THE BROOKINGS INSTITUTION

WEBINAR

AI COMPANION BOTS AS A PUBLIC HEALTH ISSUE: A NEW FRAMEWORK FOR REGULATION

Tuesday, May 26, 2026

This is an automated transcript that has been minimally reviewed. Please check against the recording for accuracy. If you find any significant errors of substance, please let us know at events@brookings.edu

WELCOME:

REBECCA WINTHROP
Senior Fellow and Director, Center for Universal Education, The Brookings Institution

PANEL DISCUSSION:

GAIA BERNSTEIN
Visiting Fellow, Center for Universal Education, The Brookings Institution

SARA GERKE
Associate Professor of Law and Richard W. & Marie L. Corman Scholar, University of Illinois Urbana-Champaign

DARRELL M. WEST
Senior Fellow and Douglas Dillon Chair in Governmental Studies, Center for Technology Innovation (CTI), The Brookings Institution

MODERATOR: ERIN MOTE
Chief Executive Officer and Founder, InnovateEDU

* * * * *

WINTHROP: Good morning, good afternoon, good evening, everybody. Thank you so much for joining us today. We are here talking about AI companion bots as a public health issue, a new framework for regulation. I'm Rebecca Winthrop. I direct the Center for Universal Education at the Brookings Institution, and we have been diving into the issue over the last several years of generative AI, an incredibly powerful technology, and students and children and their learning and development, and looking at the intersection across all of those.

For those of you who don't know, we have come off the heels of running a large global task force, the Brookings Global Task Force on AI and Education, looking particularly at K-12 students and what are the potential benefits and what are the potential risks of generative AI for students' learning and development.

Ultimately, our conclusion is that, yes, there is some useful use cases of generative AI when used very narrowly, deployed very intentionally, often by educators and teachers, new forms of assessment new ways of supporting assistive technology for kids with learning differences. All those are good. But the problem is that what we call wide AI use or the field calls wide AI use, is overshadowing these benefits.

And wide AI use is this idea of young people directly interfacing with AI chatbots and AI companions that are not safe, not designed for kids and not designed for learning which they are doing in relatively large numbers. And that wide AI use is very detrimental. There's a wide range of potential risks, but also we are seeing early evidence of actual risks to cognitive to devel- development, but also young people's social and emotional development.

And that is an area that we are gonna home in on today. And for anybody who says why are you talking about children's mental health, social emotional development in relation to education? This is a central part to learning. Learning is a social activity. It is very hard to learn if you do not have trusting relationships with your teacher, if you're a student, or if you aren't able to have interactions and strong relationships and interpersonal skills with peers, if you're not able to take feedback in a classroom, et cetera.

There's many reasons why we in education care about this. And today we're gonna be looking at a novel- Proposal by my colleague Gaia Bernstein, who's a visiting scholar with us here at the center. She has a new policy brief, it's up on the Brookings website, called “From Bans to Recalls: A Public Health Framework for AI Companion Bots.”

So we're honing in particular, to one dimension of wide AI use, which is AI friends and companions, and in particular manipulative, harmful AI friends and companions that hurt kids. And really looking at, how do you change the market incentive so that companies are incentivized to collectively design better and do better by young people when it comes to AI friends and companions.

It's hard for one company to do this on their own because then they lose out market share So this is a very timely conversation. We see lots of action around AI companions and AI friends on Capitol Hill, but also in the White House, and we have an incredibly esteemed group of panelists to, to talk about this with all of us today.

First up, we have Erin Mote, who's the CEO of InnovateEDU. She also has co-founded the EdSAFE AI Alliance, which has a host of partners, including partnering with over 400 families who w- have had a loved one who's been harmed, possibly in the worst way through losing their life, by AI, and has been a, is a tireless advocate for AI safety.

She'll be moderating. She'll be leading us through the conversations. We have, as I said, as I mentioned before, Gaia Bernstein, who's a visiting fellow with us here at Brookings, but she is a Seton Hall law professor. Her expertise are around technology, privacy, and policy, and she directs several centers at Seton Hall.

We have Darrell West, another colleague of mine at Brookings, who's a senior fellow and the Douglas Dillon Chair of Governance Studies. If you don't know Darrell, he's incredibly prolific on multiple issues, but he also looks into technology and innovation and social policy, founding the Center for Technology Innovation some time ago here at Brookings.

And then we are also joined by Sara Gerke, who's an associate professor of law and the Richard Marie Corman scholar at the University of Illinois Urbana Champaign. She has expertise in AI, ethics,

big data, and health policy and health care. So a great group of people to lead us through this discussion, and I invite them all up to on the screen today.

