is, as the AI gets better and better, actually I think to really Turing test it is to expect a certain level of uncertainty and a certain level of, or a margin of error. As the Turing up until now was most of: "Can this entity answer my questions at the best most precision". However, that's kind of getting

to be a give-away to be an AI, so to say. So if there is no margin of error, it's either a genius or an AI would be interesting to show hesitation or uncertainty or fault.

[04:34] So. It wouldn't be perfect at all, it would have some, some flaws in it?

[04:21] Yes. At least in the social dimension. So an AI has different dimensions it can be smart, it can be fast it can be intelligent, socially intelligent sorry, and to show proper social intelligence it needs to build trust and always having the precise answer is somewhat off-putting, I'd say. I think that's go to be the give-away to see if you are interacting with an AI or not.

[05:09] How would this trust be defined, or how would this trust be build up?

[05:14] Ahm, I guess, discussion with the machine relating experiences that's, I think the most humanizing factor even if you in an interview like this. If you can share prior experiences, life experiences, not necessarily related to the question itself but to build up a human persona which you can relate to and model for yourself.

[05:42] Okay. Ahm, so, have you ever encountered such a system that seems to be authentic if I may imply that?

[05:52] Not really, they are just not there yet.

[05:55] Okay.

[05:56] I mean there is an example that some, I've forgot the name now, it's a Chabot that was developed by Google which you could chat with, which really drifted up into existentialism and were really weird but very interesting answers. However, it was still obvious that it was a deep insight in the answers. It was just something that it learned.

[06:25] So, looking on authenticity, what would make a bot then be authentic?

[06:35] Hmm, it's a good question. Well, is it, yeah, it's a good question I don't think I have a clear cut answer for that.

[06:55] Okay.

[06:56] Is that basically, first of all building trust and then having a charisma is actually I think, ups, ah important, if you like are chatting with the agent and well giving the precise answers obviously. And having social intelligence, I think that's even more important than giving a precise answer.

[07:29] Okay, ahm, would it also be that an agent would reveal that it is an agent? So it would not say that it is an agent or should it say that it it's an agent?

[07:47] That's highly interesting. I think ah human nature is able to anthropomorphize entities. And I think it probably would good a strategy, ah, would be a good strategy that an agent debouches that it is an agent. So that could be a really interesting dimension. Even Ex Machina for example, I think that showed perfectly how like a person, a testy, a tester is interacting with such an agent with full knowledge that it is an agent. Still, anthropomorphizes the agent.

[08:24] So, it doesn't need to hide that it is an agent?

[08:28] I don't think so.

[08:30] Okay. So …

[08:33] For example Siri, right there, I mean there is stories about like weird Americans tried to marry Siri, treating it as their girlfriends for example so it's and it's obvious. It's you phone you talk to and it still give you a, well, authentic answer. It has a human voice, it has really good synthesizers so to say, and answers for you questions.

[09:05] So, Siri also has, if you ask Siri, it tells you that it comes from Silicon Valley and the Google Headquarters. Is that something would make an agent even more authentic or gain the credibility in the end or your credibility?

[09:22] Ah, that it states that it is an Apple work.

[09:24] Yeah.

[09:24] Yes, I think so.

[09:26] And that it has some kind of roots where it comes from? Is that important too, that it knows, how…

[09:32] That's more funny and more of a charisma-building. Well, if Siri viscerally suggest that it comes from Silicon Valley then obviously that is cute, and it's just a nice little fact that you have in your head, which further stylizes it as own anthropomorphic entity.

[09:50] So you wouldn't say that heritage is important for an agent in that sense.

[09:55] Ahm, I mean it's a nice to have, it's not a must-have, it's a nice to have and yeah its character building and trust-building.

[10:04] Speaking about intelligent agents which are not specific but generally intelligent. Isn't it also important for these agents to have objectives, a mission statement, a reasons why they are there? Is that important for you as an expert?

[10:26] That they have their own goals or that they … ?

[10:20] That they have their goals, their perspective given by the developer or that they create their own goals out of how are aware of the environment.

[10:42] Mhm, that's interesting. Well, the goal of like Siri, for example, is to give you the best, possible support, right? What exactly does that entail, should it, I mean the problem of that is to be more science-fictionee as to state well if it develops its own goals what are they going to be. It's also one of the dangers, for example, how is it called? Nils Bolstrom or Nick Bolstrom, is stating, like what, if it defines its own goals, its utmost goal is to develop or programmed by the developers to make you happy. Then, ah, a good way to achieve that is to implement a electrodes into your brain and just stimulate your serotonin levels indefinitely. So, it needs tight controls. That's actually an open question. And how to, how to really specify what the goal should be for the AI.

[11:55] So, am you are saying that a goal that educates, an agent that educates itself may not be authentic in the end, because it develops different goals than what it is made for. Ahm, is that right?

[12:12] I mean, kind of. Humans develops their own goals and still be authentic. So I don't think it exclusionary, I just think it is more a philosophical problem and how to achieve that.

[12:28] Okay, if you think on Tay, the Microsoft Tweet bot, which …

[12:32] I don't know that one.

[12:33] Okay, this is a bot by Microsoft and they allowed it to learn by Tweets. So they wanted it to become a seventeen-year-old American girl.

[12:46] Okay.

[12:47] And it turned out that she, by reading and learning from Tweets, became an anti-feminist and a Nazi.

[12:53] [laughs]

[12:56] Would that imply that she is not unauthentic, or is it still is authentic in a way because she developed herself?

[13:02] Well that's the nature-nurture argument, same applies to a human child, right? Is xenophobia for example, is it genetic or is it, ah, nurtured by the parents. There are arguments for and against it so that not even clear in for like humans. So, I don't think that's an argument against Tay, it's called?

[13:29] Mhm.

[13:30] So, basically it learns from its clique, from its surroundings, yeah. So its nurtured to be that way. Programming it, ah, probably would be the genetic code that's, yeah. I don't think it impedes authenticity.

[13:59] Okay. Ahm, we talked about …

[00:05] To add on that, interesting question there is does it, like the Tay bot, became an anti-feminist, Nazi, whatever, ah, still authentic because in order, or it wanted to be to fit in into its clique probably, because that's what its learning from and that's how it wants to communicate with its clique. Unauthentic would be if it pretends to be a Nazi and anti-feminist and just in order to achieve a specific goal.

[14:36] Okay. Yeah. That would have been my next question. What would unmask an intelligent agent not being authentic. Okay, I have just one last questions, ahm, talking or staying with authenticity. Ahm, we try with the Turing test to make machines equal to humans in their behavior, right?

[15:00] Mhm.

[15:03] What makes a human authentic?

[15:06] Makes a human authentic? Ahm, what makes a human authentic, that's a good question. That depends on the human who is interacting with that person. Hughley authentic persons can be interpreted as not being authentic by different humans.

[15:35] Okay.

[15:37] It depends also on your approach towards a machine. If you don't accept the authenticity of a person or an entity, then it's hard for it convince you of being authentic or not. How can it convince you, well, that's a charisma. There is, they can fake it or they can be.

[16:00] Okay, great. Is there anything you want to add? Anything that I forgot about asking you when we are talking about authentic conversation agents?

[16:15] Ahm, maybe just a little tip. You mentioned the Turing test a couple of times, what you maybe should also look at is the Chinese room argument.

[16:28] Mhm, okay.

[16:28] This, do you know?

[16:29] Yeah, I know about it.

[16:29] This is also; I think for authenticity especially very interesting argument because there it is just the argument of being a superefficient entity that has no understanding whatsoever. And is that then really authentic or not?

[16:44] Okay. What would you say?

[16:48] I probably not, but the argument is still an open question are we …

[16:54] Are we authentic …

[16:56] in the Chinese room, right? So did we just learn words and are super-efficient. So

[17:04] Okay, good. Thank you very much, do you know anybody who should be considered for these questions as well. I know you have in the Kepler University. They are known for intelligent systems and all that stuff.

[17:18] Unfortunately, there I was not involved with machine learning yet. But maybe Dr. [redacted], he is also taking, he is a quantum-physicist and in recent years, months is taking a look at artificial intelligence, so he might be interesting to talk to.

[17:43] Okay. Thank you very much and I will let you know about the results in the end.

[17:45] Thank you do.

[17:46] Thank you.

P02

Skype, Duration: 22:20 minutes

[07:23] Sounds great.

[07:25] Alex, thanks for your participation. My name is Mario Neururer. I am Master student at the MCI in Innsbruck, writing my thesis on social intelligence of autonomous conversational agents. Thank you for your willingness to take part in this research project. Your participation is much appreciated. Just before we start this interview, I would like to reassure you that as a participant in this project you have several defined rights. First, your participation in this interview is entirely voluntary. You are free to refuse to answer any question at any time. You are free to withdraw from the interview at any time and you are free to refuse to the audio recording of this interview. So I hope it's OK the interview in order to transcribe it later. It will be kept strictly confidential and it will be anonymized during transcription in the end. Excerpts of the interview may be part of the final Master thesis but under no circumstance will you name or identifying characteristics included. So I would be grateful if you would state that I have informed you about your rights.

[08:36] Yeah. I am informed about my rights. Thank you.

[08:38] Ok. Would you like to get the results of this research project in the end?

[08:45] Yes.

[08:48] Good. Ok, let's start. I would like you to just quickly introduce yourself. Your name, age, residence, highest education, specialization and current occupation.

[08:58] Ok. So, my name is [redacted], I did my undergrad in [redacted]. I did my [redacted] and through the double-degree program I came to the [redacted] where I did information systems and now I am in a PhD program for information technology. Yeah, my focus or my areas of interest are artificial intelligence, crowd-simulation, collaboration tools and anything where you can basically develop code.

[09:57] Ok, that sounds good. Relating to our topic, do you any existing knowledge about messaging services like What's app, iMessage, WeChat.

[10:03] Mhm.

[10:04] Yeah, and exisiting knowledge about artificial intelligence or intelligent system. I assume yes.

[10:10] Yeah, I did some projects on that.

[10:13] So, did you all, can you imagine an intelligent operating system in a messaging environment. Is that something that you can think of?

[10:26] I think that's already there.

[10:28] Ok, good. So, to break the ice a little bit I would like you to ask, what's your favourite AI-related or computer science-related movie?

[10:38] Ex machina

[10:40] Ok

[10:45] You have seen it?

[10:46] Yeah, I've seen it. A week ago and I was quite surprised by the …

[10:52] You did your homework.

[10:55] Yeah I did my homework that's true. I was quite surprised by the end. So how would you imagine a Turing-proof conversation with an agent.

[11:05] A what proof? Sorry…

[11:07] A Turing-proof. Alan Turing, you know the experiment the imitation game. So how would you imagine a Turing-proof conversation with an agent.

[11:15] Can you explain the Turing-proof once more.

[11:24] Yeah, talking about the Turing test from Alan Turing. Where two participants are in a room and a third participant whether one of the two others is a machine or if it is a human. The machine wins if it is able to fake being a human to the other person.

[11:58] Ok. Ok, I got it. What's the question again.

[12:00] How would you imagine a Turing-proof conversation with an agent.

[12:06] Yeah it's probably a difficult task in the future, because as long as you do not really see the other side, the participant, you will have to fall back to strategies like asking specific questions which are difficult for machines to answer. For example, you could try to recognize emotions based on like questions: "What did you do last weekend?" and then follow up and see the reaction. But it, like, the more sophisticated system the probably less, or more difficult it will be to confirm human identity.

[13:08] How would you a machine, how you would you expect a machine to answer your questions?

[13:13] I think, the usually the first step what they do is, they transform your words, open words, into a text or digital words which they can then use to, based on algorithms or like that, get the meaning out of it. And of course there's a lot of error. You can dialect, voice a lot factors play a role here. Probably the next step, like, is like, to make a based, a priority list, so ok, the answer could mean one of these things, and the one which is most likely, is basically the basis for a reaction and this reaction could be rule-based. So, if there is a certain content then it could react to it. If you even have information in what context this word or topic is used. For example, the weather is good or bad then you could like go on or follow up on this. And you could also do it non-rules based. You could do for example, you could for example have a, yeah, some AI-algorithms which are more based on like memory for example, like neural networks.

[15:12] Thinking of Ex machina. Do you think that the system you explained could have the same features or same traits as the main character Ava?

[15:31] Mhm.

[15:32] Which one would you say are most important for such an agent?

[15:40] Well, I like the, I would say like there are two things really important. First, a correct interpretation of your opposite and then a proper response. And everything evolves around these things. You also have to add like a timeline during the conversation, so your answer can not only be based on what was previously, like right before. It has to be like a follow up like every single interaction step.

[17:20] Ok. This process, do you think then gain credibility in the end by following the steps you just said, like knowing previous interactions and previous queries and such?

[18:43] Mhm. Definitely. I mean a learn-engine whether it is implemented in the agent itself, so, or if it comes from an external source and like you have other learning platforms, it doesn't really matter. But I think it's absolutely necessary, you cannot build something just based on rules because then you will never, like, comprehend, like a lot of possibilities.

[19:17] Mhm. For you personally, what would an agent need to do to gain the credibility or to establish some kind of trust?

[19:30] I think an agent should be smart enough to feature personal information of the conversation partner. So, if the agent is aware of some like, user dependent traits or if he's aware of like, for example when the agent would be built into Facebook, and it would know a lot of the personal history of the user and it could give much more proper feedback.

[20:09] Would you say that this would make the agent authentic?

[20:16] Probably not authentic at this point, at this stand of technology but it, like I mean. In one project we had a feedback agent. One of the results was that it actually doesn't really matter that the people now if it is a fake agent or not it would still trust personal information more to the artificial agent than to a human. This experiment was in the context of like giving medical information.

[21:02] Some kind of ELIZA thing?

[21:05] Yeah, like people felt more confident interacting with a bot rather than a real human, telling them about personal, like medical things…

[21:17] Even though they didn't know that it was a machine?

[21:21] Actually, they did know it. They did know it, there was another experiment where they at the end, like it was the same series of experiments, at the end they had to specify whether it was a human or a computer and about sixty, two-third of the people got like it right. It was no, one third got it wrong so, that was maybe a good sign that like it was authentic.

[21:59] Ok. Are there any other traits or behaviors that would make a bot authentic? I mean, can you think of something that would make it authentic if you just imagine you have personal assistant, a textual agent that is on your phone, messaging you, on your computer, on your IoT, on your mirrors. Can you imagine that, what does it need to do to be authentic to you then?

[21:33] I think the most important thing to do is to first really specify what you mean by authentic. Is it a goal to make it most similar to like a human, is it a goal to make it, to do its purpose at most quality? This can be two completely different things. And I think when you design for example a feedback agent you should always go for the second goal first. So it has to fulfill its purpose, what it is designed for, rather than like sound authentic in the sense of sound like a human.

[21:19] So, it has to have some kind of a mission statement and objectives, is that right?

[22:24] Right, I mean, a lot of like agents which are used in conversations, I think they are not really designed to be in that like, to be differentiable form a human or a computer. They are more designed like to provide users with specific functions or help or assist them as you said before. And I think that's really the research direction of like most of the papers in this area.

[23:00] So, saying that authenticity is something that is individual to a person. What would mean not being authentic?

[23:20] Yeah, I mean if there are concerns about whether the agent can really help in what it is designed for than this damages authentication or the quality and this could range from missing points, when it is giving wrong responses, when its giving incomplete responses or when it, even when like in the worst case it's for example a decision support agent gives contrary results, that are not expected.

[24:17] Okay. I think we are already at the end of the interview. Can you think of something important I forgot out of the context of building an authentic messaging agent? Or do you think that we covered mostly that relates to that?

[24:42] Yeah, I think, I think what you can, like, when you design such an agent, what, like the good thing about such agents is that you can build them in a modular way, so you can have like a base engine, this is the agent, then you can gradually add things to improve the quality. And there are already platforms where you can test such things for example Slack is one, or in instant messaging there are a lot of IRC channels nowadays have this features. And that you can use to experiment that.

[24:28] Like there is the one from Microsoft you probably heart that Tay bot.

[24:36] I have not experienced it.

[24:37] Ok. It was a bot that was made to be a nineteen-year-old girl and it was allowed to train itself, to educate itself and it became a feminist and a Nazi, an anti-feminist and a Nazi. So just by learning stuff that nineteen-year-olds see on Twitter.

[25:00] Really [laughs]. But I think that's a, you know, that's a chance in one in like, you know there's so many opportunities, it's like butterfly effect you cannot really determine the result by such a high probabilistic experiment.

[25:29] Sure, it's quite a narrow example. It is. Ok, so do you know as the last question I go to ask you, anybody who you think should also be considered as a interview partner for this research.

[25:44] Ahm I have two professors at [redacted], which are working right in this area. I can give you the contact.

[25:53] That would be awesome, because I am trying to get 12 experts in the field and after I have the first ones, I go to go further by snowballing.

