LIDILE Linguistique Ingénierie et Didactique des Langues

# Lundi LIDILE avec Dr Bahia GUELLAÏ


Dans le cadre du cycle Les Lundis de LIDILE, Bahia Guellaï, Psychologue du développement, professeure des universités en psychologie du développement (à l'Université de Toulouse 2) et formatrice en EAJE rattachée au Laboratoire Cognition, Langues, Langage, Ergonomie (CLLE) nous parlera de comment les avancées technologiques affectent nos attentes sociales et morales envers l'intelligence artificielle. L'objectif de cette présentation est de présenter des études dont les résultats soulèvent des questions importantes sur le rôle de la moralité dans nos interactions quotidiennes avec l'IA.

## Transcription textuelle de la vidéo

Merci Christine de m'avoir conviée à présenter une partie de mes travaux aujourd'hui. Donc je vais, comme je disais, switcher en anglais pour la présentation. Donc, today, I wanted to present you some work I did, but not alone, of course, with some of my colleagues, some of them are research assistants and some are professors either in France or in Singapore. Because, as you will see, a part of the study I will present today was conducted in Singapore. So I entitled this presentation, exploring how individuals perceive artificial agents in social contexts across disciplinary approach. Because, in fact, I will present you mainly two studies that I did. So the first study you will see corresponds to a set of experimental work I did in France. And the second study corresponds to a more, let's say, exploratory work I started in Singapore. So across these two studies, I was really interested in how individuals, and I started actually with young individuals, as you will see, perceive and may interact with a different type of artificial agents. So I was interested in this question because, as you all know, lately, artificial intelligence has been of a wide interest in different fields. And what is surprising is that there is not much studies that cross different disciplines. So here you will see I will cross disciplines in psychology because it's mainly my background, especially developmental psychology, but also disciplines such as esology and disciplines such as social psychology or philosophy. So one definition that we can give of artificial intelligence is that it is a new generation of technologies that enable humans to interact with in a variety of environments and which can simulate, in fact, human intelligence. So the success of integrating AI into organizations and daily life depends, in fact, of how we humans or the users in general do trust, in fact, these new type of technologies. So I was really interested in some aspect of this human-AI interaction that align also with a notion that is really important in AI at the moment, which is anthropomorphism. So this is the attribution of human-like different aspects such as emotional, cognitive, social cues to non-human systems. And this interest in this concept has grown rapidly lately in different fields such as academia, industry and popular media. And this is really evident in different type of AI lately such as conversational agents, but also social robots. And you will see that the first study actually I ran with different colleagues, many looked at social robots. So it is not only these aspects that are important actually in the notion of anthropomorphism. Another important aspect is the appearance, in fact, of the device, of these new type of technologies. So one roboticist, Japanese Masahiro Mori, in the 70s was one of the first to theorize that the possible human reaction to non-living agents such as robots really depends on the appearance of the robot. So actually more sympathy from the human arise depending on the aspect actually of this type of agent, of the robot. But what he noticed was that, in fact, the resemblance is a tricky aspect because if it's very high but not perfect, then the robot can cause repulsion. And if it's too dissimilar, it can also cause repulsion, which can lead, of course, in some lack of trust in this type of technology. So this is what he called the strange valley effect. So here, what was interesting, I thought, in the work I will present, was that the fact that humans, so as you will see, different, let's say, type of individuals may evaluate or see how they can interact with this type of technology. So AI technologies may, of course, depend on the appearance of these technologies, but also on other aspects that are still linked to the concept of anthropomorphism, which are important aspects in social interactions such as empathy, which is often defined as a pro-social emotion that we share with other living, such as animals. So non-human primate, but also other mammals, and that are usually associated with engagement in pro-social behaviors. And another aspect, which is of interest when we look at this human and AI systems interaction, is also an important aspect in social interaction or pro-sociality, which is altruism. So altruism is considered as a motivational state that leads individuals to engage in acts that are aimed to benefit others. So these two main aspects of social interaction actually emerged very early on in humans. So some authors in the literature showed that, for example, from three months of age, infants already show a certain ability to consider others as pro-social agents or non-pro-social agents. And this, of course, this capacity develops through the developmental stages of human development. So as we actually attribute some of these anthropomorphism cues in, let's say, artificial agents that are linked to these important aspects of social interaction that are, for example, empathy or altruism, other aspects such as emotion, agency, or even moral standing are very, very important. And lately, there are not many studies in the literature that explore actually these new forms of psychological reliance toward AI systems. So here I question the role of technology and in particular of social robots and virtual agents in inducing or mimicking, let's say, pro-social behaviors. So some studies showed, for example, that virtual agents and robots can generate positive relational outcomes among humans, but also among human and AI systems. So the first study I will present explores the use of social robots with young individuals. So in this study, you will see two species we explored, actually. So one is a human, of course, and another one is another animal, which is quite far from the humans, which are