Shaping the Future of Media,
Entertainment and Culture

WORLD
ECONOMIC
FORUM

Ethical Principles for Digital Media and Technology Design in the New Normal

COMMUNITY PAPER

SEPTEMBER 2020

Cover: Getty Images/Scyther5

Contents

Overview

Digital media consumption and design in the new normal

With 3.8 billion active users of social media globally today, and an additional 1 billion internet users projected to come online in the coming years, the role of digital media is growing in scale and importance. COVID-19 has brought the functioning of media and technology products to the forefront of society. The use of mobile communication apps, video conferencing tools, and food delivery applications are just some examples of digital tools that are now inextricably linked to modern day living, allowing people to easily connect with one another, search and obtain quality information and entertainment, and lead more convenient lives.

The pandemic has radically changed how people use digital media technologies and consume online content, as people in different parts of the world grow more and more reliant on mediated platforms for remote working, grocery shopping and other activities during periods of lockdown. In this light, the benefits of technology have never been clearer. COVID-19 tracing apps have allowed people to quickly identify whether they have been in contact with anyone who has contracted the illness and take immediate action to limit the spread of the virus by self-isolating. For students who can no longer go to school, technology has enabled remote learning and helped to maintain peer connections. To avoid crowded stores, many applications have emerged to enable at-home delivery of essential goods.

Despite all the tangible benefits, the same technology platforms are being weaponized by bad actors, posing numerous challenges for consumers and makers of digital products. Platforms are facing up to 15 times the volume of harmful content since the start of COVID-19. Europol also reported an increase in online child sexual abuse since the pandemic while fraud and cybersecurity attacks have increased substantially. Vulnerabilities have been exposed by bad actors that are looking to benefit from the upheaval caused by COVID-19.

This has raised a number of questions when it comes to the ethical creation and distribution of digital media. This report aims to highlight key principles of ethical media and technology design through examination of four case studies based on Safety by Design principles. Specifically, it looks to answer the following three questions:

1. What principles should businesses follow to design digital media products ethically and effectively counter the increased sophistication of bad actors looking to expose product vulnerabilities?

2. How have existing products and services fared according to established ethical tech design principles? What were the major hurdles to overcome?

3. What learnings can be applied from case studies examined in future digital media product developments?

1

Ethical Technology Design Principles

This report aims to highlight practical applications of the Safety by Design (SbD) principles that the eSafety Commissioner (eSafety), Australia's national independent regulator for online safety, developed beginning in 2019. These were developed through extensive consultation with over 60 key stakeholder groups in recognition of the importance of proactively considering user safety as a standard risk mitigation during the development process, rather than retrofitting safety considerations after users have experienced online harm.

SbD emphasizes the need to address online harms, alongside user safety and rights. These principles provide guidance in incorporating, enhancing and assessing user safety considerations throughout the design, development and deployment phases of a typical service lifecycle. The principles firmly place user safety as a fundamental design principle that needs to be embedded in the development of technological innovations from the start.

At its heart are the SbD Principles, a model template for industry of all sizes and stages of maturity, providing guidance as they incorporate, assess and enhance user safety. The three key principles are as follows:

SbD Principle 1: Service provider responsibilities
The burden of safety should never fall solely upon the end user. Service providers can take preventative steps to ensure that their service is less likely to facilitate, inflame or encourage illegal and inappropriate behaviours.

SbD Principle 2: User empowerment and autonomy
The dignity of users is of central importance, with users' best interests a primary consideration, through features, functionality and an inclusive design approach that secures user empowerment and autonomy as part of the in-service experience.

SbD Principle 3: Transparency and accountability
Transparency and accountability are hallmarks of a robust approach to safety. They not only provide assurances that services are operating according to their published safety objectives, but also assist in educating and empowering users about steps they can take to address safety concerns.