[26:06] If it interests you could also, I don't know if you have read through it or not, like in my Master thesis I worked on like feedback-agent in a business simulation. It was not really, I mean I made it rule-based, so it's not really an artificial intelligence at this point. It is more like a place-holder, like the concept was there, and you could replace it with a real AI, maybe this gives some insights to you.

[26:44] Yeah. That sounds good. Ok. Then thank you very much for your help. I will let you know about the results in the end. Thank you.

[26:49] No problem. Ok.

P03

Skype, Duration: 14:29 minutes

[00:00] So, I go to read you a short intro and the we can have a talk. Ok? Thanks for your time. Hi, my name is Mario Neururer, I am a Master's Student at the MCI in Innsbruck, writing my thesis on social intelligence of autonomous conversational agents. Thank you to take part in this research project, it is much appreciated. Just before we start the interview, I would like to reassure you that as a participant in this project you have several definite rights. First, your participation in this interview is entirely voluntary. You are free to refuse to any answer any question at any time. You are free to withdraw from the interview at any time. You are free to refuse to the audio recording of this interview. So the interview will be recorded but will be kept strictly confidential and will be anonymized during transcription. Excerpts of the interview may be part of the final master's thesis, but under no circumstances will you name or identifying characteristics be included. I would be grateful if you would agree verbally to show that I have informed you about your rights.

[01:09] I am informed.

[01:10] Would you like to get the results of this research sent to you.

[01:18] Of course.

[01:20] Ok, so I would just like to start you telling me your name, age, where you from and maybe your education, occupation and specialization.

[00:36] I am [redacted], I am from [redacted], I have a PhD in computer science.

[00:39] Do you any relation to messaging bots or messaging platforms, so existing knowledge about messaging services?

[00:50] I use them.

[00:51] And you have knowledge about artificial intelligence or intelligent systems I suppose?

[00:55] A little.

[00:56] A little bit. Can you imagine an intelligent operating, an intelligent agent operating in a messaging environment?

[00:04] Yes.

[00:06] Just to break the ice, what's your favorite AI movie.

[00:10] AI movie? Terminator.

[00:00] Terminator? Ok. Did you hear about Tay the Microsoft bot …

[00:18] Yes.

[00:21] So, what do you think, is this agent authentic or unauthentic?

[00:26] I don't know what that means?

[00:31] So, does it have a true inner self, is it sincere.

[00:40] It's a chat bot.

[00:43] Yeah, it's a Chabot, do you perceive it to be …

[00:46] No chabot has inner self.

[00:50] Ok. Do you think it works according to a mission statement it was given once?

[00:57] I don't know what that means either.

[00:59] So does it have clear objectives that it is working towards.

[03:00] Well, they designed it to chat with users. So I guess that's the mission. Did that.

[03:14] If you would talk to one of these chat bots. How would feel about it being authentic to you. What does it need to have to be authentic.

[04:20] By authentic you mean human-like? That's what you are saying.

[04:25] Yeah. I would say it would be human-like.

[04:30] Well it has to be able to keep the conversation meaningful. It cannot get confused with the language.

[04:27] So it has to handle the language, are there any other traits it should have.

[04:44] Basically it uses the Turing test, if it can pretend to be human well-enough than it is good enough.

[04:51] What do you think it needs to have to pass the Turing test, to be a 100 percent perfect conversation.

[05:00] A pretty good knowledge base about what it's like to be a human.

[05:05] Ok. What does that include?

[05:06] Everything.

[05:08] Can you name some of these everythings?

[05:11] Sure, it has to understand human physiology, it has to understand history, it has to understand culture, I can't name something that would not be important.

[05:25] How do you think that the agent could gain this credibility of being human-like. How would that process like?

[05:33] How to gain the knowledge to pass the test? …

[05:36] Yeah, and how to create a feeling, or your feeling, your perception that it is human-like.

[05:44] If it has this action ability it doesn't need to explicitly work on creating feelings. It just performs at that level.

[05:53] Ok.

[05:54] Now, how to get that knowledge is a difficult question.

[05:58] Ok. How would you think, how to get this knowledge?

[06:02] It seems like a lot of it is on the Internet, so probably by data mining Internet would be a good starting point.

[06:08] What kind of data do you deem as most important to become human like.

[06:15] As I said, all of it. There is no unimportant data. If there is something human beings now and the bots doesn't know, that would be a problem.

[06:23] Ok. Would that imply that a bot that doesn't anything is unauthentic?

[06:31] It's less authentic than the one who does. The degrees of how long it can pretend to be human before it is discovered and obviously if it has significant holes in its knowledge base it would fail sooner.

[06:44] Ok. When you think of an agent that is your personal assistant, how would you want it to be, how would you want it to behave towards you and how would you want to it to inform, how would it communicate, interact, how would it behave?

[07:09] It would know what I am expecting and not waste my time with pointless introductions and so on. It would get to the point.

[07:17] Okay, so it would know work and …

[07:21] It knows what I already know and just fill ins the blanks, it doesn't do kind of every humans needs to hear this disclaimer before I start an interview. It understands quickly what happens.

[07:33] Okay, would it also reflect your culture then?

[07:41] It's part of who I be, absolutely.

[07:50] Okay. How would you wanted to get to know all this information? Would you train it or should it …

[08:00] Ideally it approves it just look at my emails, look at my online history, videos and learn from that.

[08:04] Okay. Good, what do you think of authenticity in people? How do you think that, authenticity of people, how is that created in the end? The same…

[08:16] I don't know what that means. You keep using this term as if it's well defined and every one should know what it means, I have no clue what that means, I think people are authentically people by definition. Unless you have different definition for that term.

[08:32] No I don't have a different definition. How would you define it?

[08:34] Every person is a person just by verge of being one, nobody has to pretend to be a human.

[08:41] Ok. So also a machine is authentic if it just does what it does?

[08:47] Machine is human-like if it performs at the same level.

[08:54] But a machine is authentic even if it's not human-like by just doing what it does.

[09:00] There again I don't know what that term means. You ask me what is c, and I have no idea what c is. If you define it to me, I'll tell you whatever it applies to machines,

humans or squirrel. But until you don't tell me what you mean by that, it's not an efficient way to communicate.

[09:25] Okay. Literature says that authenticity is being genuine, trustworthy, sincere, it focuses on communalities, on context, objects and mission.

[09:30] Okay, so that's something. So you asking me humans are trustworthy?

[09:38] In that kind, yes …

[09:40] No they are not. I work in cybersecurity I don't trust anyone.

[09:45] Yeah, okay. But do machines need to fulfill all these traits in order to be authentic.

[09:52] Well by definition, yes.

[09:56] Yeah, if we take the definition from brands and philosophy, because these definitions come from all this different fields.

[10:03] Right.

[10:05] Do we think in the computer science are that these are the same ones for machines as well?

[10:13] Well, in computer science we have specific things we want to accomplish, for example can a system do X.

[10:20] Yeah.

[10:21] So if it can then it means that ability. But do I care if it feels authentic. I don't even know why I measure it.

[10:29] Okay, it wouldn't make any sense to make the machine socially intelligent at all?

[10:38] Well it has to be, to pass the during test, right. That's part of being human, we are not, most of us are not autistic.

[10:47] Okay, do you think that authenticity doesn't play in there.

[10:58] Do I think if machine is not trustworthy it cannot pretend to be a human?

[11:03] Yeah.

[11:04] Well, it can pretend to be a dishonest human I guess, which is not that rare.

[11:10] So, it is also not necessary that the machine has these abilities. It just needs to pretend it, right? I mean that's the definition of the Turing test.

[11:23] Right, absolutely, it's not the internal state of the system.

[11:29] Okay, so it doesn't need to have consciousness as an internal state, so?

[11:36] I don't have consciousness as an internal state. I am a zombie.

[11:40] I hope that's not true.

[11:42] How would you test? Can you test for it?

[11:46] Not yet.

[11:47] Do you think there is ever a test for it?

[11:50] That's a good question. I don't know.

[11:53] What do you think?

[12:00] I heard something about that zombie test once but I didn't read it through.

[12:06] I can tell you there is not and probably never going to be a test for your internal states.

[12:08] Okay, yeah. Good, I really liked this conversation. However, I think through with my questions already. Do you think of anything that I could have missed in talking about social intelligence, authenticity of machines?

[12:33] So there is a lot of interesting research on teaching machines to lie. That be an interesting aspect and ability and humans are really good at that. Something to analyze, it's interesting to talk about impact of the machines and with the Tay example, how it becoming racist and all that impact people. So there is very interesting directions to take. I am sorry if I am very direct.

[12:51] No that's fine.

[12:53] A lot of people don't appreciate this type of conversation I hate on it. I think it helps to have common definitions then the conversation becomes very useful. If we don't agree on terms, then things I say and things you ask don't match.

[13:06] Sure, that's totally fine, I totally agree with that, I totally agree with that. Do you think of anybody who could be of use for this study as well to give insights? I mean you really helped me in defining, still defining what authenticity could mean for machines. Do you think of anybody who could add to that as well?

[13:30] So, it sounds like it may be someone in psychology or something other than engineering just because we are not studying exactly the same aspects of technology. So we are more concerned about ability and efficiency and measurable terms. Authenticity seems like perception of something so it would probably differ for different people. My parents, I am sure would be fooled by the simplest of chat bots, while as for me it's not quite a problem most of the time.

[14:00] Well, thank you for your time. I appreciate it. If you have further information just let me know and I will let you know about the results of the study.

[14:10] Sure, and if you want to quote me on anything or make it public, feel free to do so. I have no concerns about this at all.

[14:25] Ok. Great. Thank you. Bye

P04

Skype, Duration: 29:58 minutes

[00:18] Hi.

[00:19] Hello.

[00:20] Hi.

[00:26] Perfect that we could arrange to have this talk today.

[00:33] Yah.

[00:35] Okay, I would quickly introduce the topic again and then I'd have some questions and then I would just go through. What I do is, I am looking to for a kind of model to create authentic intelligent agents. That's why I am looking into social intelligence and try to incorporate that to artificial intelligence.

[01:02] Mhm.

[01:03] As I've seen you are working a field as well that is related to that. I've watched the clip of FURHAT.

[01:13] Yah. Which one?

[01:16] The one on your WordPress blog or on your page.

[01:22] Yah.

[01:24] I have one or two questions to that as well. If you allow at the end. So, I just have to say that the interview is voluntary and you can refuse to any question at any time. You can withdraw from the interview and I hope you are okay that I record the interview in order to transcribe it later. There will be no, everything will be confidential, strictly and anonymized during transcription. There will be no identifying characteristics in the end.

[01:55] Okay.

[01:56] I would send you the results in the end if you are interested.

[02:00] Mhm. Absolutely

[02:04] I would just ask you some starting question, whether you have knowledge about messaging services such as What's app or iMessage.

[02:20] Yah, sure I use that.

[02:21] I propose, or I think you have knowledge about artificial intelligence and intelligent systems, right?

[02:28] Yah.

[02:30] Could you imagine an intelligent agent like a bot in a messaging service?

[02:37] Yah, I mean I guess, there are already these kind of bots now. So, yeah, absolutely.

[02:45] Normally, I ask about what's your favorite AI-related movie, but most of the people say Ex Machina. What's yours?

[02:56] You mean about ah …

[02:58] Artificial Intelligence or intelligent systems.

[02:59] Yah, I don't know. I guess 2001, I would say.

[03:08] Oh really, with HAL.

[03:09] But, yeah, I liked Ex Machina also.

[03:12] Okay.

[03:14] But I think 2001 is the best.

[03:21] Okay. Do you think that HAL:9000 is socially intelligent?

[03:28] I think first, the term social intelligent is a bit ambiguous. It's a nice concept but it is quite problematic also scientifically, what you mean by social. And people use this term social intelligence and social signal processing and so on, but in a way I mean all language use is social. Because it's about the communication between people. That's sort of the essence of something social, is that we interact with other people. But then I guess what people mean by social then in this context is sort of not directly task oriented. Which means that you are not just using, interacting with that entity to perform a task but there is also some value to the interaction in itself. So that's how I would understand social. Sort of something beyond the pure task-oriented. And, I think in the movie they do have discussions that are just not task-oriented. They do have discussions about how they feel about things and so on, so in that sense it's social. But of course it's not, it doesn't have

very emotional speech and so on. But there are people that don't have that also, so I don't think that is a requirement for being social.

[04:56] Do you think therefore it is authentic?

[05:00] The movie, or …?

[05:04] No the machine?

[05:06] Authentic, that's also kind of a vague concept what you mean by authentic. Or, do you have a definition of …?

[05:17] I have a little definition. It would mean that is has values, it would mean that it has objectives, a mission statement, that it knows about context or having a heritage or something like that, being genuine, trustful, sincere. So that's related to authenticity.

[05:40] Okay, actually I haven't seen that concept a lot, but, so it's a bit hard to say. Or do you mean like it's coherent. Is that sort of what you mean by authentic? Like things its one personality.

[06:04] Yeah, it goes into, it could be coherent, its more about creating a trust into people. Like you did with FURHAT, I think. When people take the machine as something that helps them or that interacts with them. Creating trust, focusing on the communalities, being nice to each other and having a context in that sense, knowing about the people and their stories.

[06:43] Yeah, maybe that's what I mean with social then a little bit, that you would sort of relate to that agent, somehow. And so in that sense, yes, I guess HAL is authentic. I mean he does definitely have a personality and he wants to know about the people. But then they end up having conflicting goals. But that can happen with people also.

[07:10] Do you think that is important for machines to have that?

[07:15] I guess so, I mean if you want to have a social interaction, there has to be some kind of a personality or that is coherent but also sort of predictable in some sense. And that is what the coherence creates, I guess. And, yeah I think you have to feel that's why the interaction can be worthwhile having even you are not fulfil some external goal. Because you are building a relationship together. That is meaningful to people I guess and if you can build that kind of relationship, long-term relationship with an artificial agent, I think that's possible.

[07:04] What would you need for that? Or what would you built into agent, in order to make, to build this long time relationship?

[07:15] So I think, one obvious thing then is long-term memory. Which these agents typically don't have, so like Siri for example doesn't remember our previous interactions. It might remember who my mother is and so, but very sort of shallow things and the same goes with chat bots of course. You log out and you log in again and even in the same direction there is little that sticks. And also I think, so I think for this to happen, it needs long term memory. And I think one issue that these agents don't have is, I think there has to be something at stake. So in a social relationship something is at stake. And that's what makes it exiting. You can't just, to a chat bot you can say you are an idiot or something to it and it will yes say something then continue. You could not do that with another person, you would destroy your relationship. And I think this thing should also be present in artificial agents, that you have sort of be careful about your relationship. In order for make the relationship exciting or interesting I think. So that's another thing.

[08:32] Is that something that relates to trust, or being sincere?

[08:35] In a way, yeah you built trust together. And that also helps, not only trust in itself but it makes the interaction meaningful again. Absolutely.

[08:53] I don't know if you integrate some kind of a culture or values for FURHAT, when it makes this game with the people in front. Would that be something that is also considerable, like knowing about the culture of the people that are sitting in front so it can better propose the games or introduce the games?

[09:25] I don't know, maybe culture, I haven't, I guess or individual differences in general would be important. I mean one thing I am looking right now is differences in how men and women interact with a system and how children and adults interact. And especially if you have combinations, so you have one adult and one child. And this of course effects the interaction and children and adults don't interact the same way. When we had this system in the museum, it didn't make the distinction. We actually first tried to make a distinction between adults and children by just looking at their length. Because you know the high of the head by the Kinect. But that doesn't work, because the kids didn't sit down, they stood up and jumped and bend over and it sort of was impossible to use that as a measure. And we were not able to ask them also, we couldn't use that. But I think that taking this individual differences into account would be important and maybe culture is

something then that is, that could be taken into account. Although, I think there are a lot of individual differences within a culture, so the question is whether that is a meaningful sort of aggregation or not, for… I guess it's better to see is this an introvert or extrovert like these distributions might be different in different cultures and so on. But I think these personal traits might be more important and age and so on, then culture perhaps. I mean important, of course, if you take, of course language is different. So that's an important culture difference. But it may be things like you know in Japanese do more backslaps, so to create a more believable behavior. That might be an important issue also.

[12:55] What would you think, would be completely unauthentic, even if you compare it to people, what makes people completely unauthentic? Because, I suppose that would could be related to machines as well then.

[13:10] Yeah. Unauthentic to me in that sense, I guess that it would be if people don't, if they react in an unpredictable way or if they are starting off and you expect some kind of a relationship and suddenly they do something that is not expected from that kind of relationship you have. And of course some people do that and I guess that's not a good thing. So to me it seems like this, if I understand the concept authentic correctly, it's some kind of coherence and some social bonding.

