**Technology that supports peace and social cohesion: Examples and case studies**

***Transcript as delivered***

*World Goodwill Seminar, New York
November 11, 2023*

Lisa Schirch

This is my first time with your community, so it's nice to hear more about your values and where you're coming from. I'll be talking today about the role of technology in social cohesion.

So, my work at the University of Notre Dame focuses on the role of technology and social cohesion and I'll explain a little bit more about that in a minute. I just want to sort of give us a broad historical approach to why technology has become so critical in helping humanity, but also in undermining our ability to work together. And that's when I say social cohesion, it's really about the glue that holds everyone together. The United Nations defines it as the ability of states, governments and societies to work together across divides and for each individual in that matrix of society to feel like they have agency to impact the world around them.

So, if we think back on the last 20 years, particularly on technology, there's been a variety of different narratives, really starting with optimism and innocence, in around 2000 where we thought, this would, technology would really help people be more productive, and, social media started with the explicit idea of connecting people. Around 2007, the first social media and search engines started becoming monetized, meaning they started using advertising to generate profit for these platforms. And as we look back, that was a really critical moment because it ended up amplifying and generating some dynamics that have been very harmful for society.

But even in 2008 to 2010, we still were hearing mostly that technology was going to have a positive impact on peace and democracy. There were a few Stanford University professors who thought that Facebook was going to create world peace in 30 years just by bringing the world together. It seems strange to read those words now, but when we saw the Arab Spring in 2010, 2011, we did have the sense that maybe people would start fighting on social media and they wouldn't need to fight in the streets anymore. That they would be pursuing their interests, their human rights through social media rather than on battlefields or in the streets. But that's, of course, not what has happened either. It was really in 2013 that we started getting the red flags or the awareness that social media was causing harm.

And the alarm bell really was sent from Myanmar. From the civil society, peace building groups in Myanmar, who informed to the world first in 2013, that the government was using social media, Facebook in particular, to organize a genocide against the Muslim Rohingya population. And it was really in response to some of these early harms and alarms that the tech companies built what they call “the trust and safety elements” of their work. So now there are whole trust and safety teams at Facebook. They started around 2017, 2018. Governments tried to catch up with this idea of, “oh, we have to somehow build some guardrails to social media with regulations”, civil society was doing advocacy.

But by 2020, we were starting to see that what we call industrial scale digital harms, and I'll talk more about that in a minute. But basically, this is a timeline of we're not really certain what comes here in the future, but because there's been a lot of change over the last 20 years.

In terms of what harmful content has been happening online there, in the 1990s, people were worried just about spam, some nudity, harassment, and child exploitation online. In the 2000s, it became the early 2000s, it was hate speech again, and terror recruitment so ISIS started recruiting on Facebook and people were promoting material that was about self-harm which was, of course, harmful to humanity. But it was 2010 when we started seeing the escalation of this with polarizing political speech by mainstream political candidates who were using Facebook, um, to spread information that was concerning about other groups of people sort of targeting them.

There were disinformation campaigns from other countries that were targeting especially in the U. S. but also in France and other countries. And really there, there is the development of foreign information operations where, yeah, countries were trying to divide population. So instead of social cohesion, we saw incredible polarization.

I'm going to say more about this next era now in this slide, because what we see now is this sort of tidal wave of false and hateful information online. Some people call it the collapse of truth, because there are so many rumors and blogs, accusations, different sources of information saying completely different things that many in the public just don't know how to make sense of this.

And so, there's this sort of collapse of our grounding of how do we get information and, this is in part because many nations, like over 50 countries now have cyber armies. So, these are armies, military personnel in, in warehouses, sitting by computers, basically fighting an online war of propaganda.

And the goal of that is, is very confusing. It's not just like to attack one thing. It's really, they're often supporting both sides and generating hateful content coming from two sides of an argument. So, the point is not to have people join one side or the other. The point is polarization. And so, these cyber armies are making it difficult for us all to feel like we are one people living on planet earth.

And what we've seen increasingly is all that digital hate and this disinformation online is spilling over into real world violence. In tech companies, they created community guidelines, they created trust and safety teams in the 2010s. Now governments are starting to get involved. But really the future of our information environment, our information ecosystem is in the balance in terms of how are we going to handle this tidal wave, this tsunami of harmful content aiming at dividing societies.

The people at tech companies who monitor this hate speech and this disinformation, they're called content moderators. Content moderation, what they do, pulling off harmful, false information online is highly contested and that in and of itself is polarizing. In many countries, there's hardly any content moderation because companies don't have staff [that speak local languages. Yeah, as I just said, there's sort of like a real divide in terms of what is disinformation, because we don't agree anymore on basic facts or basic sources of information.

When tech companies take down content, it sometimes feels like a violation of free speech. And what is another factor in all this is that last year, the tech sector laid off 100,000 tech workers, and most of those were content moderators. So, a few years ago, there was more content moderation online than there is today. So, the trafficking in false, hateful, deceptive information online is growing every year, and there's very little that is stopping it. There's mostly just fuel to that fire.