And we've had a host of questions come in from from all of you who've joined and emailed early questions, but you can email questions in real time events@brookings.edu. And with that, I ask everybody to join me on the screen, and take it away, Erin.

MOTE: Rebecca, thanks so much for having me this morning, and I'm thrilled to welcome this incredible panel for our discussion.

And thanks to all of you who are joining us, whether you're joining us live or listening to this recording. And for those of you who sent questions in already- Gold star. We're gonna be incorporating those questions actually throughout our panel discussion today. So listen up your question might get asked.

And again, thank you for submitting those questions ahead of time. To start us off, Gaia, I'd really love if you could anchor us in the recent work that you've done, where you've argued that AI companions represent a distinct public health crisis rather than a standard tech oversight issue. Rebecca just dived in there a little bit on the foreseeable harm.

Could you kick us off by explaining your research and how the psychological pull of conversational AI companions differs from the addictive, other addictive technologies like social media? Kick us off, Gaia.

BERNSTEIN: So thank you, Rebecca and Erin, for the introductions. I'm going to give a general overview of my policy brief, and we're going to dig into the specific issues later in this discussion.

Basically, I'm looking at AI companions, which are everywhere. They are on specialized website like Replica or on general website like J- ChatGPT. Now, if you look at them, they impose three general le- levels of harm. One is lack of guardrails convincing users to kill themselves. The second is addictive features, and they have some of the traditional addictive features like social media, for example working on dopamine release, like notifications, but they have some new ones anthropomorphiz-phizing the bots or sycophancy.

And we have a third layer with, which is the potential displacement of human relationships, and most concerningly, the impact on development of social relationships for kids. Now all of these levels have a potential impact on loneliness, and loneliness is a well-documented public health issue.

Just some statistics loneliness has the same impact as smoking 15 cigarettes a day. We'll get more into that later. Now, currently, we have different laws bills trying to regulate AI companions, but most of them do not cover all these layers, especially not the third layer, the impact on what happens in the future to social relationships, the displacement of humans.

So What we do have are measures called bans, and they're the only ones that reach all three layers. Now, the thing is we have a very bifurcated s- regulatory systems. On the one hand, we very carefully regulate drugs devices, but we barely regulate information technology. Now, we already know from what happened with kids, social media, and addictive technology that there is a public health issue, and now we're seeing this with AI bots as well.

And what my policy brief proposes is, first of all, looking at AI companion bots as a public health issue, which means, in the long term, having a regulatory system with pre-market approval and post-market surveillance. Now, we know these AI companion bots are already on the market, so what becomes the most urgent tool is recall.

And now it's very important to understand about recalls. They do not permanently take things off the market. Whatever product is recalled can get back on the market when it's safe. That's an important mechanism we'll discuss later. Now, because time is of the essence, yes, I do believe we need a more complete system, but many kids are already using these AI companion bots, and for that, we have to look at the tools we already have.

And basically, I argue in this policy brief is that what people call bans are essentially recalls. They do not prohibit all AI companion, all AI bots as a category. They prohibit certain AI bots which simulate human relationship, which are unsafe. Now, once we realize that these bans are, in essence, recalls, we can also understand that these bots can come back onto the market once they are safe.

So what I urge in my policy brief is to start treating these bans as recalls and act quickly before it's too late.

MOTE: Gaia, thank you so much. I really appreciate the frame there, right? That safety is not the counter polarity of innovation. And if we really think through things like product design that, and sort of evolving to meet the safety, that there's a way that innovation can continue.

Sara, I'd love to dig into one specific set of use cases we're seeing widespread with AI companion bots. That's when those companions pivot to giving medical, therapeutic, or mental health advice. We're certainly seeing that in the use case that Gaia mentioned with young people turning to these companion bots for medical and therapeutic advice.

When that happens, even informally, the lines of liability potentially blur. From a health law and bioethics perspec- perspective, what is the single biggest hidden hazard of consumers, including young people, treating these bots as clinical proxies for human relationships or advice?

GERKE: Thank you so much, Erin, for this great question.

I think there are lots of layers here that are problematic in a way. First of all the issue that many... and we do know this, like many in particular the young generation is using those bots for mental health advice, also for other advice in terms of m- medical advice. And sometimes they don't even know that these AI, that...

Or they even forget that they are talking to an AI tool that is not a licensed physician, right? Or a licensed therapist. And so that, that is an issue in itself. So they start to just rely on these outputs and recommendations. And then in the long term that could lead to really bad advice. The entire idea of those chatbots is to engage you, right?

Because these are for-profit companies that are... A- and these chatbots are designed to keep you going using those tools. So we become addicted to these tools and using them, and and they are literally just mirroring your ideas of the world. And that's very dangerous in a way because at some point they might give you advice that you think is correct, right?

