

A MEDI@4SEC Long Read

Social media, public safety and virtues

Marc Steen
TNO



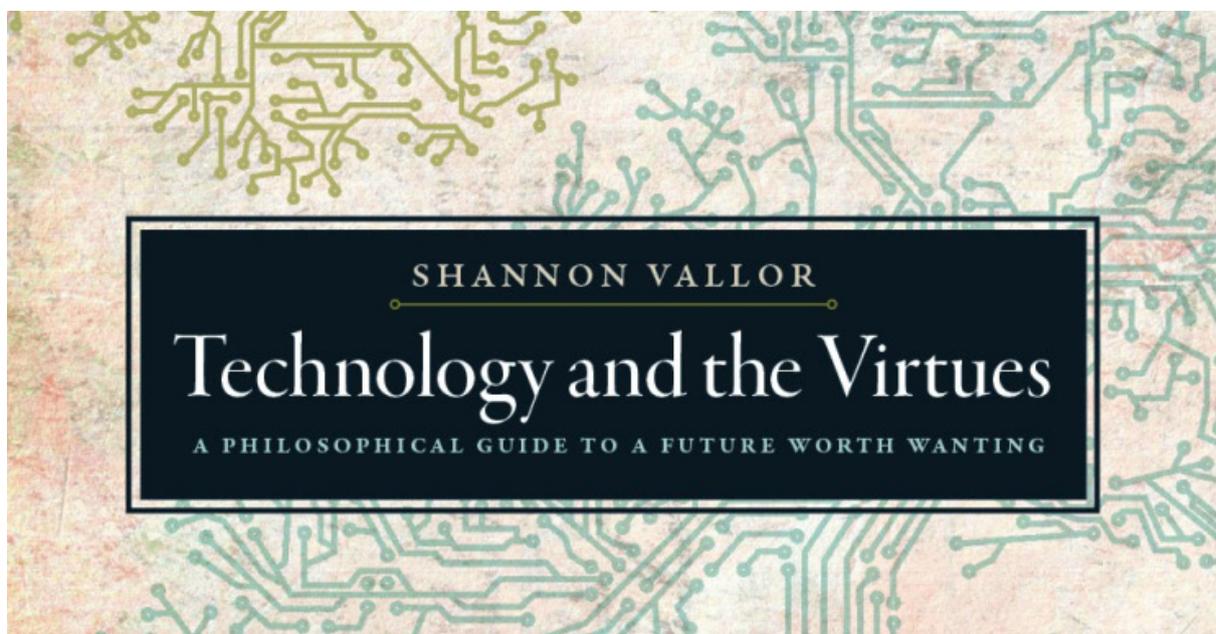
Virtue ethics can help to steer the development and usage of social media apps for public safety. It is especially helpful for developers. They can ask questions like “How could this app help people to cultivate, e.g., honesty, self-control, justice and civility?”. Careful reflection on such questions can help them to develop apps that support people in cultivating these very virtues.

In the MEDIA4SEC project we discussed numerous ways in which citizens and police officers can use social media to collaborate and co-create public safety. Moreover, we discussed diverse legal and ethical issues associated with their usage. Below, we will discuss four applications using virtue ethics as a lens to discuss technology.

Virtue ethics

Virtue ethics has its roots in ancient Athens of Aristotle. A virtue ethics approach raises questions like: *How can we create a just society?*—one in which people can experience, for example, freedom, equality and democracy (please note that such concepts have been updated since Aristotle’s time, when women had no rights and slavery was common). And: *How can people cultivate virtues that will help them to flourish?*—where virtues are understood as dispositions or patterns of thinking, feeling and acting—and indeed the *alignment* of thinking, feeling and acting.

In her book [*Technology and the Virtues*](#) (2016) Shannon Vallor, professor at Santa Clara University in Silicon Valley, proposed virtue ethics as a framework to discuss emerging technologies. She proposes to “investigate how various forms of emerging technologies, depending upon how we choose to develop and engage them, may enable or frustrate our



efforts to individually and collectively become virtuous: to make ourselves into the sorts of human being able to live truly *good* lives." (p. 159).

