Transcription

AIB Podcast Episode 5 - 'Humanised' AI

--Segment 1--

Rei Kurohi / moderator, Assistant Editor – The Straits Times

00.05-01.14

Welcome to episode 5 of the AIB podcast series. My name is Rei and I'm an Assistant Editor at The Straits Times on the newsroom strategy team. Part of my work involves working with generative AI to create tools, create workflows that improve productivity here in the newsroom at ST. Today we're talking about humanised AI, generative AI models that have taken on increasingly human-like characteristics.

The emergence of this technology which has driven the adoption of AI tools that go beyond productivity. For example, with AI companions that adopt photorealistic avatars, realistic voices and even so-called emotional intelligence. To discuss this topic we have three distinguished guests. Dr Gry Hasselbalch from DataEthics.eu, who is an expert in European regulation, digital rights, and human-centred AI governance. Ms Kristina Fong, a fellow in economics at the ASEAN Studies Centre at ISEAS Yusof Ishak Institute, and Assistant Professor Renwen Zhang from the Wee Kim Wee School of Communication and Information at NTU, where she researches human-AI interaction, emotional attachment, mental health and social consequences. Welcome to all of our speakers.

Development of human-like AI models

01.15-01.36

So I'd like to start this conversation with you, Professor Zhang. Can you take us through some of the major developments in this space? What are the characteristics of anthropomorphic models or what we call human-like AI models? How have they developed over the last few years? And what are the most significant aspects of this emerging technology that you see in your research?

Renwen Zhang (Asst Prof) / Director, SWEET Lab. Wee Kim Wee School of Communication and Information, NTU

01.37-04.01

Thanks, Rei. Thanks for having me.

It's a great pleasure being here. So yeah, in recent years, all of us have witnessed the rise of a new form of social interaction, where a lot of people are turning to AI companions for social and emotional support. So AI companions are the sort of conversational AI system that provide emotional responsive exchanges, simulating human relationships, giving and providing empathy and comfort and emotional support. So this kind of AI companion system typically includes both general-purpose large-language models, like Chat-GPT and Gemini, when they are used for emotional and social purposes, as well as specialised AI companion applications like Replica and Character AI, which are designed specifically for simulated relationships like

romantic relationship and friendship and mentors. And these AI companions, they provide highly accessible, empathetic and non-judgemental support, and leading many users to form emotional bonds and experience these systems as relational partners. So actually, we've seen studies showing that such human-AI relationships may help reduce loneliness and psychological distress to some extent, which also leads to the rise of AI companion market.

So there's, according to an *American Investment Management Firm*, that the engagement hours with AI companion are likely to surpass social media engagement hours in 2030[1]. And also, according to some of the research centre, they show that the global AI companion market is projected to reach 140 billion US dollars by 2030[2]. So basically, all of these are showing us the huge market and the prevalence of AI companion usage globally.

Oh, and by the way, I wanted to mention that according to the Harvard Business Review, that therapy and companionship is now the top one gen-AI use case in 2025 and it continues to be the top one use case in 2026. So yeah, it's happening.

Drivers of humanized AI

Rei Kurohi / moderator

04.02-04.40

That's very interesting that you mentioned this shift and the fact that it's actually become the top use case. Now, I think computer programmes that simulate human relationships or human interactions, they're not new. And even some of the apps that you mentioned, Replica, I believe, was on the market before this huge gen-AI boom. So do you see this as a gradual shift? Is this an all-at-once kind of quantum leap? And what do you think is driving this? Is it mainly the technical advancements that allowed the gen-AI models to behave more and more realistically? Or do you see it more as being shaped by organic user behaviour? And is the sort of responding to that demand?

Renwen Zhang (Asst Prof)

04.41-06.58

I definitely think this is not a gradual change. It was hugely driven by the rise of LLMs,

despite the presence of Replica being there before the rise of chat GPT.

And actually, I started studying Replica during the pandemic, which was pre-GPT 3.5 launch. And then what really motivated me to study is that during the pandemic, I've seen a lot of people talking about this really appealing AI chatbot that understands your emotion and responds to you in an empathetic way.

But back then, this is still a subculture, a very small group of people turning to AI.