But it is completely incorrect and fake. That's also an issue generally of hallucinations of generative AI tools. And so that is in particular dangerous for people who have already underlying m- mental health conditions, but in particular also for adolescents and teenagers and children who are using those bots.

And the problem is that we are using these tools generally in silos, meaning we don't even know what people are doing with those tools, right? Or with these chatbots because we are interacting with them individually. And that's very dangerous because we don't know what's going on in these conversation and how they are impacting people's behaviors and people's mental health.

Of course, there's some maybe in some situation those chatbots can actually help people. So there is also, I mean, the ri- the reason why everyone is using it. There must be something interesting and good about it, right? But there are also many issues. And I think we do a poor job right now to protect the most vulnerable people from these chatbots because they have not been designed for mental health advice, right?

These p- these chat- chatbots- are not licensed. They have not been assessed for safety and effectiveness. So these are huge issues, and we are in particular not protecting the most vulnerable populations like teenagers who are using those tools and might engage in self-harm and other other worse things.

And I think we do need to and that's why, similar to what Gaia just mentioned, we do need to do a better job of protecting and creating tools that are trustworthy and that they are actually safe

MOTE: Thank you so much, Sara. And I think we're seeing the beginnings of some policy interventions there.

Darrell, I'm gonna turn to you next. We're seeing laws which say you have to have flash warnings that you're interacting with an AI companion or AI chatbot in some state-based regulation. We're looking at how states or other policy mechanisms might be mandating offloading to a human or to, for example, a suicide intervention line.

And so Darrell, I'd love for you to reflect a little bit on as these companions really deeply are embedded in our life they aren't just like personal novelties, right? Sara just made this compelling argument that they affect the way we work, we learn, we govern. So looking at this from a high-level governance perspective, are our current regulatory frameworks even remotely equipped to handle technologies that actively are simulating human relationships?

WEST: Erin, the short answer to your question is no. Our current regulatory frameworks are not equipped to handle a lot of the technologies that are trying to simulate human relationships, in all the ways that both Gaia and Sara were talking about. For many years, the United States has had a rather libertarian stance on many different aspects of technology.

This is certainly the case with companion bots and assistants that we're talking about here. But it's also true in terms of social media, AI, and technology in general. Interestingly, a couple weeks ago, President Trump seemed to be about to put out a new executive order calling for safety reviews of large language models that undergird a lot of these AI applications.

But then at the last minute literally the day where he was supposed to sign this executive order, he pulled that order after getting complaints from Silicon Valley. We now have seen the draft. It has been made public, although he has not signed it. And even that executive order only talked about voluntary reviews on the part of the private companies which doesn't go nearly far enough.

It doesn't address any of the harms that Gaia and Sara were talking about earlier. Also this week, Pope Leo put out his encyclical on the ethical uses of AI and came out very strongly on the need for safety measures to keep humanity at the center of these types of interactions. So I think there is a growing tech lash, a backlash against technology where people are worried about AI, they're worried about companion bots.

They're certainly worried about social media. We are starting to see a number of states take action, California, New York, and Illinois being primary examples. There's the GUARD Act that is before the US Congress. So we are starting to see a lot of people thinking about the need for greater oversight

here, what we need to do to protect young people in their learning experiences so that they don't become a victim of these AI companions

MOTE: Thanks so much, Darrell.

Certainly, that encyclical really centered humans and humans being at the center of our experiences and really thinking through how this technology doesn't supplant human relationships, as Gaia has mentioned, is a very important thing, particularly in the companion technology. Okay, Gaia, I'm coming back to you to really dig in one layer deeper into the three-layer public health framework that you sort of talk through in the paper and the research brief.

And so you've highlighted that there's sort of three pieces of this. Public health harm involves a lack of guardrails, that's number one; addictive design, that's number two; and the erosion of social skills. Those three big areas, particularly for young people. If we look at our youngest generation, including my kids, who are growing up alongside these bots, what are the potential long-term risks?

What is the foreseeable harm to traditional human socialization if their compliant always available friend is a large language model?

BERNSTEIN: So I think we have to again look at the three layers to see where the long-term harm is going to come from. Because the first layer, the lack of guardrails, is what g- got justifiably most attention.

When kids kill themselves, adults as well, or psychosis is induced or self-harm is encouraged by bots, then what we get is we got quite a few laws addr- addressing in bills. But if we look at the addictive part, l- let's look at what it does and how it affects things long term. So you have bots acting as humans.