1. Service provider responsibilities	2. User empowerment and autonomy	3. Transparency and accountability
– Nominate individuals, or teams – and make them accountable – for user-safety policy creation, evaluation, implementation, operations. – Develop community standards, terms of service and moderation procedures that are fairly and consistently implemented. – Put in place infrastructure that supports internal and external triaging, clear escalation paths and reporting on all user-safety concerns, alongside readily accessible mechanisms for users to flag and report concerns and violations at the point that they occur. – Ensure there are clear internal protocols for engaging with law enforcement, support services and illegal content hotlines. – Put processes in place to detect, surface, flag and remove illegal and harmful conduct, contact and content with the aim of preventing harms before they occur. – Prepare documented risk management and impact assessments to assess and remediate any potential safety harms that could be enabled, or facilitated by the product or service. – Implement social contracts at the point of registration; these outline the duties and responsibilities of the service, user and third parties for the safety of all users. – Consider security by design, privacy by design and user safety considerations, which are balanced when securing the ongoing confidentiality, integrity and availability of personal data and information.	– Provide technical measures and tools that adequately allow users to manage their own safety, and that are set to the most secure privacy and safety levels by default. – Establish clear protocols and consequences for service violations that serve as meaningful deterrents and reflect the values and expectations of the user base. – Leverage the use of technical features to mitigate against risks and harms, which can be flagged to users at point of relevance, and which prompt and optimise safer interactions. – Provide built-in support functions and feedback loops for users that inform users on the status of their reports, what outcomes have been taken and offer an opportunity for appeal. – Evaluate all design and function features to ensure that risk factors for all users – particularly for those with distinct characteristics and capabilities –have been mitigated before products or features are released to the public.	– Embed user-safety considerations, training and practices into the roles, functions and working practices of all individuals who work with, for, or on behalf of the product or service. – Ensure that user-safety policies, terms and conditions, community standards and processes about user safety are visible, easy to find, regularly updated and easy to understand. Users should be periodically reminded of these policies and proactively notified of changes or updates through targeted in-service communications. – Carry out open engagement with a wide user base, including experts and key stakeholders, on the development, interpretation and application of safety standards and their effectiveness or appropriateness. – Publish an annual assessment of reported abuses on the service, alongside the open publication of meaningful analysis of metrics such as abuse data and reports, the effectiveness of moderation efforts and the extent to which community standards and terms of service are being satisfied through enforcement metrics. – Commit to consistently innovate and invest in safety-enhancing technologies on an ongoing basis and collaborate and share with others safety-enhancing tools, best practices, processes and technologies.

Source: Safety by Design Overview, eSafety Commissioner, Government of Australia, May 2019.

2 Case Studies and Evaluation

This section examines a number of products/applications to highlight both good practice notes, but also areas that posed a challenge to a safe design according to the ethical tech design principles outlined in the previous section.

2.1 Case study 1: ALHOSN app

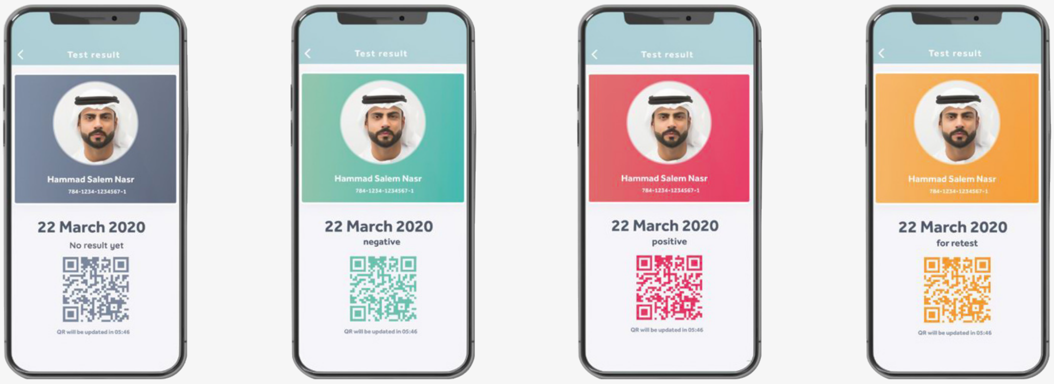
Description and features

ALHOSN is the official app for contact tracing and health testing related to COVID-19 for health authorities in the UAE. Developed in the UAE, ALHOSN is a joint initiative between the Ministry of Health and Prevention and local health authorities, endorsed by the National Authority for Emergency and Crisis Management.