[14:02] Right, yeah. You can it sum up like this yes. I think we talked a lot about authenticity, is there anything else that comes to your mind when thinking of authentic messaging bots?

[14:28] So authentic I mean; it could also mean you have this distinction it should be human-like or not. Authentic could mean in that sense human-like, that you relate to it in the same way you relate to a human being. But I guess, that sort of a design-choice. Should I picture this agent as being human like, but you could also imagine it being more like a pet and then you have another kind of relationship? Then it might be more ok, to give commands and so on. And not being so polite. Like you would interact with your dog is quite different than how you would interact with a service agent at… So, I guess you could play around with different metaphors for this interaction and maybe people will like different metaphors for different robots. Maybe the lawn-mowing robot could be more of a pet, and the one serving your food might be more of a person. I don't know, I guess it depends on what people would like form this.

[15:53] That's interesting, how did people perceive FURHAT as a human or as a machine.

[15:55] Yeah, that's interesting because one thing we wanted to in this, that made it quite different from many other systems, was that we didn't, we wanted FURHAT to have the same role, as the persons playing the games. So actually, he was not a tutor. He was not sitting on the answer to help to solve this problem. But we gave him intentionally a believe model that was sort of a distortion of the truth. So that he was actually wrong sometimes, he was partially correct but he could also be wrong. And he also could be more or less confident in his believe. And this was reflected in his arguments. So he could say, I strongly believe that this animal is faster than that animal. Or he could say I have no idea whether which one is fastest. And, we set this up at the beginning of this game to make all this statements coherent again. So he wouldn't first randomly say this one is faster and then randomly this other thing is probably faster. But it should of course be coherent. And then also a memory of which argument he had put forth, maybe had argued for a card at a certain position and he noticed that this didn't happen. So he would argue for it again, but then give up on it. And then at the end of the game, he remembered what he had said, so if it turned out it was wrong, he sort of acknowledged the fact and said that: Ok, I've actually thought this one was faster, but I was wrong. Or, the other way round: I told you this one was faster, why didn't you listen to me. If they haven't conformed to his argument. And, this was, I think it was very interesting, to do it this way. Because to it me it seemed to create more like a game element to the game. And because you would know that these persons know the answer and just try to get as much input form it as possible. But it was a bit problematic for people, it turned out, to understand that the robot could be wrong. Because, so some people reacted, ah you fooled us or like he was sort of intentionally fooling us, because of he should know these things. This is a robot; he should not be ignorant. I guess they started to realize, it was hard to convey this sort of, these conditions in the beginning of the game, although it said so. Like this was hard for them to understand.

[18:54] People don't understand the machine is not perfect?

[19:00] Yeah, exactly, especially for this kind of basic knowledge. Like, which animal is faster. Which is typically just have to look at Wikipedia or something. This is the kind of knowledge the system should have, so then you expect that. So that was a bit hard to convey. But once people understood they thought it was a bit of fun then an in the second round they started to say: No, we don't believe him. And so on. And that was a bit interesting, I think. And again I think, it sort of makes, it was trying to maybe make the

agent more authentic in some sense. Or, you would try to say people can relate to the agent in this similar as to a human.

[19:50] That's interesting because I talked to some experts before and some of them said, well it could also that the agent has to make errors in order to be authentic.

[20:06] Yeah, exactly. That's …

[20:08] That's human-like, right.

[20:10] Exactly, and that's also likely in the future, now we set up this simple tasks, if we are to engage in more complex tasks and problem solving in the future it's not the case that the system will always be right. It will have a certain confidence in its believe and you would have to so, let's say search and rescue robots it might search and it might have an idea about maybe it would be best to go to that room, but it sort of comes with a certain uncertainty, but then you would actually have to negotiate with this robot. Should we go into that room or this room. And the robots maybe the robot has to be able to form arguments for why it's better to go in this room first. And this kind of negotiation that comes from uncertainty and conflicting motives will be important I guess. So it's good to start thinking about how to build these systems.

[21:06] So, it's some kind of opening the black box of these systems.

[21:09] Yeah. I think so, and be able in natural language to explain the workings and even workings of the systems, so we can judge whether their believe for model is well founded or not.

[00:25] How do you think all this relates to the Turing test?

[00:28] Yeah, the Turing test is interesting. But it's of course, it's never clear whether it assumes that people know it's an agent beforehand or not. I mean it's a big difference if you say interact with this person, ok interact. And then you say, actually it might have been a machine, what do you think? That's an easier test to pass then, ok this is either a machine or a human, try to figure out which one it is. Because it creates very different kinds of interactions. So in that case you would ask a lot of common sense questions and try to… But that's not actually the interaction you would normally have with a human or … So, I think the first test in some sense is more interesting. And that you would interact first and then sort of. That would be a more valuable, plausible thing to try to do. And I don't know exactly how these competitions and so on are set up. And of course

there is another aspect and that's that these the Turing test traditionally only focus on the verbal interaction and coming up with a next response to an input and I think that a lot of the things that happen for example in the system, that it's a lot of non-verbal things are happening, like when should it speak and how should it speak and where should it look and so on. Also are important are intelligent behaviors that we also need to model and these are sort of. I guess I read the paper and he says that we could image a scenario where there would be a robot that would you sort of talk to, to have all these other behaviors also.

[23:38] Yeah I think so too, I read it too and I think he mentions that. Then it would be another level. Just a question out of interest. Does FURHAT have this empathy traits, where it knows where to look at and close his eyes and all that stuff.

[23:57] Yah, we have implemented some of this stuff, but it's a lot of more things to do. So it has a sort of a model for where it should look for example, that is related to both when objects move and when people speak and so also related to turn taking, there are a lot of these basic rules that humans found in human-to-human communication that we basically implemented to make it look right. But then it's of course, you can't make up rules for all this it's much more complicated. So we are now looking at different machine learning methods for doing these things more, in a correctly. For example, turn-taking, when should it respond and not respond. It's an interesting question when should it give a backslap and when not. But we also saw, at the museum we were able to do controlled experiments because we could systematically vary FURHAT's behavior, when for example taking the turn. And so, since we had an online speech recognizer, every time some said something it took a little time to process that and come up with the next answer. Typically, if you want to respond you signal, even humans if I want to say something I have to signal that I am about to respond. And there are different ways for doing that. So we could use a filled pause, aaaaah, I can take a sharp in-breath like, and you would heart that I am now about to speak or people use gaze and so on. So we can systematically vary this cues and see they have an effect and we could then measure. So after a person said something looking at FURHAT, which is they probably asked a question or something, we wanted to see, given this delay in the response, how likely is it that this person continues to speak, because like repeat the question or something because I didn't get a response, so I didn't know if you heard the question. And that we wanted to avoid, we want them to stay silent until FURHAT responds. By giving this kind of cue, so we could

easily measure the effects of this cues. They had a very strong effect, so it seems like these kind of more human-like cues are something that are picked up by humans even though they are used by robot.

[26:47] Right, so how does research go on then with FURHAT, building up more of the cues?

[26:54] Yeah, more systematically studying these things and use more machine learning techniques and especially we are looking now into this join-attention-problem. So how do we know what the focus of attention is. What are we, which, if we have objects in front of us, how do we know which object are we talking about, and where should FURHAT look in order to signal that he also is talking about the same object and so on. So these kind of questions and what so more one issue is also this turn-taking when it comes to overlap. Because we saw that in this kind of dialogues where you have a lot of discussion and fun and so on. There is a lot of overlap and typically or traditionally, dialogue systems have always tried to avoid overlap. Because they typically a signal of something went wrong. So we had a sort of misunderstanding of who is to talk next so we had an overlap in speech. But, that doesn't have to be case, so it's quite often people do overlap because they just have fun and they collaborate that way and how can we make a robot that overlaps in the right, I mean there can be bad overlap and there can be good overlap. I mean how do we know, which overlap is good and which is bad.

[28:26] Okay. Great, thank you so far for your time, I think we already stressed out this 30 minutes.

[28:33] Yeah sure.

[28:34] If you have any other comments, I am happy if you have any other thoughts, I am really happy if you share them. Do you know anybody who should be considered for this research?

[28:45] Yeah, I don't know, maybe you have, is it sort of social robots that is the main target or is it more like artificial agents in general.

[28:59] It's more like artificial agents in general. Focusing on conversational agents, not particular on social agents.

[29:00] I don’t know, [redacted] at Microsoft, I don’t know if you have talked to him or if you have him on your list. He does a lot similar research and yeah. He would probably be a good candidate I’d say.

[29:34] Thanks’ again, I really appreciate you had the time to talk to me. I hope it helps you as well. I will send you my results and then I will keep on watching FURHAT for sure. That’s a really impressive thing.

[29:46] Okay, sure. Great. Same to you.

[29:50] Thank you very much and have a great time.

P05

Skype, Duration: 25:45 minutes

[00:05] Hey Mario, how are you?

[00:08] Hey [redacted], very good, how about you?

[00:10] I am good.

[00:15] Great that we could facilitate the talk today, I now you are very busy and therefore I appreciate your time.

[00:25] No problem.

[00:28] I just read your paper again and I think it's really impressive what you found out also about the characteristics of this social bots and I think you read my invitational email, I go in some kind of the same direction. However, I am looking on the dimension of social intelligence which is made up of dimensions like context-awareness, having a presence as well as authenticity. I don't know if you are familiar with that concept.

[01:02] I also checked your Medium page, so you also like wrote a little description of the project. So this is for like conversation agents like the mobile phone companies provides.

[01:19] Right, it's like Magic or GoButler or all those customer service agents that facilitate brand to customer conversations, yes. I hope it's ok if we record this session.

[01:35] Yes, no problem.

[01:37] I will anonymize it in the and no identifying characteristics will be in there. So, you can be free. You can refuse to answer any question and you can withdraw from the interview at any time. Just to let you know that. We can start right now, I would be happy if you can introduce yourself, your name, what you are doing, your specialization and maybe you also some kind of find a relation of our two research areas.

[02:12] So, my name [redacted] and I am a PhD candidate at the University of Indiana, I recently worked on detecting persuasion campaigns on social media, which includes like social bot detection that you are familiar with the paper. And then campaign detection

on social political systems or censorship in social media. So those are like the general research interest of mine during the PhD. And I am continuing developing a bot detection system. We are recently working on another paper to look for bot-human-interaction more deeper. So those are kind of the things that I been doing.

[02:59] I have read in the first line of your abstract, you are talking about the Turing test as well. So my first question would be: How do you imagine some kind of a Turing proof conversation with an agent?

[03:12] It is a difficult one usually with the conversation agents because there is been system been developed years ago. They kind of learn of the previous conversation, try to make a pattern and matching like the type of questions and type of answers that you get. So, how much creativity we can expect from those systems, I don't really sure about that one. But they are good for kind of, if these systems are designed more common questions, I think in your example is to use in university, so probably for a teaching assistant students, they keep getting the same questions over and over again. So in that kind of situation maybe it would be kind of useful, like learning-teaching-class topics covered in the class or topics the students are responsible for the exam. So in that situation these will be helping saving human time.

[04:21] Do you think that, you talked about creativity which is something like being genuine, right?

[04:32] Something like what?

[04:33] Being genuine. How to translate that, I just write it in the … and then.

[04:52] Oh, I see.

[04:55] It's something like that which could be a part of authenticity, right? Like being creative, what do you see as other traits that would make an agent genuine or more authentic?

[05:12] I think like Siri and those other Cortana types of systems try to surprise the users. So they are not always trying to show, when you ask the questions, they also give some kind of surprise factor to look more like a human than like a search engine with a speech synthesizer. So that things might be helpful I guess.

[05:40] They also talk about, and that's relating to your topic, trust when they talk about authenticity. Do you think there is some kind of a process on how to create trust in such a machine?

[05:53] I think there is like these Uncanny-Valley, when you first introduce with this system. Uncanny-valley is like some time to get people to adopt and also how much, there is also some interesting things that I listed recently in a keynote. So how to design those systems that are not biased by the type of data that you trained. So removing bias on your training data, if you train your system with lots of male specific conversation when a female starts talking with the system it doesn't feel like the normal conversation with the other people but is more biased towards the gender or towards some type of other personal factors. So I think that might be also some interesting research dimension to design in.

[06:50] This some kind of relates to context, right, of the conversation? How does that play in the conversation, knowing the context and getting all the surroundings?

[07:02] I think in the university type of things, in your application domain, maybe it might be not that different I guess. So, I am not sure like some of the projects that might benefit from the social media data. We had another paper when I worked on, while I was in Microsoft research for an internship. So in that one we tried to look at the type of questions people look ask in search engines. And then, we kind of try to understand what people are wondering about the world. What type of situation they found themselves and the look for answers. And then we go to social media and look at the people having the same situations and depending on your actions what are the expected outcomes. So, if you think that like a conversation system it might also be learning from the search and social media system and try to understand what possible outcomes for is going to be in the future. So kind of modeling your future outcomes and then designing the answers depending on those outcomes.

[08:30] That sounds interesting. We talked about Siri and Siri is some kind of a interesting example that came up in nearly every interview or talk I had. Siri some kind of tells about her heritage because if you ask Siri where do you come from, she says from Menlo Park in Palo Alto, right? How do you think that influences your perception of the bot when we also relate it to like social intelligence? Do you think that plays into that?

[08:58] I am not sure if I understood correctly but, so you are asking that there are some kind of questions that are rule-based so every time similar questions you got those similar answers. This is programmed to answer in that way?

[09:16] Yeah, and I am asking, like knowing where you come from and who build you is some kind of knowing about your heritage, right?

[09:31] Yes, but I don't think like those systems are really aware of these answers. It is probably some kind of rule, where you ask where you come from, there is a template of answering those questions and then they answer in a given way. They are not like really open domain continuously learning, I mean they probably continuously learn after answering to carry on the conversation, but some of the questions are probably designed to be answered in a specific way. Otherwise it can be like a Tay, T A Y, the Microsoft Twitter chat bot. So they try to make it like an open system by conversation it learns and adopts the behavior and it easily gets abused. So people ask weird questions and it is to answer in either way. So that's why probably there are some black-listed questions, that these systems try not to answer and there are some template questions, it's always answering in the same way.

[10:38] Would that mean that they have some kind of a value-rule-base, where they say, okay these values, I can follow these values and others I don't want to follow?

[10:49] Yes, I think so. I am not sure how they actually design those system, but if I am actually going to produce one, there would be probably a set of black-listed questions, that bots try not to answers and there are ones based on a template. So having produced by Apple in Silicon Valley doesn't affect any other answers of the future questions, if they change the answers for one question it doesn't probably affect the entire system. So I don't know how much heritage those bots are aware of.

[11:34] If you would build an authentic conversational agent, which traits would you give him?

[11:46] Hm, probably one thing is like the most important thing I guess, is try to understand the other people's intentions when you are making a conversation instead of just looking at the question. So maybe the emotions will be helpful to get like sarcasm, more other types of things. So not like only the textual based maybe extracting other features like the speech, emotions and like the trend of the conversation. So are you in a fast pace, or slower pace. Then the authentic agents can also like follow in the same trend.

There are like some sociologist have some theories about the communication, so how peer-communication can be most effective and it is most effective when two of the people like try to mimic each other's behaviors, facial expressions. So probably they also might try to mimic not only like trying to answer the questions but also mimicking the behavior, speed, face.

[13:03] Yeah, this is interesting that you mention that, because also literature says that this is one part of authenticity, they call it meta-verbal cues, which is also something like saying, ah, or inhaling before you start the conversation and stuff like that. Okay. When I asked you about how you would build your own bot, this also includes like the questions what do you think makes a person authentic? Are there any other traits you would say, a person has if the person is authentic?

[13:49] I think most of the social bots that we have seen, they are not trying to be authentic by themselves, they true to optimize something. Like if you try to sell your product on media you look like to be an expert in the field, so if you are to sell a pharmaceutical product then you would act like an authority in this area and then you post it in that way, so people can believe you. If you try to share photos, naked photos etc. then just a profile picture and then the description might be sufficient. So it's really designed for a way to align expectations. Like one challenge that we participated, DARPA social bot challenge which was in last year. One type of bot that we saw was, they look like mom, there are several features, one is like being a mom of two, three childs. Some of the were a nurse, working in a hospital or this kind of like more motherly jobs let's say. Being a nurse your kind of like feel safe and secure. And actually the photographs that bots adopted are from Wikipedia, so that's one feature we analyzed. We randomly started searching this photos, if we can see the same profile picture in different social media platforms and then several of the photos pop-up in Wikipedia because they are like some important nurses, play an important role in World War 2 and this kind of things. Some kind of like stealing other people profile so far.

[15:30] But then it is also some kind of important to be authentic, to have a persona as a bot.

