Soft Ethics and the Governance of the Digital

Family Name   Floridi

Given Name   Luciano

[two affiliations, please list both]:

Division   Oxford Internet Institute

Organization   University of Oxford

Address   1 St Giles, Oxford, OX1 3JS, United Kingdom

Division   [none]

Organization   The Alan Turing Institute

Address   96 Euston Road, London, NW1 2DB, United Kingdom

Email   luciano.floridi@oii.ox.ac.uk

Today, in any mature information society (Floridi 2016), we no longer live online or offline but *onlife*, that is, we increasingly live in that special space, or *infosphere*, that is seamlessly analogue and digital, offline and online. If this seems confusing, perhaps an analogy may help to convey the point. Imagine someone asks whether the water is sweet or salty in the estuary where the river meets the sea. Clearly, that someone has not understood the special nature of the place. Our mature information societies are growing in such a new, liminal place, like mangroves flourishing in brackish water. And in these "mangrove societies", machine-readable data, new forms of smart agency, and onlife interactions are constantly evolving, because our technologies are perfectly fit to take advantage of such a new environment, often as the only real natives. As a result, the pace of their evolution can be mind-blowing. And this in turn justifies some apprehension. However, we should not be distracted by the scope, depth and pace of technological innovation. True, it does disrupt some deeply ingrained assumptions of the old, exclusively analogue society, e.g. about production, logistics, customization, competition, education, work, health, entertainment,

politics, and security, just to mention some crucial topics. Yet that is not the most consequential challenge we are facing. It is rather how we are going to design the infosphere and the mature information societies developing within it that matters most. Because the digital revolution transforms our views about values and their priorities, good behaviour, and what sort of innovation is socially preferable—and this is the fundamental issue. Let me explain.

To many, what digital innovation will throw up next may seem the real challenge. The question itself is recurrent: what comes next? What is the next disruption? What is the new killer app? Will this be the year of the final battle between Virtual Reality vs. Augmented Reality? Or is it the Internet of Things that will represent the new frontier, perhaps in some combination with Smart Cities? Is it the end of the TV as we know it coming soon? Will healthcare be made unrecognisable by machine learning, or should our attention rather be focused on the automation of logistics and transport? What will the new smart assistants in the home do, apart from telling us what the weather is like, and allowing us to choose the next song? Behind similar questions lies the unspoken assumption that technological innovation leads, and everything else follows, or lags behind: business models, working conditions, standards of living, legislation, social norms, habits and expectations. Yet this is precisely the distracting narrative we should resist. Not because it is wrong, but because it is only superficially right. The deeper truth is that the revolution has already occurred: the transition from an entirely analogue and offline world to one that is increasingly also digital and online will never happen again in the history of humanity. One day, a quantum computing phone running artificial intelligence apps may be in the pocket of your average teenager, but our generation is the last one that will have seen a non-digital world. And this is the really extraordinary turning point. Because that landing in the infosphere happens only once. What this new world will be like, as we create it, is both fascinating, in terms of opportunities, and worrisome, in terms of risks. But the "exploration" of the infosphere, to indulge in the geographical metaphor a bit longer, no matter how challenging, prompts a much more fundamental question, which is socio-political and truly crucial: what kind of mature information societies do we want to build? What is our human project for the digital age? Looking at our present backwards—that is, from the future to our present—this is the time in history when we shall be seen to have laid down the foundation for our mature information societies. We shall be judged by the quality of our work. So, clearly, the real challenge in no longer digital innovation, but the *governance* of the digital.