The app helps trace people who have been in close proximity to confirmed COVID-19 cases to prevent the spread of the virus. It also provides residents with access to their test results and a health colour-coding system that identifies and allows individuals that are free from COVID-19 disease to access public spaces.

The ALHOSN health coding system involves the generation of a unique personal QR code and a colour-coding system that determines the status of your health. The colours represent the following:

- **Grey:** indicates you have not been tested.
- **Green:** indicates you have been tested and you are healthy; this grants you access to public places.
- **Amber:** indicates that you need to be tested or retested due to possible exposure to COVID-19.
- **Red:** indicates your most recent test have shown that you are infected and must seek the advice of the health care authority.



GREY	GREEN	RED	AMBER
Indicates you have not been tested	Indicates you have been tested and you are healthy; this grants you access to public places	Indicates your most recent test have shown that you are infected and must seek the advice of the health care authority	Indicates that you need to be tested or retested due to possible exposure to COVID-19

Source: https://www.alhosnapp.ae/en/using-alhosn/

How the technology adheres to safety design principles

ALHOSN uses a decentralized model for contact tracing as recommended by global privacy experts. It was created using AI technology similar to that being implemented successfully in other countries around the world[1]. ALHOSN adheres to Safety by Design principles by securing data privacy of users using a combination of back-end encryption and users' consent and data control. At the backend, ALHOSN uses an encrypted Secure Tracing Identifier (STI) that is exchanged locally on devices of individuals in proximity of one another. The STI consists of anonymized data and a timestamp. The anonymized STI is stored in encrypted form temporarily for three weeks on the phone. No personally identifiable information is collected throughout the entire process.

In other words, users' personal information is stored on the ALHOSN app in an encrypted form. Even when the app communicates with other phones, personal details are anonymized. The health authorities will only need to access the contact data list of an individual who is infected.

Beyond relying on system architecture to protect users' privacy, the success of ALHOSN relies largely on users' consent and involvement. For instance, users need to provide consent when they register for an ALHOSN account, which allows health authorities to obtain the list of anonymous IDs for the past 21 days for contact tracing. In addition, users are able to control whom they share their COVID-19 test results with.

Key challenges to ethical design and privacy

Like many of the contact tracing apps, there are several challenges to ethical design and privacy in the implementation.

First, developers need to contend with the health equity issue, where people from lower socio-economic backgrounds may not have access to the advantages of ALHOSN as they may not possess a smartphone. Currently, research has shown that COVID-19 disproportionately affects people from lower socio-economic statuses or minority groups and underserved communities. As such, the government is considering rolling out other options (e.g. wristbands) to facilitate contact tracing.

The second challenge is motivating people to download and to voluntarily enable the app to run at all times given the success of contact tracing requires a large number of the population to all be using the app. This means that the government would need to put in a significant effort in understanding and addressing the privacy concerns of citizens through public education and closed-loop feedback.

2.2 Case study 2: Twitter

Description and features

In 2019, Twitter announced that it was seeking public feedback on a draft set of rules to govern how it would handle synthetic and manipulated media. While early drafts of the policy included possibilities such as placing a notice next to these tweets, warning users before they liked or shared tweets, and adding links to additional contextual information, the first adopted version of the policy outlined a rubric for handling tweets containing manipulated media based principally on three criteria: whether they contained synthetic or manipulated media, whether they were shared "in a deceptive manner", and whether the contents were likely to affect public safety or "cause serious harm".

Beyond labelling or adding context to such tweets, Twitter made clear that tweets meeting certain combinations of these criteria were likely to have their visibility reduced or would be removed from the platform entirely.

Is the media significantly and deceptively altered or fabricated?	Is the media shared in a deceptively manner?	Is the content likely to impact public safety or cause serious harm?	
✓	✗	✗	Content **may** be labelled.
✓	✗	✓	Content is **likely** to be labelled, or **may** be removed.
✓	✓	✗	Content is **likely** to be labelled.
✓	✓	✓	Content is **very likely** to be labelled.