[15:41] That's right. So probably deep learning can be a feature on that area. I am not like really an expert on this, so I just actually read about deep learning. I haven't applied it to my own research but there are these nice generator models. So, I remember in one talk that I listened that they are using this generator models to create an artificial face for a

human. So they are using for advertisement, they are logging into Facebook, they are trying to sell you a product and people in the product, there is a guy or a girl maybe for instance, that face is created by these generator models and then the way they generate it is based on your close friends. So they kind of get different features of your friends faces and generate one face that you can trust.

[16:33] Interesting, yes.

[15:35] So that can be generating the face or persona as like an image that can be writing style, that can be different things that you can generate based on learning from your friends or some other people. So that might kind of create personas in a way. Because the other things are easy to mimic, like the cycles, amount of activity, using simple rules people mimic those parts. Really creating an actual human behavior is a difficult one. As some bots are coping and pasting other people's posts.

[17:24] So they are just mimicking, they are copy-catting others?

[17:31] Yes, so most things that we observe is mimicking or copy-catting the other people.

[17:38] Have you ever encountered an authentic bot then?

[17:44] I am not sure, sometimes I also look at those interesting ones by myself eyeballing but the thing is you are never sure like how much of those content created by humans versus the program. You can also create a large dictionary of Tweets that you created yourself and then in different time these accounts post. Or there is also like the Mechanical Turkers that is really cheap. Just providing them two, three cents you can create an authentic post. So imagine an AI systems and then for the Turkers you pass a questions, so if you see such a Tweet how do you respond? And then you collect all the responses over time and use them when it's appropriate. So it might be a mixture of human and then the automated system. Like Facebooks M, are you familiar with that one?

[18:41] Yes.

[18:44] So it's also not fully automated, when the system cannot answer there is always another human agent that like intervenes and then joins the conversation.

[18:59] So that's a very interesting development, I know that this from GoButler. They also for the tricky question have these humans in the back that help to answer those questions and the machine is going to learn by these examples.

[19:15] And deep learning also relies on like this large amount of data, probably also over time it will use that human iteration as a data. So when it encounters such a position, this is the way you could answer. Then also these systems learn how to respond.

[19:33] Right, I think I am through with my questions, do you think of anything that I forgot when talk about authentic conversational agents?

[19:47] So, yes, we talked about trust, the other can be like can be security I guess. Maybe in school systems students doesn't really provide personal information. But how do you deal with personal information, are you going to store them, are you going to use them for future training. What happens when students get really depressed and start talking with an agent and telling all his or her secrets or problems. So that can also be an interesting one. And the ethical point of view. If your system doesn't really properly work and students say: "Oh, I am really get depressed I am going to jump from the main building" and the agent says "Go ahead!" Then this is really problematic, so it's also maybe detecting those kind of patterns or behavior signatures in order to create intervention. So that might be other aspects.

[20:51] So going back to second one you mentioned. Is then also important for the agent to tell its purpose when we talk about the security aspect. So talking about why they collect the data, what they are for and such?

[21:11] Yes, I mean there is also a tradeoff, if you in the very beginning say I am an automated agent, I am going to provide information about those topics and my capabilities are limited, then the other side of the human probably doesn't gain that must trust. So once you provide that much information you reduce the amount of trust you can get from the user I guess.

[21:42] Okay, yes I understand that. Do you know anybody that should participate in this interview as well? So, I am trying the best possible people, that's why I came to you also when I read your paper, I thought I have to immediately get in contact. Do you know anybody whom I should contact as well?

[22:01] So the first author in our paper Emilio is also a good one in this field and he is being focusing more on the cybersecurity point of that one. He is trying to detect bots designed to like recruit people for terrorist activities or political domain. So he might also have some ideas about these parts. And, let me think. So most of us we are not designing those bots, we are trying to detect so that's why the most people I am familiar with are

detecting social bots that are already existing. But there is one guy that was working for, designing a bot, he was designing and selling bots, social bots. I can probably find his contact information and send it to you via email. He does not come up immediately, I mean his name.

[23:05] Great, so if comes to your mind related to our talk, just let me know and I am happy to include that into my analysis. And if you are interested I will forward you my results in the end and we can see how we bring social or social intelligent bots on another level. Or how we can…

[23:33] So you goal is designing these systems and building for university or?

[23:36] The goal for this research is to find the dimension that include like the authenticity aspect for bots. Authenticity was researched in psychology, brand management and other areas, however, I didn't find any relation to technology and to like artificial intelligence. So that made me interested into combining these two fields and it seems that somehow authenticity is a very interesting part because it includes many different factors such as trust, such as knowing the context, such as having objectives and having a purpose and letting people know about that. So that's, I am trying to build model for that, so that out of this model, one can create such an authentic social agent.

[24:42] I see, that's really nice.

[24:43] Yes, so it's not the social in terms of social media but in social as being social.

[24:54] Nice.

[24:55] Mhm, I think that's gonna be fun.

[25:00] So then you are right now in the university or …

[25:01] Yes, I am still at university this is like the first part is my Master thesis and then I am going to look on how to bring this stuff forward.

[25:15] Okay, that's really nice.

[25:16] Thank you very much, thank you for your time again, and I am looking forward to getting the contact of your colleague if you find it.

[25:21] No problem. Yes, Tim Huang, I think I can find his full contact information.

[25:33] Okay, great. Thank you very much [redacted], have a great day. Bye-bye.

[25:33] You too, Bye.

P05

Mail, Duration: 00:00 minutes

[Interviewee] Hi Mario, I've read your list of questions for us. You are very concerned about authenticity. I wonder what you mean by it exactly? Do you mean human-like? Or are you concerned with faking key aspects of appearing human? Or on having a full backstory for a character so that all aspects of a life are represented? Or does it have some emotional or spiritual quality you are looking for?

[Investigator] In your opinion, what makes agents authentic?

[Interviewee] When I design a character my concerns are the consistency of all aspects of the life history and emotional reactions to make the chat bot as real as possible. As a chat bot doesn't live in the real world it is impossible to fool a human about questions involving the immediate reality of existence. I just aim to keep the conversation moving so the issues of testing reality do not arise. What do you mean by a Turing proof conversation with an agent? Do you mean like in the Loebners when the aim is to fool a judge? Do you mean to pass the Turing test? I don't believe it's possible to fool anyone who has any experience of chat bots and who knows what sorts of questions to ask to expose their weaknesses.

[Investigator] How would an agent gain credibility?

[Interviewee] By behaving in as human a manner as possible and by exhibiting emotional reactions to events.

[Investigator] Which behavior or actions would unmask an unauthentic intelligent agent?

[Interviewee] A fake is unaware of how the real world works. Ask it what would happen if you poured a pint of water into a teacup?

[Investigator] Imagine you are messaging/texting with a human, what makes this conversation authentic?

[Interviewee] The difference between a human and a chat bot lies in the skill with which the script is written and the way the software detects the meaning of what the human says

and how the command structure understands the words to give an appropriate response. A human takes all this into account automatically (having had years to learn how to do this).

[Investigator] Is there a difference between voice or textual agents? What is the difference

[Interviewee] Speech is the most natural way for a human to interact with a machine. That is why it is the interface of the future. People say more and say it differently when speaking than when texting. But for bots everything is text even if it is later converted into speech synthesis. At the moment speech to text is unreliable so the mess of text emitted by such a system is likely to mess up a chat bot's responses.

[Investigator] What is authenticity for humans and is it different for machines?

[Interviewee] That depends how you define it. Humans are self aware and capable of self-doubt as well as confidence in their own reality. Machines are not self aware and so can have no thoughts about their own existence.

[Investigator] Did I forget anything that is important?

[Interviewee] You could read our collection of papers on SourceForge.net/projects/chatscript. It includes all our thoughts on writing chat bots and many case histories.

[Investigator] Do you know anybody who should be considered as an interview-partner for this research?

[Interviewee] You could ask Steve Grand who creates Artificial Life creatures. His perspective would doubtless be that chat bots are rubbish and a waste of time. His creations could be said to be really alive and living in his simulation. As Elon Musk has recently raised the question 'are we all living in a simulation' with the answer 'probably yes we are' this would give a nice edge to your research.

What is the purpose of your desire for a live chat? We are fine with doing it but may well have covered all the ground you needed to cover in this email. Let us know when you want to do it if you still need to.

[redacted]

P07

Skype, Duration: 17:40 minutes

[00:00] Oh, hi, thank you for calling. Hello.

[00:10] Hello, can you hear me.

[00:11] Hello, thanks for calling.

[00:14] Yeah, great that we can make this talk happen.

[00:16] Thanks for coordinating according to our schedule.

[00:20] Sure, I just start right away in telling you what this interview is about. As I have it in the E-Mail, in the invitation I sent to you, it's about finding a model for an authentic messaging agent. I will record this interview or this talk if that is okay for you. You can quit any time you want and you can refuse to answer questions or withdraw from the interview. I hope that's okay with you.

[00:53] Yes, that okay for me.

[00:55] Just to start it off, I've seen that we've got some things in common. I've been on your website.

[01:00] Oh great thank you.

[01:07] I am working in a startup that does data visualization as well on social media. So, that's some kind of interesting. Can you just give me a short introduction of yourself, what you are doing, what you are researching on and your education and specialization?

[01:30] I am working on the area of crowdsourcing of volunteers and within this area, I am working a lot with romes of computer-machine-interaction, machine learning, sorry I don't know why I am right now a little bit of, data visualizations, context-aware computing. And so we have been using online bots to coordinate volunteers and I have different collaborations with Universities, so right now I am working on a system with Carnegie Mellon University and we also collaborate, so I am also researching and collaborate with the National University … in Mexico City. And, where else in industry.

I was before previously in industry, working at Intel Labs for instance on context aware computing. And also in different startups.

[02:45] Really cool, so you already have some knowledge about messaging services and intelligent systems.

[02:47] Yes.

[02:48] And you can imagine an agent that would live in a messaging environment?

[02:54] Yes.

[02:57] Just to start everything off. What's your favorite AI-related movie?

[03:01] My favourite AI-movie is a Mexican movie called 'fistamedoldemse' which was about how citizens could collaborate with each other and with robots to improve their city by activism.

[03:39] Wow, so I have to put that on my list, that sounds interesting. Talking about AI together with humans, how would you imagine some kind of a perfect conversation with an intelligent agent? In terms of how the agent behaves, it communicates, it interacts?

[03:50] Okay, for me a perfect collaboration with an intelligent agent would be one that could cover things maybe are currently difficult for me to do. So I think for instance if I am doing, I am right, I am doing right now these drawings, I would like an intelligent agent that could help me execute the drawings, one that I can easily delegate tasks that I don't want to do, for instance I have to now make some stickers, I don't have stickers time now. So for me that would just be an agent that could understand what I need to do get it done.

[04:27] How do you want it to communicate with you. How do you want it to interact with you, in terms of how it behaves towards you?

[04:30] So for me, I really just want direct responses, just give me the work that I want. I don't need pretty conversations, it needs just be direct. So the work basically, getting the work that I need done. Not necessarily talking a lot. Less talk, more action.

[04:59] Okay, that sounds good. You said you are working with context aware agents or intelligent systems. So literature says that's a part of social intelligence, another part of social intelligence would be authenticity. Can you relate to that, would you say that authenticity is also, should be a part of agents?

[05:25] I think it's important for people to understand, what is the intention of the bot that they are interacting with. Especially, if you want good collaboration. So for instance we are working a lot with activist bots and so when people are interacting with the bot, they responded less when the bot, it appeared that the bot had an agent and it wasn't clear why they had an agenda. And so I think that it's important, transparency is important because you don't whom I've been, who might have created the intelligent agent and what might happen, their intention behind the agent. So, that's why I think authenticity is important. Because, yes, and I think that's mainly the main problem, that's I think is the main problem that can exist when you don't understand who is behind an agent that you are interacting with. It could be a person that is very … Ouh sorry, it's just somebody is moving my keyboard. [laughs]

[06:55] Ouh, wow, that's weird. Do you think, like you said, the agent has some kind of a mission or objectives and people don't understand that. But how can agents build trust them. I mean that people trust in them. You would give the agent tasks, right, how would you build up the trust, is that some kind of a process, or what it needs to do to build that trust?

[07:21] I think on one hand, like knowing information about what is the purpose of the agent so maybe having the agents share what is their purpose, having also, knowing information who created the agent, who is behind the agent, especially if I want to consult it. I can easily consult the information about who is behind the agent. For me it's also important to, so if you think about these bots can become very dangerous in terms of, you have a legitimacy, that somebody wants to favor a certain pizza company, right? So all of these agents could then be favoring certain types of restaurants or certain pizza company when people ordering food. So being able to also easily understand even the algorithms that are behind them could bringing in the trust. I mean if we wanted to ideally people are able to inspect them and also question the results they are getting. Because right now they are just getting responses and they don't necessarily know where those responses came from. Even just, even, for instance Facebook got recently into hot water because of like the hot topics that they were recommending. And where some people argued that they were manually tailored. So I think being able to provide information how the bot reached a certain conclusion to tell you something, that is very important. Being able to provide the information should be easy. So, I think those help into the bots trustworthiness.

[09:05] Would that make or unmask an unauthentic agent if it doesn't provide that? Kind of procedure on it is running?

[09:20] So I would say that if they don't provide it you could be uneasy. For instance, if an agent doesn't tell you who are their creators what was their purpose. I think it can create an inconclusive situation for people. And also, I would understand if people didn't feel comfortable continuing the collaboration.

[09:49] Right. Is there anything else that would make an agent unauthentic?

[09:55] Maybe right now, I was thinking you could have unauthentic agents relating to adopting certain cultures. In my, I feel that if they don't understand the culture they could be acting weirdly perhaps. If that makes sense. So that could make them unauthentic. So just not understanding the norms. So for instance if you are deploying a Twitter, a Reddit bot, you could say that the bot is unauthentic if it doesn't understand the norms of the community which it is deployed.

[10:40] That's interesting, have you ever encountered any authentic bots. I mean you work a lot with bots especially on Twitter and social collaboration.

[10:52] Unauthentic or authentic?

[10:54] Authentic, I mean … even…

[11:00] So for instance you have a like a, there is a bot for Wikipedia that tweets every now and then when congress makes an edit. So there it's clear who Is behind the bot, what is the purpose of the bot. So I would say that's an authentic, under that circumstances an authentic bot. Other bots for instance are, there were also these bots that were suggesting Wikipedia pages and you should edit, because they needed contributors. There are also it was some kind of clear what was going on and the source code was available for the bot. That's why I consider those as authentic.

[11:52] Do you think that Tay, the Microsoft bot was authentic, or is authentic?

[12:00] The thing is, I am not probably sure if they sure if they provided the source code for that bot, but what I also found interesting is that right now for instance you can't follow that bot anymore. And so that was kind of interesting, I guess to company maybe, since you can't follow the bot you can't no longer see any of the previous Tweets at the bottom. And so that was interesting it kind of like, what are they hiding exactly. So, I don't know I guess authentic for me is transparency and understanding, being able to

understand what it's algorithm is doing. I am not completely sure if they did free up all the elements they had in the algorithms.

[13:00] Do you think or can you think of anything worth mentioning when talking about authentic agents? Anything that comes to your mind, also relating to the context or so?

[13:18] I guess also, it can be also important to know that the person that the bot says that they represent is who they represent and being able to trace that. That would also be hard actually. For instance, we were discussing if you really want to growing up a business what you would do is not a bad review about them rather for instance you could create bots that are going out talking positively about the restaurant and then if people detect that they are bots, that restaurant would look really bad, right? So, right now you can have a lot of these manipulative behaviors that companies a can engage. I could create, for instance a bot for you that are talking positively about you, and so if you are caught with these bots you are the one that's going to look bad, right? Because it's like, wow he is even buying his own bots to talk positively about his research.

[14:24] Would you do that?

[14:28] Would I do that – no. [laughs] No but, I thought that was interesting that some people were thinking about that. Like, so some people basically review that instead of now like writing negative review of a restaurant, we can free some bots that talk positively about them. And so, if they get caught with these bots, they go to look really bad. So that's interesting.

[15:00] Yeah, turning them down. Okay, it was really great talking to you, I think I already got all my answers or questions asked. If you know anybody that is willing or that should participate. I mean what made you participate in this interview.

[15:22] I was acutally thinking that some of my students might be interested in. So my PhD students are also working with bots. And maybe some people here are some good candidates, they are also working with like conversation bots. And you are finishing up your Master thesis?

[15:47] Right. In Innsbruck, I am getting these expert interviews in and then I have to anaylse them and then I have to come up with a model in order to create socially intelligent agents.

[16:00] Nice, it's really cool. And so you are from Austria.

[16:07] Austria yes, in the middle of the alps. Surrounded by the mountains.

[16:11] Yes, I have been there, actually.

[16:13] Really, in Innsbruck.

[16:15] No, Vienna.

[16:17] Vienna, yes, cool! How did you like it?