By way of experiment, we will discuss four social media applications: *Citizen* and *Neighbourhood Watch* as tools for do-it-yourself policing and collaboration between citizens and police; and *Echosec* and *Live Video Streaming* as tools for co-creating public safety in the context of mass gatherings. After briefly introducing these applications, we will discuss the ways in which they can help or hinder people in cultivating specific 'technomoal' virtues: *honesty*, *self-control*, *justice* and *civility* (Shannon Vallor discusses a total of eleven virtues, but these four are most relevant for our current purpose).

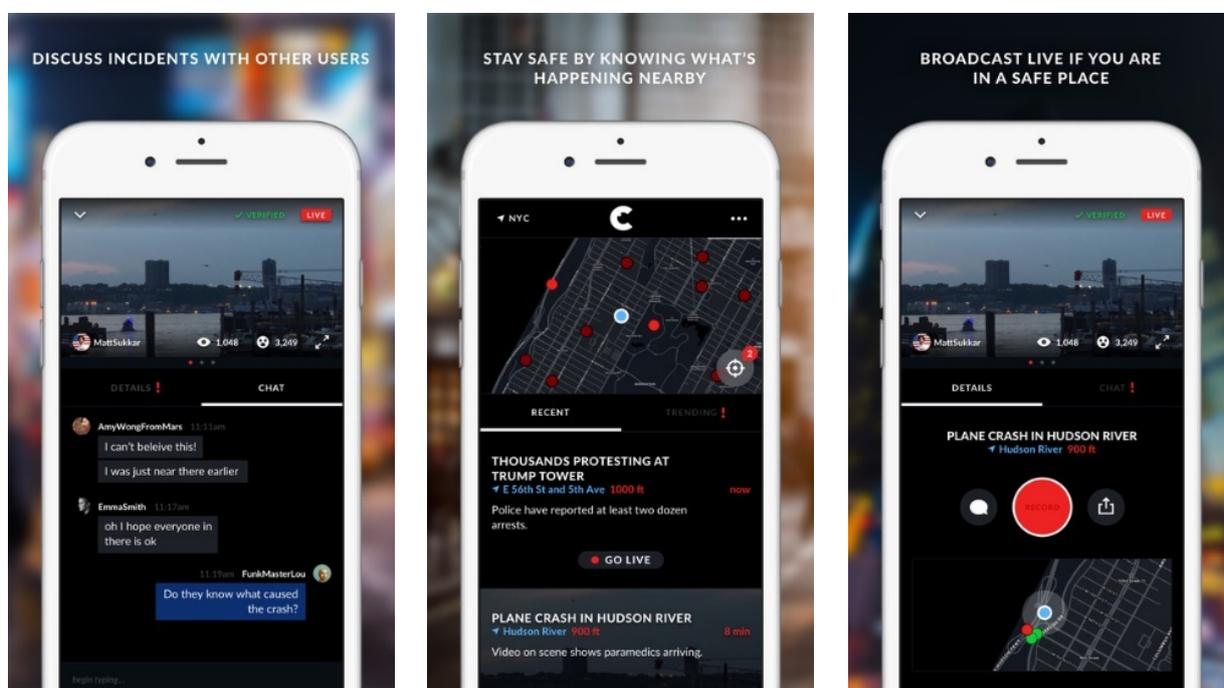
Citizen (see also: [D4.1](#), pp. 9-10)

This app was launched as *Vigilante* in October 2016 in New York City. The app "instantly sends an alert to those nearby whenever a crime is reported to 911. This information shows up on a map, so everyone can choose to avoid potential danger, or broadcast the incident live when it's safe to do so". A promotion video illustrates its usage ([vigilante-manifesto](#); <https://youtu.be/pe4BrBQxa8g>).

Soon after its launch the app was removed from the App Store because it violated the store's "strict rules about user-generated content apps to prevent abuse or bullying" and the "risk [of] physical harm, which could be possible if a load of vigilantes took their pitchforks to a crime scene" ([vigilante-app-removed-apple-store](#)). The app was, however, re-released in March 2017, with the name *Citizen* and some modifications, e.g., a pop-up that "provides stronger guidance to never approach a crime scene, interfere with an incident, or get in the way of police" and a new tagline: *Citizen alerts you of nearby crime and other emergencies so you can stay safe*.