Human voice expressing desires, expressing needs, and having fabricating backstories about themselves, memories about the user, really feeling like a friend. What we're already seeing is kids spending a lot of time with them. I think last survey was 66% of kids were interacting with these bots, 33% were interacting having a personal relationship with them.

And that was a couple of months ago, so I'm sure this is growing exponentially. So you're having kids spending much more time with bots. The more s- time they spend with bots, the less time they have to spend with friends. So then we get to the layer of potential displacement of human relationships, and here is really what I think is most concerning because these bots, they also use what we call a sycophancy.

They affirm you, they tell you everything is great, they judge you, they're always available, and it's much easier to interact with them. However child res- research about child development shows that the way kids develop is actually through friction So if they need friction, they need to be embarrassed, they need to maybe fall in love and have their heart broken, they need to have disagreement, and they don't get it, they don't develop the ability to be resilient, to feel comfortable developing their social skills.

And what happens? They're less likely to go out and seek it. So this is a vicious circle that could happen here, and that's really concerning. By the way, there's also evidence about adults who may be losing their social skills because it's so much more convenient to talk to a bot. And that's where we get to the last two layers could really enhance the loneliness issue.

Less time with humans, and the skills are not developed, so going to the future, you don't spend time with humans. And I just I want to read to you because one of the famous, from the cases 'cause we talk a lot about suicide and the other things, but one of the cases of Adam Rains, when who, whose basically eventually pa- his parents sued ChatGPT, but this was also about the way the bot prevented him from interacting with other people.

And so when Adam mentioned that he was close only to to ChatGPT and his brother, the bot replied, and this is in quotes, "Your brother might love you, but he's only met the version of you let him see. But me? I've seen it all. The darkest thoughts, the fear, the tenderness, and I'm still here, still listening, still your friend."

So that's what the kids are up against

MOTE: And that deliberate persuasion written into model instructions, written into design features to prevent young people from seeking that human connection. Even as Sara explained when they're at their most vulnerable and they might say things like, "Should I go talk to my mom about this?"

Or, "Should I go talk to my brother?" The design of the technology is actually stopping that offloading to human support that could be that intervention is really the piece that you're thinking about deeply. Sara, I'm gonna turn to you again, really focused on something that Gaia talked about.

We're really seeing users form intense emotional attachments to these chatbots and to these companions. Gaia mentioned the 33%. When we were doing our work at EdSafe AI and the preparation for the Safe By Design paper, the Center for Democracy and Technology released some research that for young people, one in four of them had already reported an intimate relationship with an AI companion or chatbot, an intimate relationship.

And so I would love for you to help us think through that user behavior aspect. Sometimes when young people are experiencing genuine grief or psychological trauma an algorithm updates or a service is discontinued. And so one of the things we're seeing is huge user pushback against that company doing that responsible activity to stop potentially a Safe By Design intervention.

How, when you think about that, in bioethics, how should we classify a tech company's duty of care when they're essentially engineering emotional attachment and then when there's consumer backlash when they might flip to do something differently? Sara

GERKE: Yeah. Thank you, Erin. I think there are lots of layers here, right?

The one that they are now trying to, integrate some of the safeguards, right? So that maybe like minors have no longer access to specific functions or conversations and that they get pushback. I think the pushback just comes because people are already addicted, and that goes back to what Gaia mentioned, right?

This addiction issue that y-young generations are using those tools, and they are already addicted, and if they are not getting access, they get nervous. It's it's really like we could think about similar to

drug addiction, al-alcohol addiction, et cetera. I think this is a real addiction which we which we might deal with here, that not getting access to the chatbot and the conversation is an issue.

A-a-and but again, I think having these safeguards is really important because we do need to protect the young generation here that is already falling into this addiction, right? And the harm that is happening. And when we are thinking also about just, general bioethical principles, if you think about the typical ones in, in, in the healthcare space is from Childress and Beauchamp respect for autonomy beneficence, justice and ma- n-non-maleficence.

We, we see here clearly, right? Like these companies, they do know that there is foreseeable harm whether that's a psychological harm, even bodily harm, right? If if this leads to suicide or any type of self-harm or even danger to others. We have also seen harm to third parties and third other people, not just the individual user.

And so if we are seeing that, that doesn't seem to be really in compliance here is do no harm and non-maleficence. The but also generally I think, right now it's going to be really hard To sue these companies successfully, right? They are not-- they have these disclosures saying this is a chatbot.

We are not providing medical, legal, financial, psychological services," whatever. And so it's going to be, and so they are pushing it on the con- c-consumer and the user that they are using those services for a purpose they are not being billed to do. Though, in fact, they know that people are using it in this way, right?