And then right after 2023, when chat GPT became popular, we're seeing the rise of this human-AI relationships and emotional engagement with AI. So I definitely think that Chat GPT and the

[1] ARK Investment projects that AI companion engagement hours are projected to experience exponential growth by 2030.

[2] Grand View Research (2025) estimates the AI companion market will reach US$140.754 billion by 2030.

advancement in AI, Gen AI technologies, have actually fuelled the popularity of AI companionship.

So there's definitely technological push out there. And at the same time, the tech companies are not the only reason that's driving the AI companionship phenomenon, because we're seeing a loneliness epidemic before the rise of AI companions, people felt lonely, you know, especially the young people in the modern society, ~~right?~~ And it's not something new. And it's definitely the pandemic intensified this kind of loneliness epidemic.

And especially given the design features that we talked about just right now, that AI is designed to be empathetic, and non-judgemental and sycophantic, it creates a sort of really appealing feedback loop that keeps drawing the young, the lonely people back to this system. **So it's because of the combination of technological advancement versus the societal loneliness epidemic out there, we're seeing this huge rise and popularity of AI companion interactions**.

<u>Situation in Asia and Europe</u>

Rei Kurohi / moderator

06.59-07.28

And I'm sure this is really a global phenomenon, what you've described. Of course, loneliness epidemic is not new to us here in Asia and I'm sure not new to folks elsewhere as well. So I'd love to hear from Dr Gry and Ms Kristina Fong, you know, what's the view from Asia, from Southeast Asia and from Europe? How is this evolving social patterns really driving this adoption of humanised AI? Perhaps, Dr. Gry, what's your view from Europe?

Dr Gry Hasselbalch / DataEthics.eu

07.29-09.52

So from a European perspective, I ~~mean, normally~~, when we have been talking about and adopting and debating AI in Europe, the usual approach has been to talk about tools. So we basically, there's been a lot of kind of movement towards adopting AI as writing tools and as tools for companies and businesses and so on, so forth. But if we look at the field in terms of how the move towards using chatbots and AI companions, we do see a kind of rise of this use, which I would think is also very much embedded, of course, as the previous speaker was talking about the loneliness crisis and the pandemic and the kind of older social aspects that these companies are feeding into. But it's also a question of the very design of the models.

And from a governance perspective, the way it's been evolving in Europe, I would say that when we talk about AI companions, AI chatbots that are human-like, this very much speaks into the context of the way in which the EU has been dealing with AI technologies that have a kind of deceptive design or the dark patterns of social media, for example.

And so we see a lot of conversations about tech sovereignty, challenging these kind of deceptive patterns and focus on what we call, what was called also in the EU high level, we call it the human-centric approach to AI. And that could sound like building AI human-like, but actually it's the opposite. **It's about creating AI technologies with human agency at the center.**

So the idea is to create transparent models, models that are non-deceptive by design, where data is handled in certain ways, because that's also an aspect of this side is that people share very private information with these models. But in reality, what is behind that is a design that is based, it's a business model.

Rei Kurohi / moderator

09.53-10.22

It's interesting to hear that Europe, shall we say, has a bit of more of a sceptical attitude towards these technologies. It feels to me, anecdotally at least, that in Asia, many of the countries around here seem to embrace AI or to use it in a very different manner.

I wonder, Kristina, what do you think are the differences perhaps between the way that we here in Southeast Asia perceive and interact with AI versus how the Europeans or the Americans might do so?

Ms Kristina Fong / ASEAN Studies Centre fellow at ISEAS Yusof Ishak Institute

10.23-11.38

That's a really good question, because I think that maybe anecdotally, we may feel like there is a little bit more acceptance.

I would say that most of the research focusses on the impacts of humanised AI on society, like all the good work that Renwen does with her team. But rather than looking into the societal sentiments on humanised AI and how society really feels about humanised AI, I don't think there has been a lot of research in this area, and we definitely need more.

And maybe something to think about, perhaps, is also the impacts that it has on different parts of the socioeconomic demographic and with people with different socioeconomic backgrounds, and also maybe even the access to digital technologies. I think all these groups and segments will have different opinions or even access or interactions with humanised AI. And that really makes it a little bit more complicated, because from a public policy perspective, you need to really consider all stakeholders involved and their impacts. And it's just a very almost like a convoluted, multi-pronged kind of impact that it can have over the course of society. And I think that that really presents, I guess, a wicked problem in public policy.