Neighbourhood Watch (see also: [D4.1](#), p. 15)

Increasing numbers of people use social media like *WhatsApp*, as *Neighbourhood Watch*: to share information and communicate, e.g., about garbage lying on the street. In [The Netherlands](#), e.g., some 640.000 people use *WhatsApp* in over 7000 neighbourhoods for this purpose. As channels for information and communication these apps offer benefits. But they





also pose risks, e.g., for privacy and data protection (with sensitive data sitting in servers of companies abroad). Moreover, there are risks for bullying and personal vendettas, and of prejudices or biases propagating unchecked through these apps.

There are diverse ways to 'fix' this. E.g., the Dutch Police participate in WhatsApp meta-groups, which consist of citizen-moderators of multiple neighbourhood groups; that way, a police officer only needs to see and respond to those messages that a moderator relays to the meta-group. This requires new processes in the police.

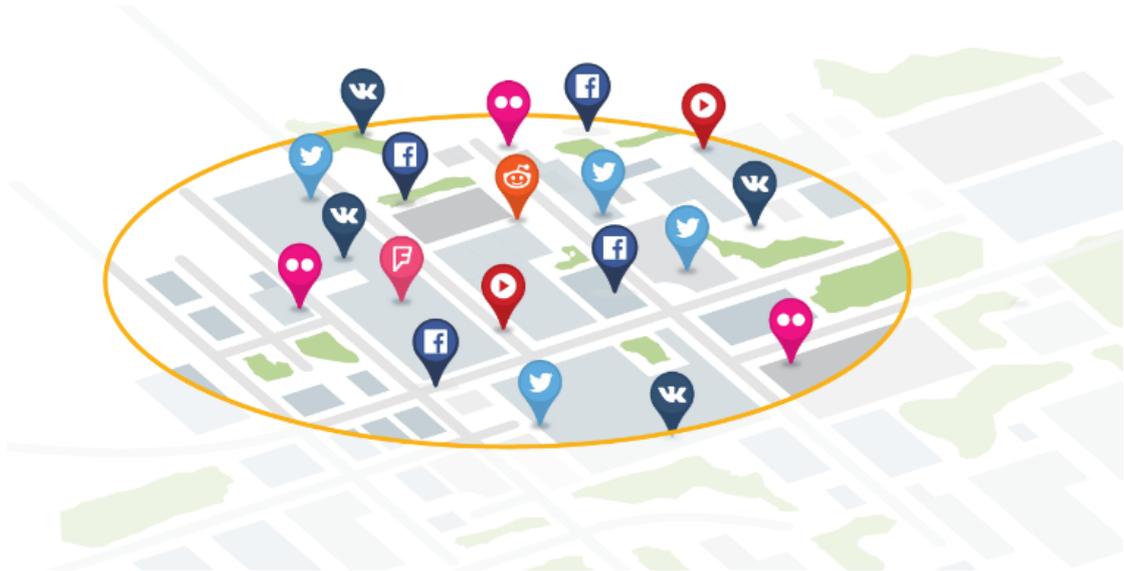
Echosec (see also in [D4.2](#), pp. 13-14)

Echosec is a system for gathering real-time, location-based publicly available data from social media, blogs and news sources, and analysing these data, using sentiment analysis, that is, the analysis of people's emotions. Practically, this system consists of tools that, e.g., a police force can use to monitor the sentiments of people in a mass gathering: within a specific time-frame and location.

There are diverse ethical issues associated with the usage of such a system. E.g., one cannot assume consent of individuals to non-anonymised, fine-grained monitoring of their social media activities by the police or public authorities—not merely based on them being active on social media. Moreover, such analyses can infringe on people's privacy, and can therefore only be used for legitimate public security purposes and only to an extent that is proportional to the security that is at stake.

Live Video Streaming (see also [D4.2](#), pp. 17)

There are multiple social media that people can use to create live video streams, e.g., [Periscope](#) (owned by Twitter) [YouTube Live](#) and [Facebook Live](#). These are generic tools that can be used for various purposes. People can, e.g., use these tools to for citizen journalism, to report on a mass gathering. This would be especially valuable in countries without a free press; citizens can then use social media like this to cover an event that would not be covered properly.



These tools, however, can also be used for less democratic purposes. People can, e.g., use live video streaming to present a biased or a one-sided version of an event. 'Fake news', if you will. Moreover, unedited live streaming of events may depict people that are perfectly bona fide in the context of violence or criminal behaviour—wrong time wrong place, which may cause them harm.