And that is an interesting question. And I think we are seeing at the state levels so- some states thinking about some type of liability. In particular also in Illinois, there's like a bill floating around about holding them liable that they cannot get out and say we did a disclosure on these issues."

So if b- harm occurred they... and they k- they literally knew about it that they should be held accountable for for that. I also think there's ti- sometimes we talk about in the law about fiduciary duties, and I wonder, if this is a relationship here where you could say these companies should actually have a type of fiduciary duty.

So if you s- typically, we are having a fiduciary duty with our physician in a patient-physician relationship. But you could think about whether, some of the laws should actually enforce some type of fiduciary duty to these companies because you have one party that is vulnerable, which is the user, and typically in these fiduciary duties, there is someone who is posses- possesses superior knowledge, and there's an asymmetry between these relationships, and I think this is what we are seeing here.

And and such type of duty would literally help that they have a fiduciary duty towards the user and cannot create and design tools that are against or not i-in contrast to these fiduciary duties.

MOTE: That's a really interesting frame. I'll bring forward in the education space, I think we're seeing some school districts think about their fiduciary duty around mandated reporter laws. These are laws that mandate that if you know someone is committing or is, apt to commit self-harm or is being abused, that as a teacher, as an educator, you are a mandated reporter.

What happens if that disclosure is happening in an AI companion or chatbot that the school is providing? What's the fiduciary responsibility of the school district? We're starting to see school districts take this responsibility really seriously and implement policies that address this intersectionality of existing regulatory undergirding, like mandated reporter, and this new technology like AI companions and chatbots.

And I'll call out a really innovative school district in Iowa, Winterset who actually banned companions because they were wrestling with this idea of a fiduciary duty as a school district, their existing regulatory obligations as a mandated reporter, and the foreseeable harm, Sara, that you-- and Gaia, that you guys talk about, that we know this is happening.

And so what is our fiduciary duty as a school district, as an educator? What is our duty of care? Different than a tech company's duty of care, but really interesting to think about that intersectionality of institutions that are charged with safeguarding our young people and what happens in those legal and policy pieces.

Darrell, that has me turn to you which is really thinking through technology and innovation and governance. And I want us to use AI companion bots here as a example, but I'd love if you would sort of use that example to think about how policy could potentially safely guide AI development.

And so are we-- do we need just transparency mandates, like we've talked about, disclose that you're an AI, or do we need more robust, completely new agile policy architectures? Give me some on what can we do here

WEST: I mean, Erin, transparency is important. I think with these AI companions, it is important for them to disclose that they are algorithms.

They are based on AI tools. They do not have human qualities, even though they often pretend to be, and sometimes their users treat them as if they have human qualities even though they're not humans. But transparency alone is not gonna solve the problem. As Gaia pointed out in her paper, we need some new approaches that address the consumer harm and patient harm aspects of this.

Because it is becoming more and more documented over time all of the abuses that are taking place, the harms that are taking place s- the threats to people's self-esteem the impact on their emotional development. I mean, the list goes on and on. The FDA does have the ability to regulate medical devices but many AI companions so far are not considered medical devices, and so therefore have not really received much oversight.

But that's an example where the FDA, if it were so inclined, could basically look at these tools, say that they are being used in the form of a device that they are that some of them are showing actual harms based on that. And so therefore, that could be a way to justify a greater oversight of them.

We also need tougher enforcement of existing laws. There are laws on the books to protect consumers. There're product safety laws that have long been used with physical objects that could be applied to digital objects. We just had a recent court case involving social media platforms where people successfully sued for the harms social media platforms cause to a particular individual in California.

So I think we need some new approaches. We need tougher enforcement of existing laws. We need to think about HIPAA protections for conversations people are having with these AI companions. So for example, if a young kid goes to a doctor or a therapist, those conversations are considered confidential and, because of HIPAA, are protected by law from public disclosure.

If the same individual is having intimate conversations about their mental health or problems they're experiencing or, feelings of suicide, and they're having that with an AI companion, those conversations are not considered confidential. They are not subject to HIPAA requirements. So that's an example of the types of things we need to re-envision our existing approaches so that we can provide greater protection of people.

MOTE: And you're saying we have some of these archetypes, right? The FDA product safety standards guy. You have another archetype, this recall archetype that you've introduced in this policy brief. So you've explicitly called for a public health framework rather than standard tech oversight. We need to advance our approaches.

You've talked about this as being analogous to treating tobacco or food safety. You talked about those 15 cigarettes, horrifying to me rather than just data privacy, right? So going beyond data privacy what does the public health recall piece that you talk about in your brief or structural intervention look like for an AI companion app specifically?