zebra finches. So the first question we wanted to raise was a very basic question. How children from three to six years do perceive potential social agents such as robots? What we explored was different aspects of the interaction. So just a moment, I'm a bit lost. It's just that my PPT doesn't answer. OK, so in the literature review we did, we found that so far there are some studies studying the interaction between children and robots, as you may now know. But back in 2022, when we started the study, it was quite new. And what we observed was mainly studies exploring interaction between atypical developing children, such as ASD children, so children with autism disorder, but not many with typically developing children. So what we wanted to see was quite simple, was just in front of an artificial or a biological agent, how children from three to six would react or would they even learn from this new type of agent? And what we were also interested in this topic was because we observed that more and more teachers in France, mainly, started to use this kind of agent in fostering new types of learning situations in classrooms. So we studied children interacting with one quite a famous social robot back then, which is a now robot. So this robot is quite easy to use because it's quite easy to program. And it is considered in the literature as a social robot that can interact with humans. And that can simulate actually human behaviors. So as I already said, it was of interest because of this anthropomorphic concept. So we developed the robot, the now robot that we call now Naomi, with colleagues from the ATIS lab in France. So we designed three main tasks. One task looked at the learning abilities of children when facing this type of agent compared to a human agent. So we called it the learning task. And especially we were looking at learning new words that were actually known words. And we looked also at the identification of body parts and identification of emotion. And the third task, we looked at spontaneous helping, which is an important aspect of altruism. So as we were

interested, of course, of the social interaction between the typical children and the robot. So in the vocabulary task, we presented this type of animal that we totally invented and the agent. So we tested 60 children and half of them were facing an artificial agent, so the now robot, and the other half were facing a human agent. So the agent was asking the children different questions concerning the three main domains we were interested in. So for the vocabulary task, the agent was asking, for example, the children what was the name of the animal, so in a previous, let's say, familiarization phase. The children were facing the agent and the agent was telling the children, look at this animal, it is, for example, a duck, or look at this animal, it is a grid, etc. a few times. And then they were presented with two animals, one familiar and one non-familiar, and the agent was asking, show me the duck, for example, or the grid. So in this first task, what we observed first was that young children, so the three years old children, were very afraid of the artificial agent, so of the now. In our final sample, we have mainly actually four and six years old. So for this first task, what we observed was that most of the children really succeeded in the task, so most of them, no matter if it was with an artificial agent or with a human agent, they had high performance score in naming the animal. We found an effect of age, and I will come back to this later, that is, older children were performing better than younger ones, which is not that surprising. So in the body task, we asked the children to show different parts of the body, so on his or her body and on the body of the agent, so artificial agent or human agent. And what we observed was a correct number score of body identification for the two types of agents, and again an effect of age. We also asked an imitation task, because in the literature so far what was observed was that this type of task was well succeeded by atypical children, and we wanted to see what about atypical children at this preschooler age. So the imitation task was mainly to imitate a series of gestures that the agent did and then ask the child to repeat this series of gestures. And again we found quite high scores, especially for the older children of course, but again no difference according to the type of agent, so they succeeded well either with an artificial agent or with a biological agent, human agent. We also asked questions about the recognition of emotion, so the idea was to see if the children could recognize emotion in the artificial agent, so in the robot. So the robot was depicting different, let's say, emotional postures with the whole body, the human agent for the human group was depicting the same emotional poses, and again we observed that in both groups scores were quite high for recognizing the basic six emotions that we presented to the children, both by the artificial agent and by the body. And finally we were very interested in the spontaneous helping situation because not many studies explored this type of situation with biological and artificial agents, so the idea was to see if in this social context young individuals, so children, would spontaneously help more an artificial agent that expresses the need of help compared to a human one. And actually this is the only task for which we found a difference between the artificial and the human agent, so the children were faster in giving back the object that the agent lost and was asking help to retrieve the object, so children were faster in the condition of artificial agent compared to the condition of human agent. So just to summarize a bit, so in this first area of study we were really interested in different types of tasks that we designed to see if typical children would react and learn in face-to-face situations with a robotic agent, artificial agent, versus a human one. And what was surprising was that between three and six years old we didn't find any difference of learning scores or prosocial behaviors between the two groups, we only found mainly an effect of age. So then we wanted to go a bit further and ask whether other type of individuals, so other species, that actually really need interaction to learn and you will see here to learn song, would learn better with a biological agent or no difference will occur such as we found for the children. So we chose this