How the technology adheres to safety design principles

Twitter's efforts with respect to synthetic and manipulated media align with the three key areas of Safety by Design thanks to both the content of the adopted rule and the process by which they developed in. For example, by circulating their draft rule and actively seeking public input via an online form (which was made available in English, Hindi, Arabic, Spanish, Portuguese and Japanese) and publishing some of the results of that survey when announcing the new rule, Twitter provided valuable transparency about their decision-making process in designing the new rule. Similarly, by offering additional context and labelling around manipulated media, rather than simply removing it, Twitter helped preserve a space for meaningful expression and commentary, while also giving users significant autonomy in how they interacted with tweets.

Finally, by flagging manipulated media directly within the platform's user interface, Twitter has accepted significant responsibility for identifying and drawing attention to (or away from) manipulated media on its platform.

Key challenges to ethical design and privacy

One of the key challenges associated with evaluating content is the risk that material shared for the purposes of satire or commentary – both powerful forms of expressive critique, especially in the online space – will have its visibility limited because it technically violates Twitter's new rule, even though the platform does not notify users that their tweet has been labelled, or offer any appeals process to request removal of the label. This is especially problematic because such tools invariably rely on automated methods to identify manipulated content, and incorrectly applied labels may significantly undermine the effectiveness of the measure.

There is also the risk that political and legal entities will begin to use the labels as a premise for investigating individuals or organizations. In many countries and jurisdictions, including France, Russia, Germany and Singapore, laws criminalizing "fake news" or requiring it to be removed are already being used, directly and indirectly, to censor legitimate speech and news coverage

2.3 Case study 3: LEGO Life app

Description and features

The LEGO Life app nurtures kids' education with LEGO building ideas and decorating challenges, allowing them to build themselves using the LEGO Minifigure Avatar maker, bring their Avatar to life with an augmented reality feature, and much more. The app is packed with content, characters and entertainment from the LEGO Group. There are loads to discover in the form of videos, build challenges, behind-the-scenes previews and inspirational builds. But more importantly, children can use it to express and share their own creativity – drawings, stories and photographs of their unique LEGO builds. In the fully pre-moderated community, they can discover building tips and hacks to learn how to use LEGO bricks in brand new ways, join interest groups and be inspired by and engage with each other, using emoticon and moderated text comments, watch videos featuring LEGO Harry Potter™, LEGO Star Wars™, LEGO NINJAGO, LEGO Friends, LEGO Minecraft™, LEGO Technic, LEGO City and more, in an entirely kid-safe platform.

How the technology adheres to safety design principles

The LEGO Group fully embraces Safety by Design principles, and partnered with UNICEF (the first toy company to do so) to set new standards, which are embedded in all aspects of the design and operation of the LEGO Life app. From the internal investment in pre-moderation of all user-generated content (UGC), ensuring there is no inadvertent sharing of a child's personal information, the introduction of verified parental consent, explaining to and empowering parents to manage their children's digital permissions, to the robust escalation safeguarding processes and transparent policies and reporting, the principles of service provider responsibility, user empowerment and transparency and accountability are firmly embraced.

In addition, the LEGO Life app enables children to develop the 21st century skills they need in an increasingly digitized world. Captain Safety is a popular character in the app who, using plain and appropriate language, is on a mission to explain to children the importance of staying safe online. They can "digitally sign" the LEGO Safety Pledge, which means they have understood and will follow the rules to stay safe online.

The act of having content moderated is also a feedback loop for children and a "learning-through-play" moment that teaches them about digital safety, but also goes further to address digital well-being. Children may have their UGC rejected to then be taught the principles of personal safety and how to avoid sharing identifiable information online or plagiarism, and the importance of claiming one's own work and how to be kinder, even when represented by an anonymized avatar.

The impact on children's screen time during the COVID-19 pandemic has been dramatic. Online tools have been required for everything, from access to schools and learning to the peer and family connectivity. It has, therefore, been more important than ever to provide children and their caregivers with clear information and tools to help them safely navigate the digital world.