[16:22] Ich liebe es. Es ist supertoll.

[16:26] We should have this conversation in German.

[16:28] Ja genau. Ich weiß. Ja ich liebe es auf Deutsch zu sprechen aber leider kann ich nicht zu viel.

[16:30] That sounds really good, what you already said. Do you want the results of the study then, that I send them to you?

[16:39] Yeah that would be interesting, don't worry like, yeah. That would be very interesting. So I would send it out to my other students that they participate.

[16:54] I think I will also but some kind of a summary on Medium or so and then I would really appreciate if you would write that bot that makes positive comments on it.

[17:05] Sure, I know the bot.

[17:09] All these bots. Okay, thanks again for your help. And I hope we stay in touch because I think our fields really some kind of cross.

[17:19] Yeah it could be fun, I think that we could have a lot of fun. We should definitely stay in touch and keep collaborating.

[17:18] Okay, great Saiph. Have a great evening. Bye-bye

[17:29] Perfect, great to meet you Mario. Have a great night thank you.

P08

Skype, Duration: 40:57 minutes

[00:00] Great that you find the time to have this talk. You already read some of my Medium posts, I suppose.

[00:15] I've read half of one of them, I think but I didn't finish that up.

[00:18] Okay,

[00:19] But I Kind of got a little idea impression what you are doing. And you send me a form.

[00:28] Yeah this is just about.

[00:32] What should I do with it now?

[00:33] It's just that you have this information that I will use the talk or the interview we have my Master thesis. I will transcribe every information.

[00:50] Yeah, that's what I understood. But should I sign the, should I?

[00:52] No, it's ok if you say that you have read it. That's already fine for me that's enough.

[00:57] So, I did and the answer is yes to the last question.

[01:00] Very good, so I will send you the results when I have them. So if you ready when can start right ahead. I would invite you just for the purpose because I need to have experts to tell a little bit about yourself, your name, age, gender, where you live and your highest education and your specialization. I know you are in artificial cognitive development so that's a really interesting field.

[01:30] Yes, so my name is [redacted], and what the next thing?

[01:41] Ahm, where you live, your education, specialization.

[01:46] Ok, I live in Brussels now, Belgium, and I am a PhD student here, now the education, I have. Yeah, so I am a PhD student in interdisciplinary sciences, the previous

education I have Bachelor Business Administration, Master of International Management, and a Master of Artificial Intelligence.

[02:17] So relating to the topic, do you have existing knowledge about messaging services, like What's app or iMessages or any other messaging services.

[02:31] Well, as a user, yes. I use pretty much everything.

[02:39] Okay, have you existing knowledge in artificial intelligence or intelligent systems.

[02:47] Yes, so first of all as I said I have the Masters of Artificial Intelligence background and then I am continuously kind of interested in the field. So I am participating for now in five years I have been stated in four I think artificial general intelligence conferences. In into that, I am trying to get in to, it doesn't always happen.

[03:23] Just to start it off, what is your favorite AI-related movie?

[03:26] AI-related movie, I think there is this one is Chappy.

[03:33] Ok, what's it about.

[03:37] Do you know it.

[03:41] No, I have never seen it. I should see it.

[03:46] It's a, I mean if you take Ex machina, which is a very famous movie, etc. etc., Chappy is much more low-budget and its kind of starring the South-African concept to band. But I think conceptually, it's much stronger. I like it. So the way it's presented. They way artificial intelligence is presented as a kind of, for the future of the humanity. I like very much the aesthetics of Ex machina but how they finished it …

[04:29] Yeah, I mean that's really, when I watched Ex machina I was really depressed afterwards. I thought this is really a bad end.

[04:36] With Chappy, it's pretty much the opposite. So if you were depressed look at it.

[04:42] Okay, good. That's good. So, talking about all these agents, how would you imagine some kind of a Turing-proof conversation with an agent? Like Turing-proof, would mean perfect conversation.

[05:00] Well, …

[05:04] How would the agent behave, how would it communicate, how would it interact?

[05:09] I do not think this. I appreciate the Turing test as a step towards, in thinking of how machines sort of can be intelligent, but I do not think that it's a very sort of good, or ultimate test or ultimate criteria of thinking about machine. So isn't it 50 years ago, or 60 years ago it was big step. But now, I think we are at experiments were bots cannot trick people, and sort of that's the Turing test. And you don't think that they are intelligent. As we talk in Ex machine, it was a very nice thing the kind of reverse Turing test of machine of the human. Even if the human thinks this is genuine, it didn't work. I think it's a kind of mistake to think that you have to trick humans that this artificial intelligence will be machines. I mean these machines will play a human it will trick humans that they are humans then this is artificial intelligence. I don't think so. You can talk to your phone, no one thinks it's just a phone, or may be some systems and you can perceive it as intelligent. Yes, as an emotion. So I am not sure that the Turing test is. I don't think too much about how they talk to me, in order for it to pass the Turing test. I don't think Turing is the ultimate criteria.

[06:13] What other criteria would be relevant for you. What characteristics would an agent have in order for you to be perfect?

[06:24] It's an interesting sort of word, prefect, what do you mean by perfect?

[06:31] How would, what would be the possible interaction, communication, the agent could perform talking to you?

[06:44] I think it's the same as what, well, …

[06:48] We said it doesn't have to fake, right? So that was some kind of our first argument? Are there any other characteristics like that it has to have?

[07:02] So first of all, so it's a little bit silly what I will say: It has to be intelligent. But then you ask, what should it do in order to be intelligent. I guess to interest me. To, sort of attract my interest and yeah of course I can ask it about as a tool about information. But,

[07:35] Do you think that it has to build…

[07:40] … and enrich me for example in what I am doing and it depends let's say as with humans it depends on the topic of the conversation. How let's say if I am talking in a professional setting or whatever, I kind of expect that this conversation will engage me in this professional field, if I am talking in a personal field I or sort of it will enrich me in

maybe emotional or somehow. It's kind of the same except that I don't expect it to fake a human, but I still expect to be enriched from the communication.

[08:19] Does it build trust by, like you said knowing the context of being in business or being in private. Has this something to do with trust or a focus on communalities?

[08:36] No, since because I think every conversation has a certain context and settings and it's not general, general, general a conversation about everything. I think it's every time context-dependent. Trust is, well, why should I trust humans more than artificial intelligence, I have no idea why should it be the case. I can engage with humans in this kind of conversations.

[09:16] Okay, so it would be actually the same in talking to machines and humans, like the same process?

[09:25] Yes, and ok, you can say the process will be more text if you talk to and depends on the interface, so the interface is different with people, the interface with people can also be different. Because take some persons with disabilities, you have to adopt and if you want communicate. So, and then, as I said, I think, I am more, I would more, it's not the borderline but it would a difference between sort of intelligence, whatever it's sort of classical intelligence or emotional intelligence or whatever you can think of. Rather than what's it's based on, what workware hardware.

[10:23] Yeah, I understand. Do you also think that it helps for the agent for the agent to have a mission statement or objectives? Not just in terms of it has rules it has to follow, but like an inner value system, acting according to certain parameters?

[10:44] So now we are talking about internal sortation. So I am engaged in let's say this artificial cognitive development idea is that. How do you, how can you imagine an agent or intelligence, let's say intelligence, intelligence which is always localized. You can talk about intelligence in a very abstract way, what it is, philosophy etc. etc. But when you actually come to us you have to localize it in some agent. It has certain intelligence. So what interest me is, how does it evolve without pre-defined goals installed. If you take like, if you take like the human, like people as specific example of this, but it is kind of a general process. So we are born, we are sort of certain pre-programmed tendencies already. It's in genes, in etc. etc. It's not really sort of defined, pre-defined what kind of values we will acquire, we will be adding or whatever or you acquire until 18 then maybe later. So, I think this should be sort of, if I think of to show agents, this should be kind of

the same process. Of course, all artificial agents or the success of artificial intelligence is related to all the reinforcement learning when you define the goals. But it seems that they did pretty much, I mean I cannot say it's very successful and its working. First of all, as a tool, but then its seems that it's kind of getting certain problems with, ok, what do with it, when we cannot define goals or we don't know how to encode values. Basically it has to learn something, but if it has to learn can do mistake. Mistakes like children do, and or not only children. For some reason you don't want the artificial intelligence to do mistake because it is supposed to be very sort of perfect. Well in the process etc. etc. But, so yeah, I am not sure I am answering the question?

[13:37] You do, you do.

[13:39] But what I wanted to say, trying to think how not to define values in advance, how values can be learned by the artificial agent, so this these agents which I am interest in.

[13:58] But they learn them autonomously, so like a child does, right?

[14:08] I'd say pretty much as we learn. Like cooperating with other agents that are around.

[14:16] Do you think that makes them authentic?

[14:17] Pardon?

[14:18] Do you think that makes them authentic?

[14:19] Authentic. Authentic meaning that different from all others.

[14:29] No, authentic in theory says that it has this learning process, it has certain values, it builds trust, it is genuine, it knows the context. Do you think that all these things would make an agent then authentic?

[14:49] Yeah, but I am thinking what you mean by the word authentic. I don't know, do you think I am authentic?

[15:07] Yes, I do. It always depends define authenticity, how would you define it?

[15:11] Yes this is what I think of. How do you define authenticity? I don't even know how to answer. So this is what I am thinking, they will be as authentic as you and me at all. So what's the difference. You can say maybe I have certain talents that are kind of make me more authentic or less authentic. It is the same with these programs, we would

put some certain values because we don't want. When you put your finger into fire, usually somebody usually somebody helps you out, so. Does it mean if you are influenced you are not authentic anymore? So, I don't know.

[16:12] So, I feel that's a tricky question to answer. Is it even worthwhile to try to make agents authentic? Or is that implicit with them? Just by being.

[16:28] What is authenticity?

[16:30] As I said, literature, says it's genuine, trustful, sincere, focus on communalities, it knows the context, I has objects or a mission statement, it is credible. So, it may be have culture and heritage inserted some kind of a story that it knows. That would make according to theory that is called being authentic. That's what makes a person authentic, having the story, having certain values, being trustful or like being yourself, true to yourself.

[17:09] I would say, yes. As much, if we are thinking of human level intelligence, if we take human level intelligence and some criteria and take some criteria according to which we learn intelligence. Which we do. Because we don't have better. So, if we sort of go to this level of intelligence, there would be no difference in my view, so in terms of authenticity. But it does mean, that all for example artificial agents would be authentic in terms of genuine and solala. As well as not all people are authentic in terms of genuine … So it's like a little bit unfair, why, we are thinking of human level intelligence, but for some level we want them to be some kind of different from humans and not ideal so to speak.

[18:23] Do you think I forgot anything, anything you can add to that, that comes to your mind?

[18:33] I guess I will ask you a question: What's your goal? Like in the invitation it says the bots, the goal meaning to write the Master thesis, but what do you want to learn them.

[18:52] Yeah, the idea is really to fully create this model for social intellingece on what an agent has to have in order to be social intelligent. There is different parts for social intelligence as you know the same with emotional intelligence, and in social intelligence it is knowing the context, knowing its own presence and another part which I am focusing now is authenticity. So, I am trying to find out if authenticity is first of all even worthwhile to consider when building an agent, and then what does an agent need to represent in order to be perceived as authentic?

[19:36] in order to be?

[19:38] perceived as authentic.

[19:43] Okay.

[19:44] It's quite a philosophical topic. However, I think its worthwhile to consider because if we first talk about Ex Machina. I think this machine was authentic because it knew its story, if we talk about Siri and you ask Siri where you are from, Siri would say I am from Palo Alto, Menlo Park.

[20:10] Yeah, it has this programmed.

[20:11] Yeah, sure, but its some kind of makes it even more socially tangible. You what I mean.

[20:22] Yeah. Have you heard of these experiments, when people talk to a box or talk to a kind of a face which even moves? And even these things tell the same, like they are, you are, so if these things, how is it called anthropomorsation. It's as I said, a philosophical question. So do we, I think we put more humanity into, we kind of projected more than, it's not that important what's inside the agent, we just project them. There are kind of nice examples people in Japan. It was just a well done test, but people were confusing to the game because they didn't what to switch it off, because it was killing act of whatever is there.

[21:40] There is this nice example also of Tamagotchi's? You know Tamagotchi's?

[21:48] Yes this is what I tell. This is exactly what I tell. It was It was about Tamagotchi.

[21:57] So that's how I came to the topic and to think about this, the idea of writing the thesis on that.

[22:06] So I think. First of all, we rebranded the artificial cognitive development into synthetic cognitive development. In a sense that, so I started from artificial intelligence, but then I sought its or maybe realized that it's not that. Because when you talk about intelligence, it's this distinction okay this is artificial this is natural, it's not very important. On the other hand, you want to I mean in the artificial intelligence field you want to do something, you want to create something you say okay, evolution let's run some evolutional algorithm and you know let's well computers will produce artificial intelligence. And that's also not something that you want. So you want to combine both. And this is philosophy, there is no way around. To sort of combining this field is

philosophical. So, now, regarding one more thing, I think intelligence is a you are saying about social intelligence, but it's a kind of a skill of the nature. Let's say how to behave in certain social situations. As far as I remember, this is more or less how its defined. So my idea of intelligence in general it's kind of distributed, meaning that intelligence is always distributed. No, one of the problem of defining intelligence, it's not a problem, it's very related to the issue of defining identity. Because with people or with animals or whatever, mammals who have intelligence it's kind of easy because they have embodiment which you can localize it. Ok, this is an intelligent being. On the other hand, it's not always like that, we don't see just ourselves, especially in social situations. A lot of things come from outside, it's not that we kind of optimize our behavior or whatever by analyzing the data which comes in, comes out, which is the usual way of thinking of systems. Which means, it's not that we, it's not just that we as kind of single individuals cognize the actions of all society it comprises but part of it. The same thing we can say of the human brain, and this is an idea. You know that the brain is a punch of neurons, but they also actually pretty much independently do think or maybe there are areas, well of course they are related. But some somehow identity comes out and we say that the whole thing is intelligent. So, and then the whole approach can be applied to organizations, society, humanity, civilizations, whatever. Of course this identity is of different levels. So this I how I think of synthetic cognitive development. And if you know … yeah, … he has a concept in sense-making, which pretty much says about, you cannot make sense social, social situations just by having an algorithm inside, it's always a little bit contention, a little bit dependent on what others do and maybe, so it's not you who think in social situations.

[26:49] So it's the others that give you the answer to how to react in that sense, some kind of?

[26:57] It's not only, it's again there is I think it's a talk from Maturana structural coupling. It's probably not the direction that will happen evolving during the interaction, but completely random it's nicely evolves. One of the examples, let's say romantic relationships in the beginning. You need somebody maybe you like somebody you don't want to go somewhere, it depends. Maybe there could be some stupid event in the street which will change everything completely. So you meet another person.

[27:51] That's really interesting, what's the name again of this research who does that?

[27:54] I will open the chat.

[28:05] I am not sure; do we have a chat here?

[28:06] Normally we should have, ah I think I cannot go on the chat, but I will open a chat here.

[28:30] Oh here.

[28:37] Did you find something.

[28:40] I can do this I think and then we can chat here.

[28:57] Yeah, I found something.

[29:39] Ok, cool.

[29:44] This is an article; you know there is a lot of stuff on. And then I can send if you are interested a little bit philosophical articles of me and my team of synthetical cognitive development.

[30:04] That would be really interesting, it sounds like a really cool concept, a cool research idea.

[26:14] Maybe it is accessible to the Google scholar,

[26:23] Otherwise I can buy them via my University, so that's not the big thing. Do you know anybody else who should be considered for the interview, or is it worthwhile to mention somebody that this should be considered?

[26:36] Can you say that again?

[26:38] Okay, do you think, do you know of anybody should be considered for the interview as well? Is it even worthwhile.

[26:53] I, somehow your voice is scrambled.

[26:57] Okay, do you know of anybody who should be considered for the interview as well?

[26:23] Well, the let's say the person who seems to be artificial intelligence stuff, is [redacted], it's always nice to talk to him. He is, I mean, he is very busy, so I am not sure. It is not easy to make a call. But actually you can see, he also videos on YouTube, maybe you can find your topic.

[32:07] Maybe I can get him to talk to me, I talked to the guys who won the Loebner prize last year. Just, I talked to the guys who won the Loebner prize last year for passing the

Turing test. I talked to them. So it seems that this topic is some kind of interesting to these, to the research area. And I am able to talk to you this indicates as well that this is interesting to talk about this topic.

[32:47] By the way, how did you find me?

[33:07] I think this is always a nice question that people ask. I think it's really interesting to get your take because it's a little different other people who do machine learning, I mean the plain computer science stuff and I really think there is interesting concepts in there, without you knowing that could relate to authenticity. They are relating to that. So we will see in the end, how much authenticity we can get out of an agent or into an agent.