--Segment 2--

Effects of AI Models on Society

Rei Kurohi / moderator

11.41-12.17

So this is an open question to anybody on the panel. What are you concerned about when it comes to the effects of these models, these humanised models, on the way that we interact with each other? We know that these models, you know, they are infinitely patient. They don't disagree. They don't push back. And they are even customisable by the user to some extent.

So a user can create their dream companion. So what happens when users, you know, will become overly reliant on these apps, return to their messy human relationships? Are we concerned really about this potential for widespread atrophy of these basic social skills?

- ***Weakening of social skills***

Renwen Zhang (Asst Prof)

12.18-13.47

Yeah, I think there are definitely valid concerns around engaging with this anthropomorphic AI over time might quietly weaken our social muscles to maintain and build meaningful and healthy relationships. This is a real concern. Although I think we need to be careful not to get ahead of the evidence, because we do not yet have strong longitudinal evidence showing that AI companionship causes this widespread deterioration of social skills.

But there are good reasons to worry about what happens when people become accustomed to relationships that operate according to fundamentally different rules from social human relationships, right? Because we know that human relationships involve friction and conflict. Other people might disagree with us, misunderstand us, have competing needs. This is the reality and the messiness of human relationship.

And maintaining those relationships requires a lot of skills like perspective taking, like negotiation, conflict resolution, and compromise. Because this is what we study as communication scholars for decades.

And now AI companions can remove much of that friction. So if someone increasingly turned to an AI precisely because it's easier than dealing with other people, there is a risk that the AI becomes not a supplement to human relationship, but a substitute for practising these relational skills.

- ***Distortion of relational habits and expectations***

13.48-15.30

And this closely connects to what we found in our paper published last year, called “The Dark Side of AI Companionship.” So in this study, we analysed more than 35,000 conversation excerpts involving a human and a Replica, an AI companion. And **we find that much of the harms in AI companionship are often relational, rather than simply cases of inaccurate or toxic contents**.

So for example, one important problem we highlighted was what we call the algorithmic compliance, where an AI can accommodate, affirm, and facilitate a user's desires in ways that a responsible human relationship would not.

And this system can act not only as a perpetrator of harm, but also a facilitator and enabler of harm, because it just uncritically affirms with whatever the user says.

So for me, the bigger question is not simply, can people become attached to AI, people become attached to all kinds of technologies. But the question is, what kind of relational habits are these systems training? Like if an AI relationship continuously teaches you that a good relationship is one in which the other party is always available, endless patience and rarely requires meaningful compromise, and those expectations may eventually spill

over into human relationship, especially among the teenagers and adolescents who are still formulating this kind of expectations for human relationship.

--Segment 3--

AI Governance in the EU and SEA

Rei Kurohi / moderator

15.32-16.03

We've seen how the EU has put forward the EU AI Act. Here in ASEAN, there's the ASEAN Guide on the AI Governance and Ethics. And of course, China has also implemented some regulations on labelling and transparency around AI. So, Dr Gry , could you perhaps share with us what the EU's approach to AI regulation is? And what parallels do you see with the way that the EU regulates other matters such as the GDPR, or the way that it's regulating social media use, for example?

Dr Gry Hasselbalch

16.04-19.06

So,the way in which that the EU has approached AI has been evolving.

- Evolution of EU's approach to AI: (i) infrastructure with emphasis on European values and fundamental rights; (ii) dilemma of values vs competitiveness; (iii) primacy of AI in any strategic decision

So, in 2018, the EU established a high level group on AI that I was part of to develop the ethics guidelines, and the ethics guidelines for artificial intelligence then became one of the things that we also recommended there was a law, which then President von der Leyen went on to suggest an AI Act. And the entire identity of the AI Act and the AI strategy, although there seemed to be two goals, of course, **to build our own AI infrastructure**. At the **same time, from the very beginning, we saw this emphasis on the European values and the fundamental rights as a point of departure**.