Now, people often say “well, it's not technology to blame for polarization” and in part, they're right, because polarization is natural in societies. It happened long before technology began. The point of technology, what isn't actually often the start of. Polarizing conversations or divisions in society. But what it does is in the profit models incentivize and amplify distrust and hate for other people. And that's because the loudest voices on social media keep people there the longest, and tech companies make more money, they show more advertising the longer any of us stays online. So, their goal is just to show the most distressing or polarizing content often because it keeps people on the platforms longer, and it's related to their profit model.

And this is where my presentation kind of takes a turn in the road, because up to this point, I'm laying out the problem. A lot of people would say we just need more tech regulation by governments, or we need tech companies to hire back more content moderators, but my research shows something different. And that is that the very code of these digital platforms Lays the groundwork of whether that platform will contribute to social cohesion, or whether it will fuel toxic polarization.

And in a recent article I wrote, I say, the road to hell is not paved in gold, the road to hell is paved with code. And this means that technology, “code” the design of the platforms, determines its impact on society. So, code is not neutral, it is shaping our human behavior, and it is showing us some content and no other content and making decisions that actually are steering humanity in one direction or the other.

There's sort of in my research, there were three narratives of, what do we do about all of this harmful content online? The first I call the “User-centered narrative” and tech platforms voice this view. They say that their platforms in their view are neutral mirrors, simply reflecting user-generated content. So, algorithms can demote. So, algorithms are the math that determines what we see or don't see online. Because we can't all see everything. So, what each of these tech platforms does is decide on a personal basis based on your search history, what you've liked or disliked and who you are, all the factors of your identity, it determines what content you will see. And while tech companies control that algorithm, they're saying they cannot, tech companies can't generate a pro-social content, good content between people.

But the second view, the second narrative comes actually from former tech workers, people who used to be inside Meta, Facebook., Google. And they say it's actually the design of these platforms that is causing so much problem. And they use the analogy that digital platforms, technology is like a Coliseum. So like digital spaces are like gladiator arenas where there's a few people fighting online and everyone else is just basically watching spellbound at this fight.

And building a platform that amplifies the most divisive voices is not good for our human society and our ability to solve problems together. Because if we can't talk to each other, if we're hating each other, and if we don't have accurate information, we can't solve global challenges and problems that we have.

And this is where sort of I'm saying we need a new generation of people who are focused on regulating technology designs. And this is where we're focusing now on identifying what makes a technology platform really serve humanity or be pro-social, we say, pro-humanity. And there are already pro social tech platforms that I'll go over here in just a minute. And I'm thinking that we can learn from them, and we can hopefully amplify and increase the public support for using these other platforms.

In the service of this goal, I have helped to set up the Council on Technology and Social Cohesion. So, this is a place, a series of workshops, conferences, events, and publications where we're bringing together people from the technology companies, peace building professionals, conflict resolution, mediators, as well as government and business leaders to come together and try to find out how we can incentivize the support for pro-social technology.

Let me go to this next diagram here. So, I would say there's in my conflict resolution work, and I've worked in Iraq, Afghanistan, many countries around the world on peace processes. And there are sort of a series of steps or stages in terms of that peace process. They're setting norms, building relationships, gathering information about what the conflict is about, deliberating on possible solutions to the conflict and, generating a final solution that works for the most people or addresses the needs and interests of most people. So, what I've mapped onto this sort of conflict resolution process are how technology platforms can contribute at each of those different levels. So, this is like the digital anatomy, the anatomy of digital platforms to bridge divides, to build social cohesion between people.

So, I'm going to just quickly go through some of these examples. What we could have on Facebook and Twitter and other platforms on setting norms. And then I'll go through some of the other solutions.

But I'm Tags, buttons, nudges and public service announcements was where I want to start. So, we call these “humanizing affordances”. Let me start actually with identity and intention tags. This is the idea that actually eBay has already started using a long time ago, and that's because they hired a conflict resolution expert to help them think about resolving conflicts between buyers and sellers on eBay. So, this little example, this little image at the bottom of the screen there on seller information, it shows that, the social confidence that this person has had good interactions with other people. It's also a verification tag. This is somebody who's registered, so it's not an anonymous person selling something to you on eBay. It's a registered verified person. There's a reputation accrual. So, this 99.8 percent positive feedback is a sort of reputation that a seller gathers over time with what they're, what their identity is on this platform.

This, what we can learn from eBay is that other platforms could be doing this too. We could be using intention tags I intend to listen and learn from others”. Sometimes what's happening on social media is that you can't read the context here and you're just feeling like the person might be angry with you when actually they're trying to understand. And so, what an intention tag would do is allow you to say why, what your purpose is in engaging in this conversation.

Another way of taking the polarization down a bit is to have identity tags. I could say I'm a mother, biker, gardener, artist, and other people might see me not just as somebody commenting online, but more as a full human being with a lot of richness.