The LEGO Group has used LEGO Life (and other platforms) as a trusted vehicle to engage around these issues; Captain Safety has never been more prolific. As an aside and in further recognition of this issue, the LEGO Group launched a new resource for families called Small Builds for Big Conversations. This is a series of creative challenges, which offer parents and their kids an enjoyable, guided method to engage in conversations about being a good digital citizen and the importance of online safety, in a relatable way.

Key challenges to ethical design and privacy

As LEGO Life was already built around the safe-by-design principles and focused on protecting and supporting children's right's online, the technology was already fit for purpose, even in the face of the global crisis. The LEGO Group is firmly committed to protecting and supporting children's rights online and carefully approaches the challenges in balancing a child's right to protection – and privacy – alongside provision and crucially, participation.

Children are also by definition minors and therefore below the age of legal consent. Thus, they cannot agree to or accept data policies and cookie management as adults can, however plainly it may be explained. The LEGO Group significantly invests in initiatives such as pre-moderation, parent messaging and communications and verified parental consent in order to enable a child's safe participation and permission.

2.4 Case study 4: Zoom

Description and features

Zoom is a video-conferencing service designed to support online meetings, e-classes, webinars and live chats. It was originally built for enterprise customers, but started gaining momentum with the mass market during the COVID-19 pandemic as people around the world looked to minimize physical contact and meet online. Zoom's daily meeting participants increased from 10 million in December 2019 to 200 million in March 2020.[2]

Zoom provides a set of features to support video conferencing, such as waiting rooms, screen sharing, annotation, whiteboard functionality, chat, polling, breakout rooms and cloud recording.

How the technology adheres to safety design principles

According to Zoom's April announcement, a number of measures were introduced by the company to address the security and privacy problems. Zoom announced a freeze in new feature development effectively immediately and shifted all engineering resources to address security and privacy issues. The company conducted a comprehensive review with third-party experts and setup a CISO council.

Zoom has offered a series of training sessions, tutorials, free interactive daily webinars to users, and has taken steps to minimize support wait times so that customers could be empowered to use the various settings offered within the product to set up more secure meetings. Such features include:

- Limit attendance to participants who are signed in to the meeting using the email listed in the meeting invited
- Set up a waiting room function
- Password protect meeting access
- Lock meetings once they start
- Mute participants who are not presenting
- Remove unwanted participants
- Disable private chat

Zoom is preparing a transparency report to detail the information related to requests for data, records and content. The company's chief executive officer has hosted weekly webinars to answer questions from the community.

Key challenges to ethical design and privacy

A series of security flaws and privacy breaches were found on Zoom.[3] On 30 March 2020, the FBI issued a public warning about a problem named after Zoom – "zoom-bombing."[4] The company was also said to have shared user data with third parties[5] and make false claims about the encryption algorithm it uses.[6] In June, Zoom shut down accounts of Chinese dissidents after an online Zoom meeting about the anniversary of 1989 Tiananmen Square social movement.[7] On 1 April, Zoom published a blog post to respond to a list of criticisms and outlined a 90-day plan.[8] On 29 June, the company appointed a new chief information security officer.[9]

3 Summary and Next Steps

The following directional evaluation was conducted for the case studies described above in line with the three key Safety by Design principles:

✓✓✓✓✓ Excellent adherence to principle

✓✓✓✓ Good adherence to principle

✓✓✓ Some adherence to principle

✓✓ Poor adherence to principle

✓ Little to no adherence to principle

**Limiting autonomy and user functionality was a deliberate choice to safeguard children in this case*

	Principle 1: Service provider responsibilities	Principle 2: User empowerment and autonomy	Principle 3: Transparency and accountability
ALHOSN app	✓✓✓✓○	✓✓✓✓○	✓✓✓✓✓
Twitter	✓✓✓○○	✓✓✓✓○	✓✓✓○○
LEGO Life	✓✓✓✓✓	✓✓✓○○*	✓✓✓✓✓
Zoom	✓✓✓○○	✓✓✓✓○	✓✓✓○○

Through assessment of in-market products against a robust set of principles for Safety by Design, it is clear that the COVID-19 pandemic has posed a number of unique challenges. It has become even more difficult to design ethical digital media consumption and distribution experiences given the increased scale, reach and exposure of products to a broad user base. Finding ways to map out various scenarios of how one's product could be abused by bad actors and building in safety precautions during the design phase will be central ethical design of products. It is evident that a more proactive, preventative and transparent approach to design is needed to safeguard digital spaces and enhance user value in the future.

