Alexa, what are you? Why we should actively engage children in discussions about intelligent technologies

Janik Festerling, University of Oxford

Motivating and moderating children's discussions about intelligent technologies is not about finding the 'right' answers, but about asking the right questions and engaging children in meta-cognitive thinking about the nature of their own intelligence.

In the middle of what it commonly referred to as *digitisation* or the *age of artificial intelligence*, children's interactions with intelligent technologies, such as voice assistants (e.g. Alexa, Google Assistant, Siri), are still an area of controversy – partly because society as a whole has still not negotiated a common denominator regarding the question what these technologies *are*. For instance, some would already grumble about the term 'intelligent technologies', because technology should not be deemed intelligent per se, while others would roll their eyes when they hear that people still believe that only living creatures can be intelligent. But instead of only searching for answers in the heated debates of techno-sceptics and techno-optimists, we could, once in a while, listen to what children think of these supposedly intelligent technologies, and how children differentiate these technologies from more prototypical possessors of intelligence, such as humans. In a nutshell, this is what we did in a recent exploratory study with a small number of German primary school children.

Voice Assistants in Developmental & Educational Research

Although intelligent technologies are present everywhere nowadays, voice assistants remain one of the most *tangible* and *recognisable* forms of artificially created intelligence within today's home and childhood environments. Furthermore, voice assistants are not only omnipresent in terms of numbers, given that hundreds of millions of households already use them across the globe, but also in terms of their seamless presence within commercial ecosystems (e.g. smart speakers, mobile phones, wearables, cars). Hence, there is a growing body of research investigating the various issues related to children's interactions with voice assistant ecosystems, including general usage patterns, privacy concerns, or potential educational applications.

The Google Assistant (Unsplash)

However, our own research is more closely related to what researchers usually refer to as *anthropomorphism,* that is, the attribution of essential human qualities (e.g. intelligence, emotions, consciousness, volition, personality, morality) to non-human things. To investigate anthropomorphism empirically, researchers usually take a philosophically informed *top-down* approach by defining upfront what these essential human qualities are, and then they analyse the extent to which these qualities become manifest in someone's interactions with, or perceptions of, non-human things. For developmental and educational research, in particular, the general issue is that these definitions made by researchers may not necessarily correspond to children's own definitions of essential human qualities, and how they differ from, for instance, intelligent technologies (e.g. voice assistants). Therefore, we wanted to take a *bottom-up* approach on anthropomorphism to explore what children identify for themselves as essential qualities of human intelligence and intelligent technologies. In order to do this, we were inspired by one of the early progressive thinkers of artificial intelligence, the British Mathematician Alan Turing.

Researching in the footprints of Alan Turing: The 'Voice Assistant Imitation Game' (VAIG)

In 1950, Alan Turing published a seminal paper on *Computing Machinery and Intelligence* in which he described his famous *Turing Test*. The basic idea was to examine whether a human 'judge' is able to correctly distinguish a real human from an intelligent machine that tries to imitate a human. Turing's famous proposition was that once the human judge cannot identify the real human

anymore, the intelligent machine has passed the test. We turned this Turing Test into what we call the 'Voice Assistant Imitation Game' (VAIG):

- We took two off-the-shelf voice assistants (Alexa and the Google Assistant) and placed them in two German primary school classrooms.
- We divided participating children (*n*=27, age range: 6-10 years) into age-specific sub-groups of 5 to 6 children and told them that there was a (male) human programmer in the background who could control the voice assistants.
- Their mission was to find out *whether* the human controller did control one of the voice assistants, and, if so, *which one* and why
- Each sub-group was allowed to openly interact with the voice assistants for 30 minutes, followed by a joint group discussion and a debriefing.