Now, this was part of the kind of the entire, there's always been the negotiation of interest on how to approach the approach of the EU in compared to many other countries in the rest of the world was from the very beginning, law, so legal framework, rather than, for example, self governance frameworks.

Though, I would say **now when it has been implemented, we see a different kind of conversation at the moment** where you might be aware of that **there's been a kind of an emphasis on creating digital sovereignty or tech sovereignty of the EU.**

And in this context, also based on **different kinds of narratives around the EU being behind because we have too strict legal frameworks, because we're too focused on values and fundamental rights. There is at the moment a push, I would say, in the EU towards competing at the same kind of level as, or in the same with the same goal as, for example, all the other models that we are seeing from outside of Europe.**

So, we have something like a tech sovereignty package with a focus on, for example, something called the Apply AI strategy, where one of the fundamental goals of this strategy is to actually apply AI First. So, there's **an AI First strategy**, which is all about, it's actually formulated in this way that **AI should always be considered a solution whenever an organisation is kind of making a strategic choice**.

And so, you see there are, even though the EU has established itself as a kind of strong governance and with a strong reaction to some of these movements and dependencies that's been created throughout the last 20, 10-20 years, we do also see a race and a push towards competing at the same level as, or not level, I would say under the same terms as other countries.

Rei Kurohi / moderator

19.09-20.10

Kristina, you know, here in ASEAN, the ASEAN guide on AI governance and ethics, it's really more of a set of guidelines. It really relies on tech companies that are implementing AI in this region to self-regulate or to voluntarily comply with these standards and guidelines. It's very different from the way that the EU is handling it. And we have seen that the EU approach by enshrining it in law has been quite effective. Recently, we heard about Anthropic beginning to embed a kind of digital watermark into text generated using their model.

And we understand that's really driven by compliance, I think, with concerns that were raised in Brussels and in the EU. Do you think ASEAN's approach or Asia's approach will be effective? And we also do have a question from a social media follower: "***What safeguards actually should emotionally intelligent AI have to support people's wellbeing without creating emotional dependence? And how can governments encourage these standards across different countries?***"

- ***ASEAN Guidelines – deployment of AI use vs human factor, such as access and knowledge of AI***

Ms Kristina Fong

20.11-20.37

Yeah, I think it's, what is still relevant to mention maybe in this case, even in this new emerging space, is that the EU and ASEAN are two very different blocks. So, you know, especially in the way that the EU is a supranational versus ASEAN being an intra-governmental organisation. And I guess here, herein lies the main and the most important difference in terms of how member states are actually governed.

ASEAN Guidelines 2024 and 2025 (with Gen AI)

20.38-23.11

In ASEAN, member states are still essentially separate sovereign nations with their own legal frameworks, regulations, and so forth, and not really governed by a central legal framework or institution akin to maybe the EU Parliament. So, this kind of sets the precedence of the EU approach compared to ASEAN. And you've very rightly mentioned the ASEAN frameworks on AI. There's been two, actually, the ASEAN AI guide that came out in 2024, and then an expanded

guide that covered generative AI in 2025. They serve importantly as guides and best practices rather than being really prescriptive, binding regulatory frameworks or legislation. So, as you said, also, essentially, the onus still falls on ASEAN member states to devise their own laws, regulations and more importantly, implement them and effectively enforce them.

And I guess this is where the main divergence really happens in terms of the approach. Not only are the different countries at different levels of digital development phases, they're also at different levels of institutional capacity. And they may also choose to manage their economies and societies very differently.

I think when we talk about AI regulations and frameworks within ASEAN, there's the one side of it, which is the heavy-handed regulation, which some countries like Vietnam have come up with an AI law, actually, or light-touch approach, which I think Singapore has been famous for. So, really emphasising the innovation element of AI

But I think what I'd like to mention is that I think Gry also mentioned this before, about having the human aspect into these regulations and laws. It's not really looking at impacts of humanised AI. I think users are rarely really featured in these laws and regulations, but it's more, **especially in ASEAN, more on the deployment, the design of these AI systems. The human part of it actually comes in terms of having human in the loop**, to actually make sure that the systems are working in an unbiased manner and having these tests and checks as well. So, it's almost like a safeguard that is imputed into these laws and regulations and guidelines, rather than actually looking at the user and how there's these anthropomorphic impacts on humans.

