**The Future of the Metaverse: Its Implications for Safety and Security**

Summary by Stephen Downes

What is the metaverse?

| | Identity and avatar manipulation for nefarious purposes | Digital human interactions with government agencies | Societal manipulation through metaverse-driven propaganda |
|---|---|---|---|
| Philip Rosedale (US) – Wen Metaverse? Are we there yet, and are we ready? | Second life allows people to build this, like housing and avatars. | Second Life has its own economy. Second life has community and group moderation tools. Second life (compared to social media) is a much more positive experience for people. Moderation costs are much lower. | Largest successful sites are populated by kids. Some sites, like Decentraland, populated by adults (but much fewer). Kids stop using them in late teens. The quality and nature of communication is unacceptable to adult but OK for kids; it doesn't have the right kind of non-verbal communications. New VR personas still not acceptable. (As it gets better) we can see each other in a visceral way. Greatest concern is the availability of data for surveillance. |
| William Swartout (US) – Using Generative AI to Make Virtual Humans a Reality | 'Light stage' allows us to create a 3D representation of a person's face. But: creating models takes a lot of work. GAI & ML used to work from library of face types (large facial model). Looks like person, behaves similar to that person (used to replace dialogue).<br><br>Also: commercial tools (eg MidJourney) text -> character creation. 11labs for voice. Also: example of 'Oppenheimer style'. (Took about an hour). | | We've already seen misinformation in the U.S. election cycle. |

| | | | |
|---|---|---|---|
| | AI character generation: Increased productivity, lower entry costs, raises risk of misuse. | | |
| Jonathon Gratch (US) – Avatar Influence: The psychology of persuasion in the metaverse | Perception of metaverse as replicating reality.<br><br>But we alter reality.<br><br>We do this because we have a social goal in mind.<br><br>'Supernormal elicitors'.<br><br>Make avatar feel closer and more familiar to the other person.<br><br>Avatar can be perceived as superhuman, eg., to get people to reveal finances. Or mix elements of your face (aka 'mirroring') to engender trust. | Can track social cues, figure out what sort of person they are, track data, trace bank account – perceive much more powerfully.<br><br>Transform social cognition, eg., a single agent can service ten people, supported with AI. AI potentially replacing service agents altogether.<br><br>How do we detect augmentation is present? | The metaverse can change how we engage with each other socially, eg., how I present myself to you (eg., by using better mic) or more complicated AI stuff ('um' removal) or make myself look more beautiful -> makes people think your smarter, more likely to fund.<br><br>Methods integrated into platforms – snapchat beauty filter, zoom summary – this will be common and socially accepted.<br><br>Reinforce bias, recapitulate stereotypes.<br><br>Mayer's trust model – ability, benevolence, integrity -> trust -> lower perceived risk -> greater risk |
| | Identity and avatar manipulation for nefarious purposes | Digital human interactions with government agencies | Societal manipulation through metaverse-driven propaganda |
| Jeremy Bailenson (US)– Emerging Risks from Tracking, Rendering, and Display in the Metaverse | VR and AR both require tracking and rendering systems. Motion is necessary for presence in XR. Example of Apple Vision pro – all the cameras. | VR & AR for training.<br><br>Cognitive and behavioural assessments using headsets – heart rate, eye tracking, skin conductance, IMU, cognitive load, etc. | 40-50 million VR headsets worldwide – but headsets don't work unless tied to social media account.<br><br>Protecting nonverbal data tracked in virtual reality. |

| | | | |
|---|---|---|---|
| | ‘Transformed social interaction’ to change the way the scene looks (to prevent people from knowing who you are) | ???<br><br>Problem of people using headsets while driving !! | Mark Miller – using tracking data (eg. head position) to identify people (511 people, 95.3% accuracy). Huge data sets collected from popular video games. |
| Emma Barrett (UK) – Trust and mistrust in the Metaverse: How and when immersive technologies facilitate persuasion and influence | Barriers / factors of use: resources required, time, money, skill. Would have taken a lot of time & money, but is now much more accessible.; reaching the audience – invites, advertising, sideloading, etc. -> | Worth it? Maybe, high risk.<br><br>Gorilla tag<br><br>AI personalization<br><br>Impact of regulation | Persuasion in VR: communication is multimodal, influence in multiple contexts (MUPs, audiences, private chat., single user apps); inhabited by multiple characters.<br><br>Tools: deceptive avatars (deepfakes), priming effects (environment), experiences (dark patterns, false memories), malware (eg. inception attack)<br><br>Coming: targeted advertising, recommenders - & therefore risks<br><br>Impacts: changing attitudes, beliefs, etc., compliance, reinforcement, recruitment |
| Shane D Johnson (UK) – Crime facilitated by the Metaverse | Metaverse: multiuser, ownership, persistent, multipurpose, etc.<br><br>Real-time photorealistic rendering of a person in real time -> coming soon | Convergence of different technologies. – blockchain, eg., digital twin, internet of things, teledildonics.<br><br>Uses/interactions: entertainment, creative, hospitality, education, work, retail & ads, health | Social media<br><br>Threats:<br>- Doxing<br>- Harassment<br>- Radicalization |

| | | | |
|---|---|---|---|
| (first two slides key and important) | Threats:<br>- Child sam, grooming<br>- Imposter scams<br>- Identity theft | Threats:<br>- Blockchain attacks<br>- Tax evasion<br>- DOS,. DOES<br>- Money laundering | - Hate crime |