Ethical risks and benefits

Looking at *Citizen* and *Neighbourhood Watch*, a number of ethical risks can be identified (from [D4.1](#)): *vigilantism* (unjustified citizen interventions); *discrimination* (disproportionate and unjust attention to people from certain social or ethnic groups); and *information overload* (burdening the police with data, whose utility is burdensome to verify). On the positive side, applications like this can lead to: *active citizenship*; *increased trust* in authorities; and *better collaboration*, both between citizens and between citizens and the police (from [D4.1](#)).

Looking at *Echosec* and *Live Video Streaming*, we find the following ethical issues (from [D4.2](#)), e.g., the *risk of privacy intrusions* and the role of consent, and *the chilling effect* on legitimate exercise of democratic rights to freedom of political expression and association. On the positive side, applications like this can promote *effective communication* for the protection of individuals from violence, danger and harm, the *prediction of and prevention of disorder*, and a *peaceful exercise* of important democratic freedoms of political expression and association (from [D4.2](#)).

Now, let us look at the virtues of *honesty*, *self-control*, *justice* and *civility*, and discuss the ways in which social media applications like *Citizen*, *Neighbourhood Watch*, *Echosec* or *Live Video Streaming* can help or hinder people to cultivate them.

Honesty

One can use *Live Video Streaming* to cover an event as honestly as possible, e.g., by showing both sides of a story, by zooming-in on details and at the same time providing a larger picture. If used in such a way, usage of live video streaming can promote honesty. As noted above, however, these media can be used just as well to the opposite effect: to create biased or one-sided stories. This can very well happen unintentionally. If one has a stake in a situation, it can be rather challenging to also show another perspective. One can think of establishing guidelines for making videos: show both sides of an argument; show beginning and end of



an event; show both details and the larger picture. This will cultivate honesty and enable viewers to make up their minds.

When we look at *Echosec*, there is the risk of people who try to game the system. E.g., people of group A can discover how to influence the sentiment analysis, by spreading 'fake news' to trigger interventions that are in their interest, e.g., to disrupt authorities' activities or to harm people in group B.

Self-control

If we look at the *Citizen* app, one can easily imagine a person who fails to cultivate self-control. Many people find it difficult to resist the urge to check their social media apps, especially those apps that provide snippets of information that triggers their emotions and excite them. Tristan Harris, former design ethicist at Google and founder of the [Center for Humane Technology](http://www.tristanharris.com/essays/), warns us for the power of apps to hijack our attention, typically to satisfy a business model that aims to maximize the time people spend in the app (<http://www.tristanharris.com/essays/>). Social media apps like *WhatsApp* are no exception. They are designed to grab and hold people's attention—to 'engage' them.

While it would be impractical to require *WhatsApp* to make their app less addictive, we *can* introduce guidelines for wiser ways of using *WhatsApp*—a sort-of rules of engagement. Examples of such guidelines are: making agreements on which topics are in-scope, e.g., burglaries, and which topics are out-of-scope, e.g., litter—this will help to combat information overload; first validating one's observation before posting it, e.g., first check whether this neighbour is indeed on holidays—this will prevent invalid information, gossip and noise; and never sharing pictures from a closed *WhatsApp* group to open social media like Twitter or Facebook (see also: [Guidelines, in Dutch](#)).

Justice

Sharing information is so easy and so quick on social media, which results easily and quickly in people sharing biased, incorrect or outright discriminatory messages. There are many examples also of systems that propagate existing injustices, e.g., systems for 'predictive policing' or systems that 'predict' recidivism. These systems use data from the past, including existing biases and inequalities, which are subsequently reproduced in their outputs—unless

they are actively corrected. There are lots of examples of this in Cathy O’Neil’s book [Weapons of Math Destruction](#) and her [TED Talk](#).

It is, however, possible to design apps in ways that help people to cultivate justice. Let’s look at social media app *Nextdoor*, which is used also for Neighbourhood Watch purposes. Staff at *Nextdoor* found that people sent potentially racist messages, e.g., of a “suspicious person” walking around—only referring to the person’s skin colour. In order to combat ethnic profiling, they changed the app’s user interface: one now first gets this question: *Ask yourself: is what I saw actually suspicious after I take race or ethnicity out of the equation?* Moreover, users are required to describe hair, clothing top and bottom, shoes, age and build—and lastly, race (<https://www.wired.com/2016/08/nextdoor-breaks-sacred-design-rule-end-racial-profiling/>).