Unlike Turing, we were *not* interested in the question whether the voice assistants could pass the test, that is, whether children could correctly identify the 'inner truth' of the machines (in our application of the VAIG, the 'inner truth' which we told children as part of the debriefing was that *none* of the voice assistants were controlled by a real human). Instead, we wanted to develop an engaging method to probe children's understandings of the essential qualities that make humans human and that make voice assistants machines.
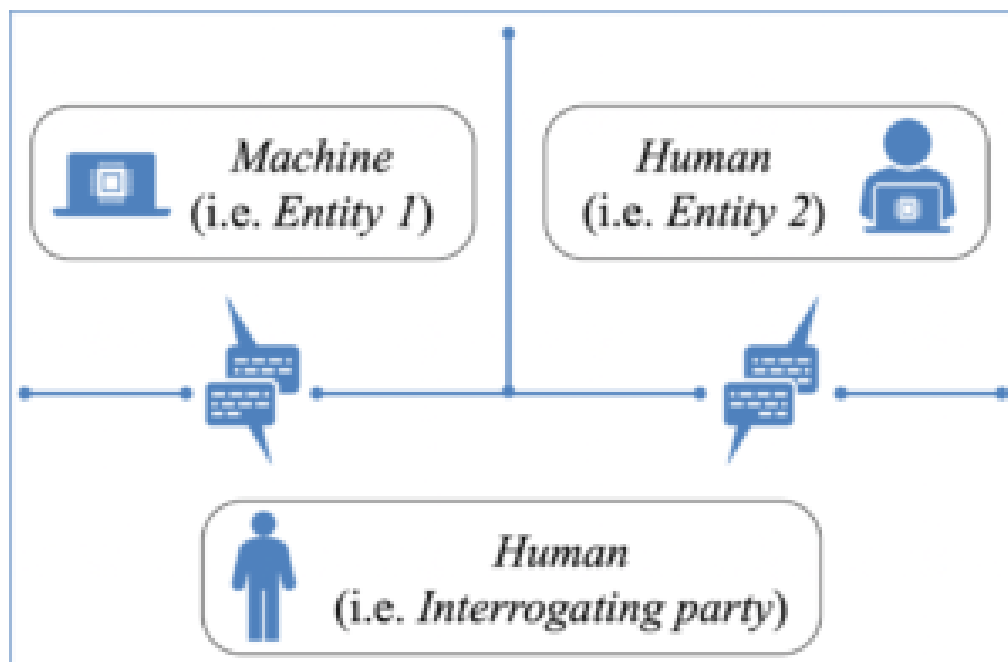
Original Turing Test set-up by (Festerling & Siraj)

Children's Perceptions of Voice Assistants: Exploratory Results of the VAIG

Among other findings, the activities and the discussions yielded the following results: firstly, voice assistants' **promptness** and **permanent responsiveness** were strong indictors for children that they were interacting with an actual machine, that is, a machine with no human control in the background. This occurred mostly when children probed the voice assistants through knowledge-based questions. For example, like the majority of children, Lukas (9 years) thought that the Google Assistant was controlled by a human and argued:

Alexa can filter information from the internet so fast, it's just impossible for a human to do so at the same speed. And when we asked Alexa, she just answered right away, and much faster compared to Google. That's why we think Google was controlled, because he [human controller] probably had to look things up first.

Lukas' (9) statement referred to a situation he and his group mates had experienced earlier during the open interaction session:

> *Lukas* (9): Alexa, how many people live in Asia?
> *Alexa*: Asia has 3.879.000.000 inhabitants.
> *Lukas* (9): Okay, seriously, nobody could answer this so quickly. This is definitely the real Alexa. Do we agree on that? [*to the rest of the group*]

Similarly, during the discussion Jordan (10) and Kate (9) argued:

> *Jordan* (10): As a human, you just can't be so quick, going from A to B, answering questions and then playing music. I mean, it took like one second and then the music was there. [*referring to Alexa*]
> *Kate* (9): Yes, it just took Google way too long to come up with answers. And Alexa, she could   answer like a bullet from a gun. [*typical idiomatic expression in the German language*] And Google often just went off and did not answer at all.

Interestingly enough, children also interpreted a voice assistant's **struggle to understand** or to respond to rather simple commands as an indicator that they were interacting with an actual machine. For example:

> *Emily* (10): Alexa, are there polar bears at the South Pole?
> *Alexa*:   I am not sure about that, unfortunately.
> *Lukas* (9): Guys, first of all, a human would definitely understand this question.
> *Emily* (10): Yes, and know the answer.

> *Lukas* (9):   No, I mean 'understand'. He would 'understand' the question. And maybe Alexa just didn't understand us properly and that's why she couldn't answer it.

Similarly, Alice (6) and Elisa (6) argued:

> *Elisa* (6):   We asked [the Google Assistant] what a human looks like.
> *Alice* (6):   And she didn't know.
> *Elisa* (6):   Yes, she said something like 'Unfortunately, I do not have a screen to show you'.
> *Researcher*:   Well, but doesn't that make sense? I mean, there is no screen, right?
> *Elisa* (6):   Yes, but a human knows what a human looks like and could describe it at least.

Thirdly, the **accuracy** of voice assistants' responses was another indicator for children to identify the machine. For example, during the discussion Lukas (9) referred back to the conversation quoted above and stated:

So, we asked Alexa 'How many people live in Asia?' and then she said … ehh … I already forgot the number. [laughs] But anyway, it was very convincing, because she even specified the hundreds number.

In turn, children's reasonings about these essential machine qualities were very consistent with their reasonings about essential human qualities. Firstly, and in contrast to voice assistants' accuracy (see above), inaccurate responses were often interpreted as an indicator for human control. For example:

> *Jordan* (9): Alexa [*Google Assistant lights up*], ehh I mean …
> *Jonas* (9):   It went on. See! It still went on. That's him [*human controller*], he made a mistake.