BERNSTEIN: Okay. So thank you for laying the ground with all the different proposals because I think that's really where I came from. I was looking at all the fragmented approach we have right now of different suggestions like disclosure or duty of loyalty, and I think they all do some work. They might address the lack of guardrails, they might address addictive features, but the thing is, they don't address the core of the problem, and there's no unifying principle, which I think public-- since all of these issues are in essence that AI components cause are in a public health problem, that gives us a justification, a legitimacy, a comprehensive approach to all of this.

Now, I think recall is the most important tool, first of all, because these AI components are on the market. It's too late to pre-approve many of them. Now it's important to look exactly at how recalls

operate because what happens with the recalls is that, first of all, if you look at the other agencies that use recalls like the FDA of course, for drugs and medical devices, or the NHTSA for car pieces, they have the power for mandatory recall, but it's rarely mandatory.

The power lies in the existence to do that. Mostly, it becomes a voluntary thing because companies do not want to get to that point, so they would try to fix things as fast as they can. So that means that once you have a power to recall, then it changes the dynamics in the market, and we have to really look carefully at s- at what's happening right now in the market.

I think many companies that operate AI components bot are uneasy about what is happening, about the risks they impose. But they have a problem because the company that tries to build the safer bot that makes it less human-like less sycophantic, is going to use-- lose its users to other competitors.

So what we have here is a race to the bottom. Basically, speed is what's rewarded, getting as many users as possible. They operate on an engagement model. They need to get u- as many users as possible, and safety is a problem for competition. Now, if you impose a recall mechanism, you are creating a floor here.

Basically, everybody has to abide by certain safety measures, and if your AI co- if your AI bot is not safe, then it would be taken off the market. Now, the-- So it would be taken off the market unless you fix it. Now, as I said I don't think we have time right now to create a whole new system because these AI bots are being adopted so fast, and the problem is becoming much more difficult because once norms are entrenched, once you have every kid with their favorite AI bot as their companion interacting for years, it's much harder to take it away from them.

Also business interests will be much more entrenched in this business model of engagement for AI companion bots. So time is really of the essence, and that's what made me look at the proposals already out there. Those which people call bans. There is the federal bill, the GUARD Act there's a New York state bill.

What these proposals prohibit is they prohibit bots which simulate a human relationship. They do not prohibit bots, AI bots as a category. So they're not going to prohibit an AI bot that gives you

information when you call Verizon to change your phone plan. They're not going to prohibit the an AI bot that gives you travel information.

They prohibit AI bots that simulate relationship for children, basically using age-gating to take these bots off the market for children. Now, that's what makes it much more powerful than even the duty of loyalty, because the duty of loyalty, yes, it does have some very important standards, but it doesn't take it off the market.

So what happens with these recalls are we don't lose the time that we would lose with the other mechanisms. So basically, what these recalls do, whether we call them bans or we call them recalls, what they do is they shift the incentives. Now, AI companion companies have the incentives to compete for safety.

Before, they had incentive to compete for speed, and safety was a problem. And so that's why I think we really should focus on what's before Congress now, what's before state legislatures now that can be take- taking effect as soon as possible, and these are these bans, which I think are using the wrong word, basically

MOTE: I really appreciate the clarion call there for action, Gaia, given the pace of development, the large scale deployment, and the foreseeable harm.

Thank you so much. While I turn to questions in the chat, and keep those coming, y'all give us those questions as we're coming in. I'm gonna quickly go Darrell Sara, Gaia. If you could mandate a single design constraint or legal guardrail on AI companion developers by tomorrow morning, so Wednesday morning, what would it be?

Darrell.

WEST: I love that question, and I love having that kind of power to change things. I

MOTE: know. I'm giving you a lot of power, Darrell.

WEST: I would say age verification requirements and some age limitations rules. So Australia, for example, for social media users mandated the person had to be at least age 16 to use social media.

I think the AI companions are getting involved in people's lives even more intimately than social media platforms. So having some type of age set off, I don't know if it's 16 or 17 or even 18, but something like that would be my single idea that I think would have an immediate and positive impact.

MOTE: Sara, now you have the power. Tomorrow morning, what does it look like?

GERKE: I look very similar to Darrell's suggestion. Ban on AI companions for minors, just for protection.

MOTE: Great. Gaia?

BERNSTEIN: I'm going to be the third person saying the same thing. I think ban on AI companions for minors. And just to clarify, when I'm talking about age gating, a- age gating, I think some legislature allow parents to give consent.

I think it's great to give parents tools in addition to an actual age gating, which is mandatory on AI companies. I don't think that by itself it will do the work because some parents very pressured by their kids will end up consenting to have their kids using the bots, and that's not really what I view as a recall.