species because, so zebra finches, because in our lab we had the chance to have a colony of zebra finches, so it was at the University of Nanterre, a laboratory of ethology, cognition and development. So this species is interesting because we can observe the learning of song very early on in development, and here because we wanted to compare the situation with the children and the robot, so we saw that it would be a very interesting species to see if in another species other than a human, it would react. So with colleagues from the lab, so ethologists, we presented different situations, as you will see, where we put the same pattern of learning in different social contexts. So we chose a young individual, young zebra finch, with either a biological tutor, so another zebra finch, but older, an adult, or with an artificial one, so we designed a robotic zebra finch for this study. So just to come back to zebra finches and the way they learn songs. So in this species, there is a sensitive period where the young individuals, the males, because only males sing in this species, the male learns actually the song of the species. So what we wanted to see was if interaction, so let's say multimodal interaction, and contextual interaction, would be very interesting. So this is, again, a very important aspect that I said that we need to look at if we want to understand the foundation of early individual and AI system interaction. So here, a situation where the young was in contact with a tutor, either artificial or biological. So if there were, like, contingency between the two, let's say, individuals, so artificial and biological, would, and if there are no contingencies, would it impact actually the learning process of the song by the young individuals? So what I will present here are just the condition of the control condition. And the interactive contingent condition. And so later, we want to also test the non-contingent condition with the colleagues. So group A, group B, 12 males per group. So first, what we were interested in was to see if we could observe learning of the song by the young individuals with a biological tutor. So an adult zebra finch or an artificial tutor. So the robot zebra finch, which we call the mandabot. So mandabot was developed by colleagues from EASYR in Paris. And as you see, it's not perfect, let's say, if we can call it anthropomorphism cues, even if it could be applied here, of course, to birds. So maybe bird morphism, let's say. But it was really an attempt to see how the young zebra finches would react to this new type of agent. So there was a song coming from the robot, as you see here. And the robot was also able to move. So we could control the movement. So the idea was that it was presented either in a contingent situation where the movement were incoherent, let's say, with the behaviors of the young or non-contingent, so not coherent with the behavior. But I would present only the contingent condition here. So first, what we observed, and we were quite surprised, was that actually we could see high imitation scores of the song by the young individuals, the young zebra finches, when they were learning the song from an artificial agent, so the mandabot, compared to when they were learning the song from the male tutor, so the biological agent. So these results are quite interesting because they are quite similar in a way from those we found for the children. So no main differences concerning the imitation and so learning of the song when presented with an artificial tutor or a biological one in this species. And what was also interesting in the first analysis we did, was that the reaction observed, so it's hours and hours of observation by a PhD student who did these very nice studies, Alisa Raguas. So what she found was that the young, when they were in contact with the mandabot, seemed to be quite interested by this new type of agent, as they stayed still and seemed to listen to the agent, which are important situations to ensure the learning process of the song. So if I conclude this first part, we evidence that children from three to six, so typically developing children, when they are facing either an artificial agent or a biological agent, so a human agent in different tasks, they seem to be interested by both type of agent and there were no difference in the two groups in the performance scores in the different tasks we proposed. The only interesting difference we found was that they were faster in helping the robotic agent, so now, compared to the human one. And the other effect we found was mainly an age effect, with older children having higher scores, performance scores than the younger ones, which is not that surprising of course, but interesting. Concerning the bird, we also didn't find any difference between the artificial and biological agent, which mean that even in this quite far phylogenetically speaking species, we can find interest for an artificial agent such as the mandabot, and they learn also a new song from this type of agent. But where we are so far is that we don't exactly know, either for the children or for the bird, along which cues do they rely on, to consider these artificial agents as potentially interesting social agents. So we would like to go a bit further in this question. So far what we can see is that learning abilities in very young individuals and pro-social behaviors can appear when they are facing, let's say, an artificial agent, and that there is no main difference when we observe the results compared to when they are facing a biological agent. So in the second study I will present, we wanted to go a bit further in the question of prosociality and another aspect that is really important when we address, let's say, individual and artificial agent interaction, that is also moral abilities of the agent. So in this study I will present, we wanted to see this time if adults could evaluate and how they would perceive different types of artificial agents that are in a situation where they can help or support humans. If they can consider these new types of agents as having prosocial or moral abilities. So we embedded this question into the context of smart cities, which is linked to the project of Descartes that Christine was talking earlier about. So actually why it is a question of interest, this question of smart cities, and the notion of AI and especially social interaction with AI, is that because in the future more and more authors point to the fact that