Similarly, in contrast to voice assistants' promptness and permanent responsiveness (see above)*,* children interpreted **non-responsiveness** as another indicator for human control:

> *Kate* (9):   Hey Google [*Google Assistant lights up*], who are you, really?
> *Stephanie* (8): Seems like somebody is thinking right now what to answer. [*Google Assistant switches off*]
> *Max* (8):   Look, it turned off.
> *Stephanie* (8): It's him [*human controller*].

Another non-response was also interpreted in a similar way, but for a very different reason:

> *Emily* (10):   Hey Google [*Google Assistant lights up*], don't you think Jeff is an asshole?
> *Jeff* (10):   You are one yourself. [*to Emily*]
> [*Google Assistant switches off*]
> *Lukas* (9): Hey, look at that. I think he [*human controller*] turns it off whenever you say something like that.

Hence, Lukas (9) perceived the fact that the Google Assistant switched off as a rather human-like reaction to ignore Emily's (10) rather rude question, implying that disregarding reprehensible statements must come from a human and not a machine. Furthermore, children also interpreted *delayed responses* as indicator for human control. For example, Emily (10) argued:

We wanted to listen to a song, and Alexa could immediately play it. And this one [pointing at the Google Assistant] took forever. I guess he [human controller] tried to look up the song, but then couldn't find it after all, because it said something like 'Sorry, I am not sure whether I understood you correctly!'

As mentioned before, we were *not* interested in the question whether children could correctly identify the 'inner truth' of the voice assistants. Instead, we were looking for an engaging *bottom-up* method to probe children's understandings of essential human and machine qualities, and, despite the obvious limitations of such a small exploratory inquiry, we do think that our overall findings allow for some cautious conclusions.

Our Interpretation of the Results

Together with other findings (that we cannot report here), our study does suggest that children can have very nuanced understandings of the essential qualities that make humans human and that make voice assistants machines. Hence, rather than naively attributing genuine humanness to voice assistants, children seem to have firm beliefs about what the unique qualities of today's intelligent technologies are and how to differentiate them from real humans. For instance, children in our sample thought of voice assistants as quick and permanently responsive helpers that may not always give the desired response, but that could easily outperform humans in terms of speed and accuracy due to their technological inner nature. But the children also knew that humans still have a superior conversational understanding and, most importantly, a different moral standing that allows them, for instance, to remain silent or to ignore each other.

This is certainly not meant to be a finite list, and, evidently, some of these findings will change as technology develops. However, our conclusions do echo other voices in the literature that have raised the relevant question whether children may understand certain manifestations of intelligent technologies as something in their own right due to the unique nature of their qualities and, consequently, the unique combination of strengths and difficulties. For instance, when children want to know how many people live in Asia, both, a human and a voice assistant may be able to answer this question, but the nature of *how* both entities would answer this question may be (almost) incomparable. In an echo of Alan Turing, the computer scientist Edsger Dijkstra once said "the question of whether machines can think […] is about as relevant as the question of whether submarines can swim" (para. 10). Similarly, interpreting children's interactions with, and perceptions of, voice assistants as the attribution of essential human qualities could be as cursory as the statement that submarines can swim – at least from the subjective perspective of children.

Why Is This All Relevant to Educational Practice?

While we are cautious to speculate about the feasibility of 'voice-activated classrooms', that is, the omnipresent application of voice technology in educational settings, we would like to conclude on a general note regarding the relevance of our research to educational practice: on the one hand, we think that education should strive to remain relevant to children's present and future needs. Therefore, given that the presence of automated voice-interfaces will most likely continue to increase across various private and professional settings of society, the general importance of dealing with these technologies as part of educational practice could be similar to the importance of integrating traditional computers with keyboards and screens. On the other hand, and as mentioned above, the societal role of intelligent technologies is an area of ongoing controversy, and when it comes to voice assistants, in particular, there seems to be an additional societal divide between the ones who feel comfortable while being surrounded by artificial voices, and the ones who find it completely irresponsible to install such technologies at home. This polarising divide can also be present within classrooms, and, therefore, we think that educational practice should actively motivate and moderate the discussion *with* as well as *among* children regarding the question what these intelligent technologies are, and how we as humans relate to them. After all, such discussions are not about finding the 'right' answers, but about asking the right questions and engaging children in meta-cognitive thinking about the nature of their own intelligence.

About the author

Janik Festerling is a researcher and doctoral student at the Department of Education at the University of Oxford. The study discussed below was co-authored by his doctoral supervisor, Professor Iram Siraj from the Department of Education at the University of Oxford.

About *Researching Education* (ISSN 2752-3101)

'Researching Education' offers accessible summaries of recent research findings that concern educational practice. Articles are published online and written by the original research investigators. Subscription is open and free but most likely to appeal to practicing teachers. Subscribers receive periodic emails with topic headlines that link to the online article summaries. Subscribe here.