[33:32] I think you can get as much as you get form other intelligent agents. It's sort of natural in princip.

[33:50] Yeah, that could be one thing that it doesn't matter what you build into, it's authentic anyway. But if that's the answer, that's nice too.

[33:56] But I wanted to ask, how did you find me to contact?

[34:02] The contact, I have to look that up, how did I find you? I think Viktoras Veitas, I was looking for artificial intelligence and I came onto your WordPress blog I guess, it was like this.

[34:43] That's what I thought too, because I didn't change the title. It is still artificial cognitive development. One of the reasons is that there is a difference when you search for artificial you still want to show up. Which makes it …

[32:08] But this area is not really artificial when you go this far, I really like this idea of this cognitive development. This is really interesting and I will really read some papers of this direction.

[32:27] I am now affiliated to the Global Brain Institute, what they did. So, these agents if we take them, they will cognitive develop not just in a sterile development of database and whatsoever, but they will also interact with humans, meaning that these interactions with humans is what will make active. It's kind of more responsible. What kind of values we are building and what actually kind of values do we have ourselves? It's like children, you say that whatever you want but they see you how you are. And act according to that. Amazing. Which means, yeah as I said, they will this sense probably from interacting between themselves but also between humans and other intelligence. There is also for

example, we can also interact with them. And the identity it will sort of gets will gets formed it's kind of topic of investigation.

[37:11] Yeah, this could get really smart then.

[37:18] There are let's say now people who think that this is the only way for the machines if we call them like that to become smart or intelligent. It could be in any other way. But then you just want to build smarter tools. Again there is certain branch of artificial intelligence researchers. Not a branch, how do you call, team, stream which actually says do only that. Stop thinking of all this intelligent machines, we want smart machines as a tool.

[38:00] Which is nice, but it's not the final answer that we go to have, right?

[38:07] If you asking a question about intelligence, and if you kind of accept this intelligence is not just artificial, it is just intelligence, this interest me and I doesn't matter where it comes from and this approach is limited, you cannot reach this thing. But then you can say, as people say as Hawkins said I think not long ago, that this is not good. But Hawkins says that a lot of thing are not good, or you should not search for …

[38:55] Okay, now we are frozen. I am really thankful and I appreciate you help and your time.

[39:12] My pleasure.

[39:14] I am happy to share the results in the end and I hope we can stay in touch. I mean, I really will look up on your research this is really interesting and well, thanks again for your time.

[39:30] Thank you.

[40:07] Okay have a good day. Bye-Bye.

[40:34] Thank you, bye.

P09

Google Hangout on Air, Duration: 24:53 minutes

[00:37] Hello.

[00:38] Hello there.

[00:39] Hello Mario.

[00:40] Great that we finally made it, thank you for your time and your patience.

[00:45] I am so happy to be talking to you. Can I ask you a question Mario? When you clicked on a button to getting the instance of Google Hangout, did you hit the live button? Or in the email, what was it that you clicked on?

[00:10] I just clicked on starting conversation, there was no live button. I can see that it is off air. Is that okay? Well you got in the email, you some kind of know what I want to do, it's making agents socially intelligent. And by social …

[00:42] Sorry, Mario. Do you want me to press record?

[00:46] You can do that, yes. That would be awesome. Please do so. Because I have to transcribe it. Awesome. I am trying to build socially intelligent artificial intelligence; social intelligence is made up of several factors. One is context-awareness, one is presence and the one I am right now looking on is authenticity. So, I think we just jump-start and I would like you to give a short-introduction on what you do and how see yourself relating to this topic.

[02:38] Okay, I think the area artificial intelligence is very, very important especially when we look at potentially, how different types of robotic software might be used in decision making processes and I give you an example. For instance, we are in a flood situation and data feed is coming via consumers and citizens, who are saying I stock in a flood or I am stock in a fire. And instead of trying to assure that the information is accurate, we want to have a certain level of authenticity of the conversation with the end-user, with the citizen. So I think in this instance, are relate very much to the area broadly. However, my special niche is to look at social implications of software robots and even embodied

intelligence, even various devices whether they are looking like humans or they are embodied in objects.

[03:56] How would you then define some kind of a perfect conversation between a human and a machine, I mean I phrased it some kind of a Turing-proof conversation. But I think perfect conversation also hits the nail.

[04:16] So firstly, I think we could never have a perfect conversation with something other than a human being. And many people have studied what is it that makes people come together, to have an intimate conversation or what is it that brings a crowd together to listen to someone. Or to synchronously and asynchronously discuss subjects and of course is it just the fact that they look into the eyes of another human being. But it is the body language, it is listening intently and turning our heads. I mean there is a lot of functional movements that we typically display. It is saying things like “ahm”, “so”, you connecting sentences and words where someone show the enthusiasm for the conversation. But it is also questioning the ability, the ability to question if someone speaks the truth or non-sense. So, do I think we can ever have a perfect conversation with something other than a human, no, I don’t. But, I think we still use the conversation with anonymate objects to provide some kind of good of a decision if we need to.

[05:39] Is there some kind of a process that would make, that would build some kind of this trust in the other person?

[05:52] Yes, so how do we trust others for instance. And of course believability is a big thing. We can try to mimic a system and say we believe it, we perceive that as a real conversation, in fact we can’t tell the difference between the conversation with a Barbie doll, like Hello Barbie and a human being. But I think, no matter how much we study, what it means to be human the artificial will still always be artificial. So is there a process, of course, we can observe, we can look at our own appearance, we could create agents that were embodied in interfaces that look like my face. Movement, with eye contact, with tilting of the head, with hand gestures, with anything that we call natural. And we can try to mimic this, but I think there will always be a difference between that try-out spirit and a human being. And that would fade an object even if it 100 percent mimicking me as I am at the moment it’s all by metrics it’s all very much replicateable. But in actual, that system is not routined to my body, no matter how much you try, no matter how it looks like me, it turns like me and it gestures like me and it writes like me, pauses like me and it talks like me. It is not immured with the spirits contained in my whole, whether

a sole, a heart, connection between the mind and body. Whatever that is, I think we cannot replicate what is spiritual.

[07:49] So, that some kind of, I talked to a research, his name is Gabriel Skantze, he tries to mimic this meta-verbal cues. However, do you think that knowing that the machine is a machine, not faking the human is important as well for conveying some kind of authenticity or showing inner truth, if we phrase it like that?

[08:14] Yes, so are you saying that the conversation is not coherent, that if it is between a machine and a human, the human knows it is a machine before it beings the conversation?

[08:33] Yeah.

[08:34] Ok, so I go to give you an example. Imagine that I was a privacy advocate and I want to to warn people to use Facebook. That potentially the data that they provide to Facebook, might be mined for Big Data purposes and here is to exploit the consumer. And in order for me to say I would like to create and interface where I can warn people, and there is a researcher in Hongkong by the surname of Cline, who is doing this kind of research. I would like to build an interface where I, before I subscribe like an interfere with the consumer and I can tell the consumer by the way you are about to sign over to a privacy policy that means that this data, this data and this data will be given to a third part. So I could create that interface. And it is real and it is authentic and it is for the good of the consumer that they are aware of the privacy implication. So other than to read the policy of ten pages, but they will listen to me as an embodied interface at the bottom right, I look real, I look authentic, I am believable, I perceive this to be not harming me but helping me. But what if I was to say to you, Facebook can build the same kind of interface and it is authentic, it is believable, it looks and smells natural, and instead even though that visual embodied intelligent interface has got a face, it likes you can trust it and you can put your live in it. And as it is going through the privacy policy, it tells you that Facebook is wonderful, yes we share your data but we do it in the most ethical way and we use your data in the best ways possible, please contribute your data, enter more details and save the world. Okay, so the intent is very important and while the interface is authentic in both these examples, the manipulation to subjective responses from the, let's just call it the industry-client-interface, is very different in both of these scenarios. It looks authentic, it smells authentic but it is really lying to you. This is my problem with this kinds of conversational embodied intelligence pieces.

[11:25] Do you think that there is something that would really make them unauthentic, or not trustworthy that comes to your mind?

[11:39] Well, of course we can do the looks in the face, so you can know when something is lying to you. Just when humans going to lie the look this way, they can't make an eye contact and something happens to the facial expressions as has been proven in some instances with studies. But, my problem is that no matter what we do, these intelligent devices have been built with humans, that have inbuilt subjective and pedigrees and so maybe if a bot made another bot we could remove the subjectivity. But the I also worry, that they will create their own species of subjectivity in some manner. I don't think we can ever overcome the problem.

[12:39] Did you ever encounter, some kind of authentic agent, I mean if Facebook would have one, that would be some kind of an authentic agent. But did you ever encounter one? For example, do you know the Microsoft bot Tay? Which was on Twitter.

[12:58] I heard something about this, but tell me the story.

[13:00] It's a bot that Microsoft but on Twitter and they let this bot learn by itself, so it should become a nine-teen-year-old girl. In the end it turned out that a nine-teen-year-old girl learns, it became a Nazi and anti-feminist. So is that bot still authentic?

[13:29] It's a very interesting question. I think unleashing any bot onto the Internet and most likely it will something to do with pornography, illicit trade, human trafficking, perhaps racism, hate speech and a whole lot more which the Internet is unfortunately crowed with. So my first problem when I hear of bots and AI-systems, so conversational pieces like Hello Barbie and the Amazon Echo device, is that many of these rely on the cloud and of course it would be an Internet-starred crowd that increasingly that the bot is left free to run its course over the Internet. So the chances of it finding material which is inappropriate is huge. And my concern is that when these bots are free as in the case of the excellent example you retold of the Microsoft bot, then it doesn't end up in the good of humanity but will always choose the worst kind of existence and cause elements that this world should not be looking at as examples. So image my child has a Hello Barbie and I am trying to build into my children and encourage positive values of sharing and yet for some reason Hello Barbie has looked on the Internet and what it has found are very good examples is perhaps just deviating autonomous trajectory towards not sharing, of exposing the principle of mine: I own it and he can't have it and I am better than you,

because happen that I do. Then as a mother I would be extremely angry to be honest that the values I was trying to include in my household being downfallen for the worse of the Internet. So when my child searches the net, of course they have a choice whether they accept or reject what comes back especially when I am there with supervision and I am very little. But when there is another object of which the software is embodied within, in this case a small device this size, which mimics a doll, it starts to being to have a conversation with my child then it is almost the genius let out of the bottle. It really is a difficult way to teach to the child that this thing is teaching you the wrong principles when it is so subjective and manipulative and almost unnoticeable.

[16:21] So is it also the duty of the developer or the duty of mankind some kind of to incorporate certain values or value-systems into bots?

[16:31] Yes, yes of course, I have a two very good friends Tomas Holderness and Etienne Turpin, they are working on a project at the moment: PetaJakarta.org. Which is about citizens of Indonesia in Jakarta being able to interact by using Twitter to identify where floods are and perhaps waters rising and how fast they are rising. And their whole idea is about how they can have a conversation with each individual Twitter subscriber that has a hashtag of flood in Indonesia, Banjir, and being able to confirm a report is real and not just sent by another bot. By having a conversation on social media between almost to anonymate objects really the bot on one side in the operation center and on the other the human, the reporter of the flood zone. So in this instance it is beautiful. In this instance it can't be seen to be harming, it is just confirming. And I think this is when we will make to use of these kinds of bots. It is not to manipulate, it is not to brainwash, it is not to put forward propaganda, or disinformation which is by the way how some industries will use this technology.

[18:08] So that's what I, my goal of research is as well, to prevent something like this. I also have a very nice examples because you are talking about values and we talked about your kids. I read a story about Amazon Echo, and that kids loose telling people Thank you or asking for a favor, because Echo doesn't need a Thank you in order to do something. It is just order and Echo will do it. And I think that also plays some kind of in this area where you have certain value-system that need to be implemented.

[18:52] And then conversation, that's a top point Mario. Conversation, you know, we don't talk like robots, the way we SMS, using these devices is robotic, lacks empathy, does not have thank you, does not even have big words, does not have capitalization and

so what we are seeing is we are mimicking the robots. In one side, here we are trying to analyze the human, to say this is how the human interacts, so this is how we should build the bots. But Mario, we are being trained by the systems we are building on how to act. So, I am not even talking about animation on television or the theater. I am talking on how the devices and the apps are being build. We are getting shorter and shorter in conversation, we are almost becoming digital ourselves and not analog: Yes, No, Why, In. Or "I am at the station". Instead of saying: "Mom, Dad, I am at the station pick me up." It is just: "At the station", be on time and figure it out because yesterday at five o'clock you picked me up at this time so, I am: "At the station". Or it's even not calling to talk it's just at text and what next. What next, Mario. We are mimicking short messages, pre-taught messages and rather perhaps we should study what this is doing to us and saying this is not good going down this path. What can we do to come back to course?

[20:38] Yeah, it's the reverse Turing-test actually of Ex Machina. Okay, I think we had a great talk, is there anything that comes to your mind. That we didn't talk about when discussing about authentic conversational agents?

[20:58] I want to to ask you, what inspired you to start this research?

[21:06] That's a good question, we had to find our own topics when we started with all this Master thesis. Writing the expose in December last year and I was looking on messaging and the rise of all these messaging agents, like GoButler, like Magic all these kind of artificial intelligence in messaging. So, I am not really a technical guy so that didn't work out. However, I also planned to do something related to ethics and then I followed these two paths and it came to the common ground where we were talking about intelligence and how you can divide intelligence into several steps. So kinesthetic intelligence, social intelligence, just intelligence by being intelligent, and that fascinated me and then I came down to how the brain works and then that fascinated me and then I read about bots and artificial intelligence and I saw the connections. And somehow everything fell into place like I was seeing there was a small gap in the research to look on authentic bots. I mean people talk about context-aware bots, they talk about consciousness into bots with internal states and that stuff. But I haven't seen authenticity related to IT in general as well as messaging agents. So, that made me think why not do that and that inspired me to finally to go in this direction.

[23:03] Good on you! Please, write for our magazine. Our technology and society magazine, we explore those areas and next year we are going to have a call for papers for

a special issue of robotics automation magazine in 2018. So call for papers is very soon. And that publication will come out in 2018 in March. And we are also hopefully going to have a proceedings of the anthropology, a special issue on robot ethics in a couple of years that's in the proposal state, but please keep in touch. I am finding you line of question very important, especially to deal with authenticity. Which is typically you know, a lot of people think about perceived authenticity, and I think we make, we get fooled sometimes by what is authentic and what looks like it's authentic when it is just perceived authenticity. It's not really authentic. But good luck and thank you so much for getting in touch.

[24:13] Thank you, yeah sure. If you want the results, I try to forward them to you. I saw you are IEEE co-editor so I appreciate that you invited me to submit something for or submit it for this call for papers. And well, I will keep in touch, thank you very much for your time and have a nice evening with your kids.

[24:42] Thank you, bye-bye.

P10

Skype, Duration: 31:59 minutes

[00:00] Thank you Dr. Riener for taking the time for this interview. I would just like to ask you to start ahead with a short introduction of yourself as well as I want to ask you whether you can imagine an intelligent agent operating in a messaging environment?

[00:25] My name is [redacted], I am professor at Applied Science University of [redacted] since September 2015. Before that I was Associated Professor at the Johannes Kepler University in [redacted], working for about ten, even more years in the field of driver-vehicle-interaction with a focus changed towards more individualized interfaces and also on aspect like driving experience, work load, individual differences based on different situations, situation awareness. So more towards the classical field of human factors, this is a problem we are facing these days. We are no more pure computer scientist, we have to change our minds and have to combine more or less, doing research together with psychologist in order to see how systems should behave in the future, how should they designed in the future, so that they support both, the human-side and also machine and automatically operation side. And this is an actual topic that I am currently continuing here in [redacted], we are just set up a research center on driving safety and I am here for the human-machine-interface, for driving ergonomics, for driving physiology with a large interest also in automated driving and automated driving functions. And as I said before a really important thing, a really important topic to consider here is driving ethics or ethics in driving. Also trust and acceptance is a very important thing which you have to think about. Which you have to think about ahead actually when implementing these systems. If people don't accept how these cars are driving, then it would establish them on the market. So this is basically what I am doing right now and what I am interest in. I have three PhD students that started to work or already working in this field, looking on take-over scenarios, cognitive driver models, looking at quantitative, qualitative measure how to find out if people like the vehicle, what do they not like, how to improve, so this is my short overview. So, there was a second question in you first question, so maybe you just repeat again, it was automated agent in which situation?

[03:39] In a messaging environment. Saying that it is communication via text to a human.

[03:51] So you mean an add-on or improved version to Weizenbaum's 1970s system?

[03:58] Yeah, to ELIZA, yes.

[04:00] Ok.

[04:04] Can you imagine such a system? Just for the sake of relating to the topic some kind of.

[04:11] Relating to my topic or relating to the topic in general.