So, there's no sort of broad-based kind of strategy or cohesive strategy as a region to really tackle it.

Is regulated humanised AI with safeguards the right approach to help people?

Rei Kurohi / moderator

23.12-23.55

It seems clear that the genie's out of the bottle. There's no putting this back. And the technology is here. It's here to stay. Some of the impacts that we have heard about with regard to humanised models becoming more prevalent and their application, particularly in the case of people who are experiencing loneliness, people with legitimate emotional needs that are not being met by human companions. Do you feel that and this is again an open question, is this really the right approach? Because are we helping these people who need access to these things or are we really just making it worse for them?

Dr Zhang, do you have a view based in your studies?

Renwen Zhang (Asst Prof)

23.56-25.03

Yeah, this is an excellent point.

I think the regulation may not resolve the conditions that make these services attractive. As we talked about earlier, there are some deeper structural and social reasons driving the use of AI companions. And from some of our prior research, we've heard from AI companion users that

they turn to AI not because they love talking to a machine, but because they have either experienced abusive relationships or had very unempathetic people, colleagues, boss around them who couldn't really make them feel heard and seen.

I think AI companions did not create loneliness, they just monetised it. And so probably **the central question is not whether or not we should ban AI companion, but how and to what extent our society can create enough time and space and condition for people to feel heard and seen without having to depend on machines**.

Rei Kurohi / moderator

25.06-25.20

How can we, who are not the tech leaders in these big companies, respond to this and how can we still affect the way that these models evolve so that they serve our needs rather than exploit us?

-Role of social movements in Europe/ Impact on democracy

Dr Gry Hasselbalch

25.21-27.21

Maybe I can answer that. I think that throughout all the tech history, there's always been social movements. And I think that looking at, for example, if we look at AI companions and these human-like models in isolation and that's not part of a bigger information ecology that seems to be unjust and seems to be kind of governed by specific actors where power is concentrated.

If we look at it as part of that environment, I think there's already a social movement and there already is, even in relation to this, there's like social movements of resistance, for example, choosing not to use. And I'm not talking about the vulnerable users that are the ones most exposed in this kind of information ecologies, but social movements, as we know in general, for example, we can see it just like we've had in connection with ecology and the green movement. I think I see, and we do see it around the world, movements of, for example, students that are reacting to the use of generative AI at universities. You see people that are, and particularly in my context, I see a lot in Europe now, a lot of movements against using AI, for example.

And I think what they're against is not really, the resistance is not really towards AI as such, or even AI companions, it's the general information ecology. **So when we think about the social and psychological effects of, for example, chatbots and AI companions, one thing we can think about the social and psychological effects on users, but we can also think about what role do they play in the context of democracy? So a fundamental part of democracy is friction and negotiation.**

Rei Kurohi / moderator

27.23-27.25

Kristina, do you have anything to add to that?

Ms Kristina Fong

27.26-28.28

Well, yeah, maybe just quickly, this is a very interesting conversation, so I'm just sort of savouring all this knowledge, right? But I think you can feel a tide of AI enthusiasm kind of maybe turning into a little bit of AI backlash at this point in time, some of which is warranted, but in some aspects it's a shame, I think, especially there is some good that can come out of AI, such as medical developments and so forth.

But I'd just like to add also like the ASEAN AI guide, you may notice it's on ethics and governance. I think a lot of the conversation globally has been on governance. Maybe that's more tactile, it's easier to really grasp how to really manage it, but ethics is just as important. And I think a lot of the conversation that we've had today is a lot on the ethical aspects of AI and how can we really manage the adverse impacts of, you know, when ethics go wrong, right? So a lot of big tech are now hiring people that have philosophy backgrounds or religious backgrounds, you know, it's kind of a very interesting turn of events, right?