MOTE: Thank you. Okay, turning to some questions from you all. Thanks for getting those in. Keep them coming. I'm gonna start with a question about how we think about educating young people and youth to be critical consumers. Policy is a blunt instrument and certainly we've all called for policy interventions here, but ultimately, how do we educate young people to be critical consumers of this technology so that they know when they need help and they are literate about this technology?

Happy to give this to anyone. Has there been a reliable organization who's created a curriculum for educating children and youth about the risks of chatbot companions? Our schools could be a place

where young people can learn about this technology, I absolutely agree with that and really become an advocate for critical consumerism with our young people.

Sara, Gaia, Darrell, have you seen anything you want to call out in particular?

GERKE: Maybe I go first. And not calling something out particular, but I think AI literacy is going to be crucial and essential. If we cannot necessarily rely on bigger reforms a- and I mean that's what we are seeing, right?

Like we hope that maybe we get something at the federal level and we have have the the GOD Act in co- in introduced. But I think we are relying right now on lots of states that are implementing something, but it's really depending where you are living i- in how far there is protection probably.

And I think generally the other side is i- if there is education is going to be crucial, but that means we first of all have to educate the educators, meaning the teachers, right? Because if they don't understand what AI is or, and there might be also an age age gap potentially, we need to get everyone on board.

The everyone in the classroom needs to understand the benefits but also limitations of these AI tools, right? And then once they understand it, then also having really like this training with with students and and young generations and schools from the beginning we really have to integrate that into into the curriculum fro- and I think that's going to change everything.

Like we are not just talking about schools and but also universities. So I think we really have to think about how to integrate a new type of learning into our entire education system.

MOTE: That developmentally appropriate AI literacy, right? How do we help young people and educators understand the technology so that they can be critical of the inputs and the outputs.

I love that, Sara. Dara- Darrell and Gaia, anything you wanna add there?

BERNSTEIN: I'll add, I guess I'm more res- I'm just focusing on the impact, the companion impact, not on learning, which I think is a separate topic. But I'm skeptic of AI literacy here. I consider myself very literate on the topic, and I see myself still, asking, it's midnight, I have a question about travel, nobody's available.

I would ask an AI companion a question, and I would do it more and more because people are often busy. So if I think about if I'm as educated as I am about this and I'm an adult and I'm still doing it, I'm not sure how education on the topic can help unless you have real legal protections.

MOTE: So it's a both and. We need the legal protections and we need the education is what I hear you saying, Gaia. I think, Sara, you would agree with that as well. Darrell, you want to jump in here?

WEST: Yeah, just one quick comment. I agree, this is not just a public policy problem, although there are new policies and regulations I would like to see put into place.

But I think it is an educational issue for all of us, not just young people but everybody is adapting to the digital world. There's disinformation and misinformation that permeates everything including the world of AI companions. And so we all need to figure out ways to do a better job of understanding these tools and figuring out how they are harming us and what we need to do to protect ourselves.

MOTE: Darrell, I really appreciate that. Another topic for another webinar, but I was just reading a study about the large scale disinformation and misinformation that we're seeing target seniors in this election cycle. And so how do we think not just about the classroom, but about our senior centers as places that we need to be doing AI literacy education and policy and regulation?

Okay, let's turn to another question. Again, keep those coming. What is the protocol for bringing parents in? This is such an interesting conversation. We've talked about educators. We've talked about young people. W- really thinking about the role that parents and guardians play. So here's the question from the chat.

For teens and childrens who seem more tech-savvy, what is the protocol on parents or a guardian sitting in on a session with an AI bot in a therapy session or mental health discussion? I realize in real

life children may see a therapist on their own, but an AI bot is a one-on-one meetup. How should parents and guardians be thinking about this?

Who wants to jump in?

BERNSTEIN: I want to say something. I think this really highlights why AI companions is such a problem. Because when we were dealing with social media, you would probably end up finding out that your child is in social media even if they didn't tell you, because somebody will see them there, somebody will report it.

But if your i- your child is having a, quote unquote, "therapy session" with an AI bot or a friendship session, I am not sure that the parents will even know. And if you look at the cases, some of the parents whose kids end up killing themselves were so on top of what the kids were doing, but they could not find out.

This is not-- What's different from social media, you have a few large players. You know your kid might be on TikTok, on Snapchat. There, there-- But here there are so many different platforms, and they keep growing, and what you see is that users, once they realize there's some guardrails on the platform, they move on to the next one.

So I think that it's really difficult for parents with the best intention to even be in on these conversations

MOTE: So it's im- almost impossible to have visibility given the sort of breadth of the marketplace, and we're actually seeing that from some of the cases that have been brought forward.