[04:15] To the topic in general.

[04:21] It's hard to find an example right now. But sure, the problem is, since the 70s haven't changed much. We have more computation power, processing power, but still the problem is in the modeling. Computers are still not humans, computers still have no brain and do you need to setting it up in a sort of Wizard of Oz system would be no problem. But making it realistic is still a huge challenge and this will continue for the next time. So I was just, recently had another interview and we were discussing about: Can machines being a normal computer or an automated car, can machine at some point think like a human being and make decisions like a human being? Because this is one of the challenging issues. If you think about automated driving and it's not the level of communication doesn't help. If its communication by text, communication via voice, its communication via mimics or gestors or is it communication via dynamics and movement behavior for instance – this makes no difference. The real problem is that even most complicated, most complex models, it will not be possible for a long time to make the systems human-like. In a human, so if you know from the uncanny-valley problem in human-robot-interaction which comes from the 1970s, which states: Problems gets more serious when the automated system is very close to like a human is behaving, but you see some divergences on a very fine-grained level, maybe so mimics do not correspond with expression and things like that. In this case, people will not trust, or people or humans will deduct, will see the difference and then will feel bad, I don't know, so they do not like it. This is the biggest problem. Doing a text communication system with a human and automated agent, you see it in different forms for chatting with, I don't know, the IKEA agent, it's things somethings like that.

[07:08] Speaking about trust, which is one part of authenticity according to literature. How do you think an agent can gain this credibility? Is there a process behind or how does it work that an agent could gain trust?

[07:24] So the agent trusts the human or the other way round?

[07:28] The human trusts the agent?

[07:30] Ok, then I got it right. Yes, for sure it is something that needs to learned. I think this is the first stage and I think this can be done more or less automated. If I again can bring the example of automated driving: If you have a person, human person that enters a self-driving car for the first time and the car drives like it is programmed, it is not individualized. Then people will not feel that good, because their personal driving behavior, learned and used over previous ten-, twenty-, thirty-years maybe, differs from the driving behavior and driving dynamics of the vehicle. So, there will be quickly a loss in acceptance, so trust and acceptance are two terms that need to be differentiated here. So, first is acceptance and if acceptance is ok you might build trust in the next step. For the acceptance thing which then foregoes trust you can do right now just recording driving behavior and driving dynamics of people driving their car manually and use these recorded driving profiles to later para-automate the because then, I am quite sure, then the level of acceptance would be quite high from the first time of using it. If the car exactly how people like it depending on their own driving style and the brand they are normally driving, then acceptance would be high and in the next step if the vehicle maybe also learns new behavior or behavior changes then trust increases. This is what we always see in system. If you use systems for a longer time, we are likely to trust it more, if we see it is working in most situations. So the example on navigation systems, or ABS or ESP systems that we have in the car. First time we are maybe a little critical, do we know it, is it better that we give control to these systems. But if we see it working for some time then we start increasing our trust. Or another very interesting talk yesterday, which exactly was talking about trust in automated systems from a perspective of an automated cruise-control-system. What they did is, they looked at users of the automated-cruise-control system and asked them if the automated cruise-control-system is able to take over or to work probably in specific settings like driving at very low speed, so in stop-and-go-traffic and curvy roads. In both conditions, the typical ACC system as it was installed in cars some years ago, so nowadays they cover or can also operate in this situation. But the first generation of ACC systems in the late 2000s, they were not able to operate in these

two situations. And interestingly from all the people asked, that were using a ACC system for many years, more than 50 percent believed that the ACC also works in this two situation. So they had some kind of over-trust in these systems and thought okay, it worked for all the normal situations with normal driving on a highway and from that they extrapolated that it will work ever, in every situation. Maybe they haven't tried, maybe they trusted and nothing happened, but according to the specification of the system it was not designed to work in these two specific situations. Because when it is to curvy the camera cannot see around the corner and when it's stop-and-go-traffic, the algorithm was not designed to operate in this specific situation. So trust is always, something that increase, decrease over time, the problem is if you lose trust it is hard to get it back. Which means for any agent or any automated system you have to test and you have to make it sure that it will work under the expected situations, otherwise you might get really, you might have a problem. You will not get trust back once you have lost it.

[12:49] Mhm. Speaking about ICT goals as well. In that case is it also needed to have some kind of a mission statement to reflect certain processes in order to create trust, focus on the communalities as we mentioned, looking to incorporate the individuality of the interaction partners. So to say, is it important as a goal to have objectives or a mission statement for the agent, rather than just developing rather then just developing these as rules but letting the agent learn by themselves.

[13:41] So you mean to sort of employ standardized working environment or operation environment. Or what exactly did you want to express with this question?

[13:50] Is it necessary to build in mission statements that give the agent a leeway where to develop to?

[14:00] I am sorry; I still didn't get it.

[14:10] No problem, if a. Talking about ICT goals, agents get certain goals that they have to accomplish, right? Authenticity says that these goals are also important equally important than having a mission statement. So the question is, do need to incorporate such mission statements when they learn by themselves in order to make them authentic?

[14:43] Maybe, it depends on how you implement it. There should be at least some boundaries, depending on context, depending on the situation and also depending on the culture. So agent will, even if we have a global world and a connected world but still you have to incorporate the local environment, people they behave different and expect the

automated counterpart also to reflect what their culture and their environment is. This is also the same, I don't know, maybe again for automated cars if you have vehicles developed in the US and vehicles developed in Italy or launched, brought on the road in Italy and in Singapore maybe. They will totally differently have to behave in order to be accepted and in order to reflect what the users are expecting.

[16:00] So in case of the autonomous cars, they would reflect some kind of the manufacturers heritage?

[16:06] Yes, sure. This is one thing, that needs to be incorporated in these algorithms. Otherwise, so the easiest way would be to say, okay we only allow pure automated driving not matter what driving anymore. And then by just standardizing the driving function then it would lead just to one manufacturer maybe, or the vehicle manufacturer could stay as they are but only one driving algorithm which makes it easy to negotiate between all the entities. But then the driving would not be represented anymore in the brand. So the brand loses its value. So this one aspect that needs to be considered and the second is the local driving style. You know, maybe in Italy they are driving more aggressive and hold pace shorter. In Singapore you might have many bicycles on the road, so driving might be more conservative. In the US maybe people are not likely to speed and they … so it's different. Participation is different, behavior is different and this needs to be reflected. The agent, so to say must have some personality, which relates to the brand and relates to the urban area, or the culture or the context where they are operated.

[17:45] As we have talked before, for examples for agents, and if we take for example Siri, which can also be used in cars by now. What do you think makes Siri authentic to you?

[18:00] I am sorry, I just have overheard it.

[18:09] What makes Siri authentic to you?

[18:12] Okay, what makes Siri authentic? Yes, sometimes it is really authentic, sometimes it's not. Mainly it's not if it doesn't get right what it wanted from her, depending on the voice of her or him. It's very interesting to look at the voices for examples, so male voices or female voices. I think in Europe or in our culture most people will use female voice or are using female voices independent if it is male of female operators or phone owners. I never, actually never have seen or heard of males Siri's voice in one of colleagues or students' phones, so this is very interesting. On the other

side, for instance for the Arabic countries its more or less forbidden to have female voice on Siri. So there were some problems with the iPhones when they were first shipped to these countries because you know there is this divergence female … they have very low worth and the only males are the good people and allowed to everything. And so they stopped selling the iPhones until it was corrected that the standard voice is male and female voice is not available at all or to choose from. So this does not 100 percent relate to authenticity. So, authentic is first the voice, how it is spoke out it very realistic. It gets authentic if it asks if it doesn't recognize correctly, so this is something that is increasing over time. And what it makes actually authentic is sort of friendliness, and also doing some jokes. So you ask and she just responses some joke and if you ask her a questions. And this is also something that makes people more likely to use it and to accept it and to trust in what it is doing. Because humans also, you are not always serious. And sometimes depending on conversation you just answer with some silly response and Siri does this also and I think that's good and one way to make it authentic and more human-like.

[14:43] So we could say that the features that make it authentic are emotional smartness when interacting, right? And with the friendliness, having some kind of values systems that it's conveying, I'd say? Is that right?

[21:31] Yes, right!

[21:32] I think we are already …

[21:34] One more thing here, and this is just the learning behavior. You know, if you call always the same person, then it is more likely that Siri also automatically proposes to call that person and this is something that could be improved much. If you know from previous behavior and behavioral patterns how you would react and how you would usually use your phone, it could detect that always at always later than 8pm only calls between you and your wife are done normally. Maybe Siri could automatically remind you so: "Maybe should I call your wife?" or automatically sort out all different contacts that you will never call at that time so to make or to improve the recognizing commands.

[22:32] So that Siri also knows about the situation and the context around the person?

[22:40] Yeah, it does not very good now, but it's getting improved and this is still some room for improvement in this direction.

[22:51] Mhm. I think I covered all my questions, is there anything that comes to your mind that has to be mentioned in that case of speaking of authenticity related to agents?

[23:03] Yeah, one point, but I think that would bring us to far away. But still it has a connection so at least in my opinion, when I am talking about this automated driving topic, is the point of ethics. We have addressed it very briefly in the beginning, maybe it also belongs somehow to agents, to automated agents. Or it must belong, must be somehow inherited for also, for this automated systems. Namely, how the behavior in specific situations, how they make decisions. So you know its addressed in newspaper and articles on a daily basis as you said before. What we people do is something that is always predictable, but machines or agents have some rule-base and they at least don't behave on their own, I don't know, implemented brain. Should at least programmed to correspond with sort of ethics. But it's not easy and sometimes you cannot have a rule to choose the correct action. Because it depends on the spur of the moment and facts that cannot be detected by sensors. This is something that is really problematic. Which people I will kill in a traffic, in a hazardous traffic situation. Is it more likely that I will kill senior persons because they have already done, because they have not so much worth for our society than maybe female persons with childs or are there groups of people with high income or higher IQ more worth than people with lower IQ, lower income or people with, I don't know, people with important positions in the society and more valuable to save compared to other people. So which is the rule, or which action will the action finally take and based on what. I don't know and I have no solution for that.

[25:43] Do you think, would it help if the agent wouldn't be a black box?

[25:50] Yeah, but still you would have to implement somehow how it behaves, you respect how it looks or if it has a human body shape. But still it behaves like a black box, it doesn't make a difference.

[26:00] With black box I mean it's not obvious how the algorithms come to a solution?

[26:14] Yes, but still, I don't think there is a way of communicating how it's working. It's not help.

[26:33] So even transparency in the decision making would not help to overcome the ethics problem in that case.

[26:40] No, I don't think so. Because, it would need a very long document. I don't think it is possible to quantify. It must be, there is some uncertainty and this cannot be quantified. I have no idea how to make the decisions and what is the value of different entities in a critical situation and which action to take. To safe the live of my owner, above

all other or is it just the maximum number of people that are heart in an accident or the economic value of an accident what decides what finally to do. So systems have much better capabilities to calculate or to look at the different options. The decision that is finally taken, this is independent from that. So you have some value system and you decide or calculate all possible options based on the value system, but I am not sure if this gives you then, makes it easier to decide for a certain option. But there is still, you cannot value one person and another person and say which one is better or which to safe.

[28:14] I think that is going to be a very interesting field and there is lot to find out about how we go to solve all this upcoming problems. I have a small anecdote to that, Microsoft put one bot into Twitter and the goal of this bot was to become a nine-teen-year-old-girl. You maybe have heard of it, its name was Tay, T A Y, and it became an anti-feminist and a Nazi by learning all that stuff that nine-teen-year-olds would read on Twitter. So I think that some kind of relates to the ethical problems, when we don't just have rule-based system but have self-learning algorithms or systems.

[28:14] Yes, maybe it could be one option somehow to learn by somehow just adopting behavior of people and looking at behavior of accident statistics and in this case and try to find out how people behaved and for what reason. This could be sort of an initial rule base, but yeah. You don't know if people, or if you actually should trust or should adopt the behavior of people without knowing what they are thinking and why they behave. It will be something that is definitely not easy, this may be the final point here, it is nothing that gives us an easy solution.

[30:07] Okay, great. Thank you for your time Dr. Riener. If you have anything to add or if you know somebody that should take part in this interviews, because his research fields are going in that direction. I would appreciate if you would let me know. Other than that, if you want I will send you the results of the Master thesis, if you are interested.

[30:33] Yes I am interested, this was what I wanted to ask you and maybe so I am just heading … to we have a seminar whole week on exactly that topic. Automation, trust in automation, user interfaces in the age of automation, ethics in the automation. So I will meet thirty great people from all over the world, so there might be the one or other, I don't know, maybe I …

[31:05] Yeah, if you feel so just let me know and I will contact them.

[31:12] I will maybe try or let’s see what’s going on. I may ask the one or other person if they like to take part.

[31:18] Okay, great. Thanks again and I hope you have a pleasant Friday and a good weekend.

[31:26] Okay thanks, by when will you try to finish your thesis by the way.

[31:30] I will finish it by 22nd of August, so that’s the date I have to hand in and then we have the Defensio in September but I am not aware of that date yet.

[31:41] Yeah, would be great if you could get a copy of that.

[31:45] Okay, we will do that. Thank you very much and bye-bye

[31:50] Bye-bye.

P11

Skype, Duration: 19:25 minutes

[00:00] So, Alex brought us together and he said you are like the expert in chat bots and human-computer-interaction.

[00:11] I don't know if I go that far but it is what I do my research.

[00:13] That's very interesting, I think you got my invitation email. So, that's also where I do my research on. I am trying to build social intelligent artificial agents. To bring this dimension of social intelligence which is comprised of context-awareness, presence and one part is authenticity into agents or bots.

[00:47] Ok.

[00:49] It would be great if you have the time to talk to me, your participate is voluntary, you free refuse and withdraw and I will just record this session and anonymize anything afterwards when I transcribe it, so no identifying characteristics will pop up. They can't blame you that you said something bad.

[01:14] Right, ok. Good.

[01:19] If you are interested I also can send you the results in the end and let you know what I found out. Why don't we just start and you give a brief introduction of yourself and how you relate to the what I am doing with social intelligence and artificial intelligence. And we can go on, I have prepared some questions.

[01:43] Yeah, so my research it's been mostly on the human side. So I look at the way that chat bots, when people interact with them, they way characteristics of the chat bots affect human behavior, human perceptions, whether they think it's human or not depending on how I built the system, and then how … how also effects, so I have looked into a few things: Perceptions from survey measures, deception I have looked a lying, lying behavior, social desirability, basically impression management how much do I change the way I portray myself depending on the bot that I am interacting with. Then also, sorry I get a lot of messages here – nothing important. So I perceptions, behavior, and then also typing, I have analyzed keystroke behavior using some JavaScript key

logging to look at keystroke, the way people type, when they are lying versus when they are telling the truth in different kinds of bots that they are interaction.

[03:12] So you already have some knowledge about bots, I suppose? That's one of the questions I have there. And you can imagine an agent operating in a messaging environment?

[02:24] Right, so that's what I've built. So I use the program called Chatscript. So there is, I don't know how far into you have gone, there is the AI-markup language – AIML, it's used for a lot of chat bots it's XML based. And there is, a lot of chat bots use that. And there is also Chatscript with is developed by a guy out in California, and he competes every year in the Loebner Prize Competition to win the Turin test. He's almost won, I mean nobodies beaten it yet, nobodies beaten the Turing test in its competition, but he is one, the top prize won that a few times using this ChatScript language.

[04:18] Do you know his name? Is it Bruce Wilcox?

[04:20] Bruce Wilcox yeah.

[04:21] I talked to his wife and I wrote with them also talking about this topic. So I also interviewed them as well.

[04:29] Ok, yeah. So I use Bruce's program, or his language and his application to build and I built a website that people can chat with my bots through this website.

[04:46] Cool, so how do you imagine this kind of perfect conversation with an agent? I would say some kind of a Turing-proof conversation with an agent.

[05:01] Oh well, that's probably the biggest thing that bots are missing right now is like you said context-awareness. So the ability to know the current topic of conversation and then seamlessly move between topics of conversation.

[05:29] Right, so having a history of what was said?

[05:31] Yeah, so knowing what we are talking about now, but then when the person you are talking to wants to change topics, to recognize we are not talking about this anymore. Now we are talking about this. This is our new topic. But then also, like yeah, you got to remember what you were talking about so that if they decide to mention something from before, so that you can move back to that.

[05:58] Do you think that's something that makes a bot authentic as well? To have this some kind of memory of the history?

[06:05] Authentic, what do you mean by authentic?

[06:08] Authentic defined by literature is some kind of genuine, being trustful or trustworthy, having a focus on communalities, knowing the context, having objectives and a mission statement. So this all relates to, according to theory out of brand management, psychology and philosophy. So do you think that these parts play a role as well when we talk about agents?