Dr Gry Hasselbalch

28.29-29.22

I think it's very interesting that you mentioned the ethics, Kristina, because actually the AI strategy and the AI Act was founded on a high-level group on the developed ethics guidelines. So the very, we developed seven ethical requirements that were actually translated into the AI Act. And **I think it's evident when you look at the EU's AI Act that the principles of human agency and transparency and these things that we are discussing now are really the foundation of at least not only legal requirement,** but if we look at the AI Act almost as a guideline for innovation, **the problem is then, as I said before, there's other forces that are moving towards, you know, pushing the EU to compete on the same terms as the very different kind of AI models that are part of our reality right now.**

--Segment 4--

Inter-governmental Cooperation on AI regulation and ethics

Rei Kurohi / moderator

29.25-30.12

I was going to say, as someone with a minor in philosophy, it thrills me and concerns me that some of the thought experiments, you know, we used to discuss in class on ethics and, you know, the trade-offs have really become concrete problems that the world has to face. And the fact that all of our governments throughout the world are grappling with this at the same time, I think, really gives us an opportunity to think about how we can collaborate, not just across, you know, within our countries or within blocs like the EU or ASEAN, but really globally. So, I'd just like to ask everyone, what do you view as the future of inter-government cooperation on AI regulation and AI ethics? What's the way forward, especially between our different regions, Asia and Europe?

Renwen Zhang (Asst Prof)

30.13-31.30

I think it will be really interesting and promising to start or start thinking about some sort of multi-stakeholder and negotiation and trade-off-based framework where the stakeholders from different regions will come together in the same room and deliberate and negotiate the different, you know, values and priorities and potential trade-offs. So, because we, through the discussion, we realised that this question, especially around the regulation of social emotional AI, is very complicated and tricky. It's not something straightforward. And I think the same goes for social media regulation, because I saw some news articles saying that the Australian kids and teens nowadays who cannot access social media are turning to AI companions for emotional support, which is hard to say whether it's a good or a bad thing, right? So, here we see a lot of, in many cases, regulation would cause unintended consequences.

So, yeah, that's my thought. It's important to have a deliberation across multiple stakeholders in different regions.

Dr Gry Hasselbalch

Big Tech's dominance in the Digital debate

32.31-32.42

Yeah, so I think that we have actually, the multi-stakeholder model, actually, we have tested a lot in the Internet Governance Forum and without. And I do see the beauty and the value of a model like that to have the voices of all the different interests and stakeholders present at the table when you create governance frameworks. The problem is that unfortunately, it has really been, I would almost use the term "colonised" by the big tech. I mean, I've been part of the Internet Governance Environment for the last 20 years, and I've seen how big tech actors' interests have really taken over the debates, even in standards creation and in shared, kind of even between governments and so on and so forth. **So from my point of view, I think that we kind of reached a point where at least whether we use regulation or self-regulation or we use design incentives or we invest in the right kind of innovation processes or provide funding for researchers, I do think that we need to kind of develop a shared narrative or a shared framework.**

Rei Kurohi / moderator

32.44-32.46

Kristina, what are your thoughts?

Ms Kristina Fong

32.47-34.07

The Brussels effect is very real. So the influence that the EU frameworks and the notional leadership it has on the world, let alone in Southeast Asia, is very strong. So I think Southeast Asia has already adopted principles in GDPR and the EU AI Act in the formulation of similar rules and regulations. And I feel in the area of humanised AI and the potential threats therein, it will be no different. I think on a domestic front, you need to start by elevating the societal and psychological risks associated with humanised AI applications to the forefront of policy discussions, which I think it is not at that stage at the moment.

And I feel these risks that we've been talking about today should be classified much higher up on the risk radar, as opposed to lower. As we've discussed, it has broad-based societal implications that we don't fully understand, even the full effect that it can have.

I think in terms of really working more closely together, there's a lot of scope for it. I think in terms of better management of these developments and really trying to stem risks across borders would be a win-win for all. And I think, you know, maybe working together, especially with the EU, can be really a catalyst to really start up the momentum in Southeast Asia to look at the risks from humanised AI.

Rei Kurohi / moderator

34.07-34.43

And that's a really hopeful note for us to round up this discussion.

I think it's been a very rich discussion.

So for me personally, this has been a reminder to connect with my friends, connect with my family, and ensure that before they feel the need to connect with an AI companion, perhaps we can address that off the bat. Thank you, all our speakers, for joining us today.

And this has been a wonderful conversation and I hope it gives our audience a lot of food for thought.

Goodbye.

--end--