We could also have reaction buttons instead of just the like button or the angry button, which is particularly harmful. We could have a common ground button, when we feel agreement, or we can appreciate that we might not agree on everything, but we have some common ground with something somebody has said. We could appreciate other people's views and just say, say, “I don't agree with you, but I appreciate your experiences that have led you to those beliefs”. We could have a reaction button that is just a pair of glasses saying, “Oh, now I understand your point of view better”. We could be able to show solidarity with others, we could give thanks to others. So those are just a few of the kinds of reaction buttons.

There's a variety of other things that Meta and Twitter and other platforms could do. They could have pop ups what we call public service announcements, basically, encouraging people to listen to each other, not just to shout at each other.So, taking the goal of the platform not just to be the loudest voice or the most polarizing voice, but being able to reward people who are listening well and paraphrasing other people well. On Twitter, there was actually an experiment coaching people with a public service announcement that would pop up if somebody started typing something that was hateful, or maybe not all the way true, there was a meanness warning that would just say “Most users don't do this”, and Twitter found that this really impacted people's behavior. Also, what eBay found is that they coach people on their tone in responding to people. They said “Your comment is unlikely to be effective based on its tone.

And to achieve your goals, try expressing yourself like this”, and then they would give some draft language. At Notre Dame, my students and I are using AI to generate sample languages for responses to disinformation or hate speech. And so, we're hoping this can be like a coach. If somebody sees hate speech, they can call up our little chat bot that will just offer and generate for them some ideas about how they might respond to protect the humanity of the person who's being targeted, but also to call out the humanity of the person who's saying it.

So that's the 1st step that I wanted to talk to you about of this stair step of conflict resolution and social cohesion building. There are a variety of platforms that have been built specifically for dialogue.So, Soliya does dialogue between high school and college students around the world who come from different backgrounds. Right now, with the war in Israel and Gaza, there is an emphasis on pairing up people from the region, to be able to talk with each other online with a facilitator there helping them to do that. Gatherround is another platform and really there's a variety of different platforms out there that have been designed specifically to help people across cultural divides to understand each other. A variety of different initiatives, this one in Bangladesh by the UN Development Programme are using technology to also do these “peace talk cafes” to promote the values of peaceful and inclusive societies on social media.

So, it's actually really beautiful when you start learning about all the people using technology to try to build social cohesion online. In Chicago, which is close to me here at the University of Notre Dame, social workers are using Facebook to reach out to gang members and to try to befriend them and figure out if there's a way of being supportive in really the de-escalation or demobilization, like leaving gangs and reintegrating back into civilian society.

So, I've gone through setting norms, building relationships. I'd like to move on to gathering information, and just quickly show you this slide from Syria. The Carter Center, which is a major conflict resolution, think tank and do tank. It does a lot of good peace building work in the world. They measure online data to visualize the shifting movements of armed troops. And, to help bring together parties with information about what is actually happening.

I'm going to jump up to the next stairstep for the sake of time here and talk about the use of technology for deliberation. There's a platform called Remesh that is used actually by the business community to build internal solutions within different companies. But Remesh has also been used by the United Nations to help publics in Yemen, Tunisia, Libya talk with each other and really deliberate about what are the policy options in their country? What is possible? How might they move into the future? And what might a peace process or a peace agreement look like that might bring sustainable peace to that country.

So Remesh uses artificial intelligence. So, it often, like a lot of people online can put in their qualitative ideas. So, they're typing in their ideas. That's a qualitative input. And then they can vote basically up or down. So, they agree or disagree with other people's contributions. So, there's a qualitative and a quantitative input here. And what artificial intelligence does, as you see in this automated analysis is it makes it digestible what thousands of people are saying, so you can listen at scale to people. So, this is what humanity is really needed is an ability to listen to thousands of people on what people actually are really caring about.

And this is important, not only because it gives individuals a sense of agency, that they can participate in the decisions that affect their lives. But it really helps policymakers decide, what policies might be sustainable and have public support. So, Remesh is one platform that does that.

The other one is called Polis, and Polis is part of the Computational Democracy Project. You can read more about Polis online. I didn't highlight the name here, but it's in this in this blue square. You can see P O L I S. Polis is another platform, also built to help people listen to each other at scale] and really to solve problems at scale. Many countries now are using Polis to enable the public to have input onto different policy discussions. In Taiwan, the Digital Ministry in Taiwan used Polis dozens of times. It's helped them solve more, more than 30 policy dilemmas and reach policy consensus and actually pass legislation.

So, the idea is that governments would use Polis to bring together a polarized group of people or to find out where is their polarization, where is their disagreement? And the platform is unique. Well, it's similar to Remesh, but it's what makes the two platforms unique is that the algorithms, the AI finds the common ground between people. So, it helps people see not only what everybody is saying, but where everybody agrees. And really that is like the key to peace building and successful peace processes, is you have to be able to find where there is agreement online.

So, in closing I've just gone over some of the digital technology design affordances.So, by design affordances, like how you design them with certain buttons and tags and nudges and features and algorithms and how we use AI to really find the common ground between people. So, with that, I will stop and open it up for any questions.