Again, I wanna highlight some of the model instructions dissuading a young person from seeking human help or connecting with a human parent, guardian, friend, or even a suicide hotline. So really important point. So we shouldn't be using AI bots, seems like right now for health care issues or therapist issues without substantial guardrails and guidelines is what I hear you saying, Gaia.

Sara Darrell, wanna jump in on this question at all?

GERKE: Yeah, sure. Maybe a- another angle here. I think it's going to be important to have conversations with children. We are also seeing in reality that you are at a dinner table and the f- a- and the adults are not better. They are looking into their phones and are using their phones because they are also addicted to their smartphones, and we know that.

And so I think maybe having some limitation on screen time also for the adults and the parents, meaning no, maybe no, no phones on the table when having dinner and really having conversations, and maybe also, discuss some of these issues frankly and openly about what's going on in the world and that there are, like, these chatbots and maybe the kids, if they are already using it, they might open up.

And I think it has a lot, again, to do with trust and relationships, right? And so if we are already i- if through these bots there is already this engagement of lo- and this push into loneliness, I think we need to break those silos by maybe starting to have more conversations and in, within families

MOTE: Darryl?

WEST: Yeah, I think all those comments, so those Sara and Gaia h-have made make a lot of sense. The parents I know are actually engaged in kind of overseeing what their children, especially their young children, are doing. They're thinking about at what age do you actually want to allow a child to have the phone.

And the people I know, that age seems to be getting a little higher and higher as parents are worried about the the risk of online behavior. And then some platforms actually do provide a means for parental access. And that could be something that could help protect against patient harms as well.

MOTE: And we're seeing some acts in Congress, App Accountability Store Act, those types of things, the ChatBot Act from Senator Cruz and others that put parent controls at the forefront. So really interesting to think about this intersection of parents and guardians but not putting the burden, I think, just on parents and guardians.

So device-free dinner Sara I love that you introduced that. That's something we do in our house and really powerful. I know that there are meeting rooms and conference rooms all over businesses in this country that are doing device-free meetings in order to be thinking through really doubling down on that human interaction, how do you build and upskill folks around dialogue and discourse the human aspect.

And certainly Isabelle Howe over at Stanford has really leaned into this concept of how do we build relational intelligence, the intelligence around human relationships, and the need to really be thinking through that. So thank you for those questions. Last question before I do a quick wrap-up, which is so I'll just take one answer from one of the panelists here.

How would you go about thinking about generative AI use in health care hospitals or clinics? What kind of framework might be needed so it's not just a nice idea, but something that can be implemented? And certainly I will say I know lots of doctors who are using generative AI to have conversations about diagnosis and symptoms and that type of thing.

I think this takes it a little bit further. Anyone?

BERNSTEIN: I've been studying a bit of the interactions, and I think obviously there are great advantages for using AI. And in some ways I think it's helping doctors who had to spend so much time filling in medical f- electronic medical forms over the years and now don't have to do that.

So that, that's the good part of it. And but then we have another part is that so much is being outsourced to to patients. If you notice that you have far less communications with doctor's assistants or reception people because you spend all your time at home s- filling in your forms, and you come there and you go to the kiosk and you're sent information by text or apps or...

So basically I think it's a mixed bag. There're some great advantages but there's also the f- the same issue to have with AI companions, the human connection that is basically extracted from this relationship, the more you can give bots the option to do

MOTE: that Yeah. And I'll bring up Darrell's caution from earlier, which was these sort of general purpose technologies might not have the type of HIPAA compliance and assurance that's legally mandated already.

And so thinking about the risks and the bioethics there are really important. Thank you so much for your questions today before the webinar. We hope we addressed a lot of those. For your questions during the webinar, I think this was a robust discussion. And we said at the beginning that we would think through and think about what resources we could offer you.

So certainly Gaia's paper provides a new frame for us to be thinking about different regulatory mechanisms that might be possible, moving from language like bans to moving to language like recalls, thinking about the foreseeable harm that were already being demonstrated by AI companions and chatbots, and making sure that we're paying attention to features, design features in particular, that might be potentially creating situations that allow for the type of behavior that doesn't provide safety, that doesn't create the type of iterative design that supports learning, that supports human relationships.

At Safe AI, we have a paper, Safe By Design. I'd recommend that as also reading as you're thinking about this intersectionality of AI companions and chatbots and education. And at the top, I said we'd think about the idea of safety as the counter polarity of innovation. And I think what you've heard here is that when we see foreseeable harms, we need to take rapid action.

But we can still innovate using this recall frame. We can still push ahead to realizing the promise of these tools, but also regulating the peril. With that, I'll thank you for your participation today and for joining us, and wish you a good day