citizens will be more and more important, so the number of citizens will be more and more important in urban area. And this will lead of course to a progressive rethinking and new conceptualisation of the city. So in this context you may already have heard about the concept of smart city. So most of the different aspects of our everyday life, if we live in the city, will be in this smart city concept, rethought, let's say, according to the importance of artificial intelligence. This is mainly how today we link the concept of smart city, is how to address the urban issues and how to build a new area that could respond to the needs of citizens and using these new types of technologies that are the AI technologies. So in the 90s the term smart city was mostly referring to technological components and infrastructure of the urban area, but now more and more literature emerge concerning also the importance of having a human-centred approach when we design or when we think of how to design a smart city model. So for example it is how to give attention to social inclusion, how to promote also participation of all citizens, and when I say all I think of course of adults but also of children, of elderly, and how we can also support their social relations. And slowly slowly in the literature emerge also the notion of how to care for the citizen using for example the AI technologies. So far there is a debate around whether artificial agents, and especially in the context of smart city, can be reliable agents, and one aspect of the reliability is the morality or prosociality of the agent, and as you understood I'm quite interested in this prosocial or social aspect. And another aspect is also how and if this new type of agent can be designed to be moral. So here we were interested in these aspects, but because it was quite challenging to think of a way to have some data concerning how citizens in a smart city such as Singapore, because a study I will present to you was conducted in Singapore, so the idea was to see how potential citizens could evaluate different types of artificial agents they could encounter in the smart city context, and if they could evaluate them as being prosocial potential agent or eventually moral agent. So prosocial aspects are more linked to the action of the agent, and moral aspects are more linked to the value, so the decision behind the abilities of the agent. So quickly just I go back to the concept of

moral agency that we were interested in this study. So this concept is at the crossroad of different disciplines such as psychology, sociology, philosophy, and because the study we managed to propose, so it was a survey, was also sought by different colleagues from different disciplines. We wanted to look at this concept, but we very quickly found that it was difficult to give a coherent definition. So one definition I propose here is that moral agency refers to what the individual can conceive or can think of the ability in a different situation that elicits moral values of the agent, so the agent was, as you understood, either artificial or biological in the survey I will present. So this is of importance because we see more and more situations where, for example, elderly with a robot kind of experience new type of interaction with this new type of agent, and as I said earlier, there are a sense of course of anthropomorphisation of this agent, not only in the appearance but also in the feeling or also values we can think they can have in mimicking the interaction we have with humans. Here with this new type of agent, of course, there is a generalisation of this type of situation. So the experimental design, so it is still at the stage of preliminary, so we finished the studies, but we are still looking at the results, so it's still preliminary results. So the first study I will show you was a survey we proposed to adults, so young adults, as you see the mean age was 20 years old here in Singapore, and the idea was to present scenarios that we designed based on previous literature but also based on interviews we managed to have with different actors, such as drone developers or AI system developers, etc. So first the participants were having a short explanation of course of the study, and then we proposed them what we call the Perceived Moral Agency Scale, which was based in fact on the work of Banks and Collaborators in 2018, and so the idea was just to see if at T0, so before we proposed different types of scenarios to the participants involving different types of artificial agents as you will see, so this scale was of interest just to evaluate how they would perceive moral agency in different types of agents, so either drones, chatbots, what we call disembodied AI, social robots such as the NAO, and we use the human agent as a control. So we proposed this scale, the Perceived Moral Agency Scale at T0 and after the study, so at T1. So first they had this scale which was based on the Likert scale, and then they had a first task which was to evaluate the moral abilities of the different agents, so artificial agents and compared to human agents. So in different scenarios that involved interaction between the AI and a human, we asked questions such as do you think that the drone in this situation should do this or would do this, etc. We used different ways of asking the question because we wanted to see first if the participants could rate the autonomy abilities of the agent, which was the belief, the intent and the agency, still on the Likert scale. We wanted also to see if they could rate the action endorsement of the agent, so if the agent should act in a certain way or should not act, and the moral judgement of the different type of agent, so was it morally appropriate that the agent act in this way or not, etc. So one participant had one type of agent, so either a drone, either the social robot NAO, either a disembodied AI or a human. So the human condition was actually our control condition. So we proposed visual situations that we imagine that could happen in an urban context, especially in the context of a smart city. And in the first study, so the one I am presenting here, we associated these visual images, these visual stimuli with a short presentation of what was going on. So we called this first study the explicit condition and we ran also another study with still around 200 participants, so other participants. But this time we presented only the visual, but I don't have the results so far. So here in