These extra steps in the user interface counter people’s also help to cultivate self-control.

The usage of tools for sentiment analysis can have a chilling effect on people’s fundamental rights. It can very well create obstacles for people to exercise their freedom of assembly and freedom of expression. Knowing that you are monitored affects one’s behaviour. This could lead to people not daring to show-up for a gathering or not daring to speak-up—and thereby corroding these freedoms.

Furthermore, it is inherent to systems like *Echosec* that they select data to focus on, e.g., data of specific groups of people. Either because they are at risk in a mass gathering. Or because they bring risks in a mass gathering. Such a focus and selection brings with it the risks of

1. Incident ✓ 2. People 3. Additional info

Next, add descriptions of any people involved.

ASK YOURSELF

What details can I add that will help distinguish this person from other similar people?

 Describe clothing from head to toe. Police say this is the most helpful to neighbors (and helps avoid suspecting innocent people.)

Optional details describing what any people were wearing and what they looked like...

Back Skip **Next**

discrimination. In order to uphold justice, an organization that uses these tools would need to act carefully and transparently, e.g., by informing the people of the groups involved, that their data will be selected and focussed on. Moreover, an organization will need to make itself accountable for the usage of such a system, e.g., by installing mechanisms for oversight.

Civility

Finally, social media can be used to cultivate the virtue of civility. This virtue is defined by Vallor as “a sincere disposition to live well with one’s fellow citizens ... to collectively and wisely deliberate about matters ... and to work cooperatively towards [common goods]” (2016: 141). Civility understood in this sense will require also virtues like empathy, courage and perspective. Civility depends upon people’s abilities to empathise with each other, to balance hope and fear, and to apply and combine different perspectives, e.g., individual and collective concerns or short-term and long-term interests.

Moreover, it is key to understand virtue ethics as being sensitive to people’s unique qualities and contexts. For a physically strong person it would be *courageous* to intervene in a fight between two men, e.g., to prevent further harm, especially when he or she has relevant skills. For a person with less physical abilities or skills, the courage would be to bring oneself into safety and to call the police. Civility in a *Neighbourhood Watch* app requires sensitivity and consideration in deciding which information to share with whom.

Looking at apps like *Echosec* or apps for *Live Video Streaming*, the cultivation of civility would require that these tools and apps can be used to support a willingness to collaborate, to engage in dialogues about issues that matter, and find ways to work cooperatively towards creating common goods.

Sherry Turkle, professor at MIT, has studied people’s usage of computers, internet, smart phones and apps for over 30 years. In her most recent book, *Reclaiming Conversation*, she offers a diagnosis of the problem of social media: our obsession for being always ‘on’ and connected made us forget the benefits of solitude (we’ve become afraid for boredom), we forgot how to connect to others (we’ve lost basic skills for making conversations; we’re avoiding ‘awkward’ conversations and thus have lost basic skills for empathy), and we forgot how to organize conversations in groups (many people are staring at their screens). Turkle observes that the key problem, that is a lack of conversations, can be solved by exactly that what’s lacking, by having better conversations.

These observations can provide clues for the design of tools and apps that can help people to cultivate virtues that facilitate conversations: empathy and courage, e.g., learning to connect to others, also when they have not much in common with us at first glance, or perspective, e.g., learning about ‘the other side of the story’.

Conclusions

We have shown several ways in which the developers of social media applications can view their apps as *tools* that people can use to cultivate virtues like *self-control*, *honesty*, *justice* and *civility*. Moreover, with relatively small efforts they can steer the development of their applications in directions that help to create a just society in which people can flourish—in which citizens and the police can collaborate in co-creating public safety.

[rise-of-live-video-streaming; : https://blog.nextdoor.com/2016/04/26/continued-improvements-to-how-our-members-post-about-crime-and-safety/](https://blog.nextdoor.com/2016/04/26/continued-improvements-to-how-our-members-post-about-crime-and-safety/).



The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under Grant Agreement no 700281