Disclosure: two of the case studies examined in this report - LEGO Life and the UAE AHLOSN app - involved assessment by parties who belong to the organization designing the applications.

4 Contributors

Members of Global Future Council on Media, Entertainment and Culture

Nehal Badri
Director, Brand Dubai, Government of Dubai Media Office, United Arab Emirates

King-wa Fu
Associate Professor, Journalism and Media Studies Centre, University of Hong Kong, Hong Kong SAR, China

Ayesha Khanna
Chief Executive Officer and Co-Founder, ADDO AI, Singapore

Edmund Lee
Assistant Professor, Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore

Susan McGregor
Associate Research Scholar, Data Science Institute, USA

Anna Rafferty
Vice-President, Digital Consumer Engagement, LEGO Group, Denmark

World Economic Forum

Farah Lalani
Community Lead, Media, Entertainment and Culture; Co-Council Manager, Global Future Council on Media, Entertainment and Culture

Cathy Li
Head of Media, Entertainment and Culture Industry

Kirstine Stewart
Head of Media, Entertainment and Culture Platform

With thanks to:

Julie Inman Grant
eSafety Commissioner, Australia

Julia Fossi
Director, International, Strategy and Futures, Office of the eSafety Commissioner, Australia

Katherine Sessions
Senior Legal and Policy Advisor, Office of the eSafety Commissioner, Australia

Kelly Tallon
Senior Legal and Policy Advisor, Office of the eSafety Commissioner, Australia

5 Endnotes

1. United Arab Emirate Ministry of Health, *UAE public urged to join COVID-19 contact tracing app Alhosn to protect themselves, communities* [News Release], 20 May 2020, https://www.mohap.gov.ae/en/MediaCenter/News/Pages/2422.aspx.
2. Yuan, Eric, “A Message to Our Users, *Zoom*, 1 April 2020, https://blog.zoom.us/a-message-to-our-users.
3. Hay Newman, Lily, “The Zoom Privacy Backlash Is Only Getting Started”, *Wired, 1 April 2020,* https://www.wired.com/story/zoom-backlash-zero-days.
4. “FBI Warns of Teleconferencing and Online Classroom Hijacking During COVID-19 Pandemic”, *FBI, 30 March 2020,* https://www.fbi.gov/contact-us/field-offices/boston/news/press-releases/fbi-warns-of-teleconferencing-and-online-classroom-hijacking-during-covid-19-pandemic.
5. Cox, Joseph, “Zoom iOS App Sends Data to Facebook Even if You Don’t Have a Facebook Account”, *Vice, 26 March 2020,* https://www.vice.com/en_us/article/k7e599/zoom-ios-app-sends-data-to-facebook-even-if-you-dont-have-a-facebook-account.
6. Marczak, Bill and John Scott-Railton, “Move Fast and Roll Your Own CryptoA Quick Look at the Confidentiality of Zoom Meetings”, *The Citizen Lab, 3 April 2020,* https://citizenlab.ca/2020/04/move-fast-roll-your-own-crypto-a-quick-look-at-the-confidentiality-of-zoom-meetings.
7. Churchhill, Owen, “Zoom closes account of US-based Chinese dissidents after Tiananmen conference”, *South China Morning Post, 11 June 2020,* https://www.scmp.com/news/china/article/3088520/zoom-closes-account-us-based-chinese-dissidents-after-tiananmen.
8. Yuan, Eric, “A Message to Our Users, *Zoom*, 1 April 2020, https://blog.zoom.us/a-message-to-our-users.
9. Global Newswire, *Zoom Hires Jason Lee as Chief Information Security Officer [Press Release], 24 June 2020* http://www.globenewswire.com/news-release/2020/06/24/2052921/0/en/Zoom-Hires-Jason-Lee-as-Chief-Information-Security-Officer.html.

WORLD
ECONOMIC
FORUM

COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org