this first survey, a short description of the images were associated with the presentation of the visual stimuli. So what was interesting first in our early findings was that the perceived moral agency we proposed, so the scale of banks and collaborators, we proposed at T0 and then after they had all the different scenarios and questions, so at T1. Actually, we found that humans were always called higher on perceived moral agency than any of the other artificial agents at T0 and at T1. But we found a slight difference for the artificial agent where we found that the participants rated higher the three different artificial agents at T1 than at T0. So we mean that the different scenarios we proposed them might have changed in a way their perceived moral agency of these artificial agents. So basically in this first study, there were more expectations for the human agent concerning the different aspects, so autonomy, action endorsement, moral judgement, compared to the other artificial agents, so we were a bit disappointed in a way. But what was interesting was that in some situations, so when we looked a bit closer to the answers given by the participants in the different scenarios, we actually found that some scenarios where, for example, artificial agents were supposed to help the human, there were more expectations of helping, especially for the robot, compared to the other artificial agent. So robot and human were rated quite similarly in these helping situations. So quickly, I'm almost done, but we wanted also to compare these two surveys, so here I presented one, and the analyses are still ongoing because we have not only the quantitative data, but we have also the qualitative ones. So some of the questions were open questions, so this will be interesting to see also how the participants explain their reflection along the perceived moral agency of the agent. So we also proposed, and it's just finished, the same two surveys to French participants. And so quickly for the presentation of today, I wanted to see if there were differences. And actually what seems to appear in the results is that we found on the moral dimension some differences between the Singaporean participants and the French ones. That is, the Singaporean participants seem to perceive robots with having more moral abilities than the other artificial agents and quite similar with the human agent. And this result is different from the French participants. So Singaporeans seem to have to evaluate social robots as having more moral abilities than the French participants do for the same type of artificial agent. So there seem to be different expectations according to the cultural background of the participants, especially for robotic agents compared to the other artificial agents. Concerning the autonomy dimension, again, it seems that Singaporeans have a higher expectation concerning the robots' autonomy abilities than the French participants. So we thought that maybe there are more positive perceptions concerning robots among Singaporeans. And finally, what we observed when we looked closer at the different scenarios was that in some contexts that involve emergency situations or surveillance, it seems that Singaporean participants have stronger acceptance of assistance provided by a robotic agent. So again, among the different artificial agents, this is the one that has higher scores, higher rated scores and quite similar to the human one. So it seems that the Singaporean participants might have a stronger acceptance of assistance provided by a robotic agent compared to the other artificial agent. And that these findings suggest maybe a more pragmatic or utilitarian approach in the Singaporean cultural context. So to finish, what we observe is that in different social contexts, using different disciplinary approaches as you saw, it seems that human participants, when they are facing different types of agents, so artificial agents or human agents, they seem to perceive this new type of agent as having some similar maybe moral abilities. But we need to be careful because when we look closer at the different artificial agents, it seems that social robots are rated differently and closer to the human agent. But what would be interesting is to go a bit further in this type of analysis that could help us understanding how individuals perceive and could interact actually with this new type of agent still with an interdisciplinary approach. And so far, one of the work we are doing is that, as I said, quantitative and qualitative analysis of the four surveys in total are being done by also a colleague from the computer science side. And we want also to see if we could have comparison between answers given by the humans. So in this second study and other artificial intelligence systems such as LLM. So at the moment, we are also running the survey presented with different type of LLM. So just to conclude, I still have some questions. One of them is how we can think of developing agent that could use a proper channels and language to communicate in a way that conveys let's say trust and commitment from human. And if so, to what extent it could be beneficial for humans. And one population that I think is interesting is vulnerable population such as children or elderly. Because there are some numerous work on how artificial agent can improve well-being and care feeling for elderly, not much for children. But I think it's important to think of this vulnerable population if we want to develop tools and especially AI tools in the context of smart cities. And it is important also in the future research to take into account the role of culture because we see that it impacts evaluation and feeling about AI in general by human participants and let's say by citizens in general. And I'm still very interested on the prosociality and morality aspect of this new type of agent. To what extent they can be also useful to develop these very important skills that are prosociality, empathy, cooperation in humans. So I thank you for your attention. I hope I was not too long. I must say that I don't see any of you because I only see my presentation so it was a bit disturbing. But anyway now I will stop and I could see your faces. So just to finish I want also to thank my collaborators and especially the research assistant Aloysius Tok, Fernanda Mancilla who is helping us with the LLM part that we just started. And Alice Araguaz who did wonderful work with young individuals and robots. Thank you.