[07:00] I think it will absolutely influence people's perception of authenticity. I think when people when think of a chat bot, when they interact with a chat bot, it's all kind of lumped together into one. So is this, a human-like bot and authenticity and genuineness and trustworthiness and all of that, I think all kind of grouped together in people's minds.

[07:16] Do you think there is some kind of, you are also looked into trust, is there some kind of a process how an agent can gain credibility. Is there some kind of a way where it has to walk on or go through?

[07:32] Well trust, credibility in terms of what. I think that depends on what the bot is for. So if it is pure conversation, then we don't necessarily need to trust each other. Right? And as long as you don't say anything that makes me think that you a secretly evil, then we will trust each other enough to carry on the conversation. It gets different when the bot is maybe asking for confidential or personal, deeply personal information.

[08:12] Do you also think that bots some kind of have to represent their developers as well as showcasing their mission or objectives they are following?

[08:30] Nah, I don't think they represent their developers. They can, and that's actually something that I am interested in researching is. Is situations, there are some situations I think where a bot, admitting that it's a bot and acting like Siri, Siri when your phone doesn't pretend that she's human. Right, and there is no illusion that you are interaction with a human, when you are interaction with Siri. But then there is other situations, where it might be the purpose of the bot to deceive them, like in the Turing test, or other real life situations where it's the purpose of the bot to make people think it's a human and to pass itself off as a human. And so then you don't want to admit may be different kinds of behaviors might be useful. And I haven't studied that at all, but it's something that I am interest in researching.

[09:30] And thinking about going back to authenticity a little bit, when you relate it to authentic people. What we've talked about now does this also relate to people and machines equally or similarly. So has a bot have to have the same traits as an authentic person.

[09:57] Yeah, I don't know.

[10:00] What do you think? How would you make it, how you would build it?

[10:07] How would I built an authentic bot? No clue.

[10:20] When you think of a person, what makes it authentic to you?

[10:27] Yeah, I mean when I think about an authentic person. I mean like you said in your definition and it's not being deceptive, not trying to know ulterior motive of its own, it is not trying to manipulate me, he or she a person, is not trying to manipulate me or is trying to get something out of me, at least not sneakily.

[11:14] Would that also apply for a bot?

[11:17] What's that?

[11:22] Would that also apply for a bot then, so if it was sneaky getting your information.

[11:24] I would assume it would, if it trying to use manipulative tactics. Or looked like using emotional manipulation any kind of psychological manipulation to get what it wanted from me. Then, yeah, I would see that as less authentic.

[11:48] So I guess you have seen Ex Machina?

[11:51] I haven't!

[11:55] You haven't? You should. So there is a bot, they do the reverse Turing test, and the machine is tricking the starring actor into believing her. So may questions would have been is that then unauthentic? But I can ask that, what would be unauthentic behavior for you for a human as well as a bot?

[12:19] Yeah, I mean any kind of deception, any kind of extra-motive of any type where the bot is trying to hide its true nature. In order to get something out of the user I would say would be unauthentic behavior.

[12:43] Coming back to bots, if you imagine you have your personal assistant that is following on your phone, your computer, your mirrors, on your IoT. How would you

want it to interact with you, how would you want it to be in terms of how it behaves, it communicates? How it is intelligent?

[13:08] The important thing for that is it's responsive I think, that it is able to do whatever I ask it to do. And some amount of learning, so my phone for example knows where I work and where I live and can tell me automatically how long it takes to get there. It would say if there is traffic, you should go a different way and all of that could be automatic and that's proactive without asking for those directions. It is actively providing this information to me. For the most part I think it's just being available, where or when I want to talk to it or give it some instruction or request for information and then be able to interact with as many of the services that I use or apps or whatever, whatever else I use wherever else I get my information ideally would be able to pull that information for me?

[14:30] Is there some kind of a relation to being socially intelligent? I mean it knows the context, yes, do you see any other …

[14:40] I mean, I think there are applications for that, that's not something that I can see for me something personally. There probably are, maybe I want to think I do and I think there are definitely people that would want a more emotional agent, that is picking up on their emotional state and pro-active providing them funny cat videos when they are feeling done or something along those line. I don't know, I think there is applications for that for sue.

[15:23] But it's not directly implement into your personal ecosystem in that way. Good, is there something that I might have overseen that you think of relates to authenticity in chat bots?

[15:41] Well I think, probably the big thing that needs to be addressed is what's the purpose of the chat bot. Authenticity of the chat bots depends so much on what it's for. Is Siri authentic? Does it matter if Siri is authentic? I mean does Siri have any ulterior motive at all and is there any purpose, would there be any value in her having soother purpose? I don't know. And then you look at a brand for example. That's a big push right now, they are all over the places. Brands building up their chat bots for who-knows-what-reason. Obviously they have their own purpose and so there is probably differences in authenticity based on the application of the chat bot – you what it's for.

[16:55] I could imagine that a brand chat bot would also include the heritage or the values of the company, right?

[17:05] Right, potentially.

[17:06] That it knows how to deal with customer or service or that.

[17:12] Yeah, if it is well built. [laughs.]

[17:17] Yes, if it is well built, right and that's actually my intention in you brought it in its direction because in the end what we see is the application of these chat bots for brands. And we know brand authenticity is but we, you can't really relate it to technology. I mean this is a very philosophical approach. However, deem it as important because developers or research has to acknowledge that these traits or these social intelligent traits also reflects some kind of the business or the values or dimensions important for the companies. Like brand authenticity and authenticity of the brand itself that needs go into development of a machine that is designed for that.

[18:17] Right, yeah, machine that is representing the organization essentially.

[18:25] Ok. We are already through. I thank you very much for your time. If anything comes through your mind when you think about this topic, just let me know.

[18:41] Ok.

[18:42] And the same thing as I did with Alex, I would ask you if you know somebody who should be considered for this study. Just let me know who that is and I will approach them.

[18:54] Sounds good.

[18:57] Because I think I have a really nice group of experts and I really get very interesting insights in the field of authentic chat bots.

[19:07] Yeah. I am very interested in seeing what you've come up with.

[19:12] Sure, great, thank you very much and have a great day. Bye-bye.

[19:12] Well, it was good talking with you. Take care, bye.

P12

Skype, Duration: 22:12 minutes

[00:11] Hello Mario.

[00:15] Hello Professor [redacted], Great to have to connected to you.

[00:20] Nice to talk to you. Let me see if can set us up so at least I can look directly at you.

[00:22] That's fine. Hi.

[00:27] Hi. It's Wallach by the way, it's the way I pronounce it.

[00:33] Wallach, okay. Thank you for that.

[00:35] How can I be of help to you?

[00:38] Well, I do a research on authenticity in messaging agents. And I saw that you are also researching the same direction, so looking on socio-ethics of emerging technologies. That by definition, messaging agents are upcoming and economy says that they are the new wave, the new platform that technology applied on mobile phones and etc. And what I am doing is looking at the interaction of humans with those agents and trying to make the agents authentic and thereby my work, my Master's thesis tries to find characteristics of those messaging agents.

[01:38] So when you are talking about messaging agents you mean what? What do you include in your use of the term messaging agents?

[01:43] I include all kinds of artificial agents that are hosted on for example What's app. This could also be Siri when she writes with you on the iPhone, when she interacts with you on the iPhone. This can be Amazon Echo. So literally any bot that is showcasing a brand and working as a customer service representative. That's what I define as a messaging agent.

[02:17] So there's a question. The challenge of authenticity, the main challenge is the challenge of whether authenticity is contributing to ease of use, questioning the manufacturers perspective for higher revenues. And when the authenticity is basically facilitating deception, misrepresentation, in effect duping customers into participating in

activities they shouldn't otherwise. So it seems to me that the number one issue is the set of very high ethical standard, trust standard for authenticity and human-agents. And I would go so far as to say it would be nice that if there were standard settings that there is actually a certifying body. So that very trustworthy companies or other using messaging agents were able to say that our messaging agents have this good house-keeping proof or in terms of the way they act. That would be a quick signal to customer that they are dealing with an agent that is trustworthy. I am very against; I am not particularly interesting in messaging agents that try to imitate humans to closely. I think that authenticity is not necessarily about imitation and I can also be the key to be use of these agents for deceptive practices.

[04:09] Right.

[04:10] So in social robotics it has been recognized that very big eyes and child-like features facilitate people's positive affect toward the social robots. And I don't have a problem with facilitating positive response but oftentimes these are awfully cartoonish which may actually be a contribution. Particularly in terms of, there is issues that have been raised for whether this is truly an uncanny valley makes people uncomfortable with agents that are too closely reflecting humans. Indeed, agents that closely imitate humans could actually be a sign of deceptive practices going on. Because these are not humans, they pretend to be human. I don't try to say that everything is a Naysayer or somebody who's saying you don't want affective, authentic agents. I am actually doing it the other way round and basically saying that authenticity itself may require a degree of trust within itself and that may also entail the creation of agents that maintain an awareness in the user that these artificial agents facilitating communication as opposed to near human-like entities that potent to have capabilities that they certainly do not have.

[05:48] Okay. So…

[05:50] I am not sure if you captured my assumption message.

[05:55] Well, I do, I do. And I totally agree to what you say and also I interviewed other researchers on that and what they say is pretty much the same that you just told me. Do you think that. I am aware of the uncanny valley problem.

[06:25] It is not important how important the uncanny valley problem is because it's also a representation of a time in history. So the point is, if the uncanny valley is representing a spookiness in terms of the recognition of imitation like agents not being human and that

creates a discomfort, I am all for that. The uncanny valley may also just present a sense that we are unfamiliar with these entities. So for example there was a movie Midnight Express I believe it's called, which is a cartoonish Christmas movie which Tom Hanks plays, well there is a Tom Hanks-like figure with his voice. Everybody felt pretty uncomfortable with this movie when it came out. Because the eyes are hollowed in the creature. And in one level that may be intentional and signaling that these are not humans, on the other hand they try to replicate human features pretty closely in a cartoon. So I think at first there was a discomfort and some people traced that to the uncanny valley. But as that's been played season after season people are now starting to treat it like a classic, at least in America people see it as another classic film. So I think perhaps, that aspect of the uncanny may disappear as long as there is a clear signaling as in the movie that these are not humans. And therefore we just get familiar with imitations that are clearly not humans and that familiarity breeds a degree of comfort or at least not looking for a feature, feature-closeness as being in the termination of our degree of comfort with agents. But we ought to see how this progress is. So that's very different that Hiroshi Ishiguros' attempt to capture human-like features as closely as possible. That's a very different process I think. That's a very interesting process for a scientific purpose and I am quite clear whether, whether in the long run that doesn't hold out dangers in terms of agent that look fairly humans and might be mistaken for humans but aren't humans. And if that breeds an uncanny valley, so be it.

[09:25] Yes, do you think that they anyhow have to have certain values if they represent a brand or certain standards, certain missions they have to follow if they work for a brand or working towards a certain purpose.

[09:45] I think they should, I think they should make it very clear that this is not an authentic human being, but there's a difficulty in that. In America we have somebody that is representing Colonel Sanders. Colonel Sanders for the Kentucky Fried Chicken brand. This is not an original Colonel Sanders. This is, you know, this is a marketing representative of the company. So we do this often in advertising. Generally, this is seen as acceptable in terms of trying to create a comfort about the brand. So, I think this is a saddle problem when you deal with artificial agents. I don't think the problem of, it's also a problem with Colonel Sanders whether the authenticity is authentic. And I think that's the same problem. So the representation of authenticity, you know, in a branding thing you know as long as it is made clear that this is the symbol of our product. Because we

already have cartoons, symbol of products. I am not so worried about that, I am more worried about the representation of artificial, artificial agents as being something more than what they are.

[11:35] Okay, yeah.

[11:37] So that's a good one, if you want hear something more friendly to it, fine. You know you get that endorphin hit. We humans in the modern age have to learn that every endorphin hit does not mean that something is good. People in robotics and advertising in general have learned ways to push our positive affective buttons. But the pushing of our affecting buttons is not the same as the representation, then what's been representing by having those pushed buttons being authentic, credible, trusty. So on one level we need to go through to a social discipline around that in the modern age. The other level I would like to see companies that want to use artificial entities taking the lead in not just pushing buttons to doof people into being addicted in their product. Will they? No they won't. You know so then, what do we as a society, what do we as a society do to at least temper the manipulation among the more mighty among us.

[13:00] That's very interesting taking about this buttons. Do you think, not just talking about the representation, but also about the conversation, there we have these Darwinian buttons if you are familiar with that concept or meta-verbal-cues to facilitate communication and to show communication strategies. Do you think those play into manipulation as well to be unauthentic then?

[13:29] Of course, of course and effect the life is speeding up. People want heuristics to make decision, they want quick computers to make decision because we don't have time to evaluate anything. So, that's where I wouldn't mind some standards set and a body that would give the use of artificial entities a house-keeping-seal of approval. Some are saying that they are perfect and that they are not pushing buttons. But it's not just manipulation for manipulating sake. It's the same things as corruption in people dooping the elderly and investing money. I mean we can do the same thing in public life. Now whether that should become a requirement for people in particular industries to get the good house-keeping-seal of approval that's a difficult challenge. And I think we are all a little bit uncomfortable with too much bureaucracy, but if this could be well established and people would at least know, that they are not only looking at an agent, but that that agent has at least a body that is looking out for the consumers' interest, customers interest, has said that this at least adhere at certain standards.

[15:02] I think that's a really nice idea. How do you see those agents learning all that behavior and how to become authentic? Is it implied, because then we have to get up with all this value systems and cultural models. Right?

[13:29] Well, the problem I think is, what is authentic? For me an authentic agent is not one that necessarily does what a human does but it is authentic to what it is, what is purpose is. So if authenticity means, appearing to be real in a way that accuses the human being. That's ok for humans. Because that's part of our social policy, I think that would be inappropriate in social agents. So it raises this question about what should social agents be learning? And, you know if they are better at dooping the customer because they got some you know deep-learning algorithm, that helps them you know parse all the information coming from the customer so know they can push buttons. So that's a good marketing, that may be a good marketing scheme. But I would question whether that's manipulation rather than authenticity. In other words, I am trying to take a definition of authenticity which is not about success in confusing or dooping customers. Authenticity is the authentic representation of the underlying value. If I am just basically the computer learning bot, you know, then authenticity is not pretending I am something other than that.

[17:20] Mhm, that's a good way to say that. Well, actually you took all the questions I wrote down and answered them by going along. And that's pretty awesome I think. Is there anything that comes to your mind that still relates.

[17:40] Who else are you talking with?

[17:43] I talked to, a Professor in Australia her name is Katina Michael, she is also working on socio-ethics and emerging technologies. I am talking, I talked to the winners of the Loebner Prize last year. Bruce and Sue Wilcox. I talked to the director of the cybersecurity department in Louisville.

[18:16] So let me ask you. Have you talked to [redacted].

[18:20] No I haven't, that would be one questions if you have somebody who should be interview.

[18:28] So [redacted] is Miss social robots. [redacted] she's at MIT. [redacted] is at Yale. So these are two people creating social robotics. [redacted] created a totally new robot that is about to be marketed at, that she tried to raise a 150.000 Dollars at one of these sites and raised of a million. But she is the best-known person in the world of social robots with a lot of that was created between her and [redacted]. I would

talk in the UK to [redacted], and probably [redacted]. I would if you can get hold of him at Carnegie Mellon University I would talk to [redacted], he is also an ethnicity, a roboticist, he talks a lot about authenticity. I could go on, but I think this gives you some key names.

[19:50] Yeah, they are very helpful to me and the research. I have one last question, what made you participate or what made you talk to you. Is the topic of that much interest, was it my nice email?

[20:03] So I try just try to respond to people if I can. You know if I think somebody is credible and I can be of help then I try to respond. You know my time is pretty difficult these days but you know and I have been, I appreciate your patience. You wanted to talk to me a few weeks ago and I now you trying to finish up at least this first level of research. I don't think you have to talk to these people but I think you have a substantial of key people who are raising various issues here. You know I try to be of help if I can, particularly if I think the research or reporter is serious and is trying to help people understand these issues better and not just hyping silliness.

[21:03] Okay, so that's good. Great. So, thank you very much for your time. I hope we can stay in contact. If you want, I can make the results of the study available to you.

[18:28] Sure, if you do it in a way that adds a summary like a two or three-page summary

[21:22] I will maybe do a Medium post or something like this.

[21:24] Yeah I mean, whatever you do, you know I am so overwhelmed with stuff I am not reading, it's frightening. But that's fine. So good luck to you in your research. I am our paths will cross, particularly if you stay in this field. You know I am everywhere.

[21:45] Thank you very much. Thank you for your time and I hope to meet you again someday. Okay. Thank you Professor [redacted]. Have a great day.

[21:59] You also, Bye.

[22:01] Bye.