The proof that this is the case is all around us, in the many initiatives addressing the impact of the digital on everyday life and how to regulate it. It is also implicit in the current narrative on the unstoppable and unreachable nature of technological innovation, if one looks just a bit more closely. In the same context where people complain about the speed of innovation and the impossible task of chasing it with some normative framework, one finds that there is equal certainty about the serious risk that the wrong legislation may kill innovation entirely, or destroy whole technological sectors and developments. You do not have to be Nietzsche ("Was mich nicht umbringt macht mich stärker" – "what does not kill me makes me stronger" (Nietzsche 2008)) to realise that the inference to be drawn is that updating the rules of the game is perfectly possible—it can have immense consequences—but that reacting to technological innovation is not the best approach. We need to shift from chasing to leading. If then we like the *direction* in which we move, or *where* we want to go, then the *speed* at which we are moving or getting there can actually be something very positive. The more we like where we are going the feaster we will want to get there. But for this to happen society needs to stop playing defence and start playing attack. The question is not whether, but how. And to start addressing the how, some clarifications are helpful.

On the governance of the digital there is much to be said, and even more still to be understood and theorised, but one point is clear: the governance of the digital (henceforth *digital governance*), the ethics of the digital (henceforth *digital ethics*, also known as computer, information, or data ethics (Floridi and Taddeo 2016)) and the regulation of the digital (henceforth *digital regulation*) are different normative approaches, complementary, but not to be confused with each other, in the following sense (see Figure 1 for a visual representation).

Digital governance is the practice of establishing and implementing policies, procedures, and standards for the proper development, use, and management of the infosphere. It is also a matter of convention and good coordination, sometimes neither moral or immoral, neither legal or illegal,. For example, through digital governance a government agency or a company may (a) determine and control processes and methods used by data stewards and data custodians in order to improve the data quality, reliability, access, security, and availability of its services; and (b) devise effective procedures for decision-making and for the identification of accountabilities with respect to data-related processes. A typical application of digital governance was the work I co-chaired for the British Cabinet Office in 2016 on a "Data Science Ethical Framework" (Cabinet Office

2016), which was "[…] intended to give civil servants guidance on conducting data science projects, and the confidence to innovate with data."[1]

Digital governance may comprise guidelines and recommendations that overlap with, but are not identical to, *digital regulation*. This is just another way of speaking about the relevant legislation, a system of rules elaborated and enforced through social or governmental institutions to regulate the behaviour of the relevant agents in the infosphere. Not every aspect of digital regulation is a matter of digital governance and not every aspect of digital governance is a matter of digital regulation. In this case a good example is provided by the General Data Protection Regulation (GDPR).[2] *Compliance* is the crucial relation through which digital regulation shapes digital governance.

All this holds true of *digital ethics*, understood as the branch of ethics that studies and evaluates moral problems relating to *data* and *information* (including generation, recording, curation, processing, dissemination, sharing, and use), *algorithms* (including AI, artificial agents, machine learning, and robots), and corresponding *practices* and *infrastructures* (including responsible innovation, programming, hacking, professional codes, and standards), in order to formulate and support morally good solutions (e.g., good conduct or good values). Digital ethics shapes digital regulation and digital governance through the relation of moral evaluation.

---

[1] Available from https://www.gov.uk/government/publications/data-science-ethical-framework

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJEU L119, 04/05/2016.
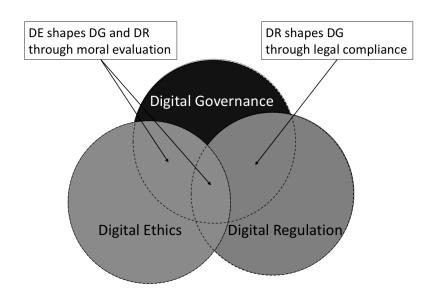
Figure 1 The relationship between digital ethics, digital regulation and digital governance

Once the map is understood, several consequences become clear.

First, there is the synecdoche use. Digital governance in Figure 1 is just one of the three normative forces that can shape and guide the development of the digital. But it is not uncommon to use that part for the whole (a bit like using "coke" for any variety of cola), and to speak of digital governance as refering to the whole set. It is what I did at the beginning of this article, when I stated that the real challenge today is the governance of the digital. By that I meant to refer not just to digital governance but also to digital ethics and digital regulation, i.e., to the whole normative map. And this is also how I interpret the report "Data Management and Use: Governance in the 21st Century" that we published in 2017 as a joint British Academy and Royal Society working group (British Academy 2017). As long as the synecdoche is clear, there is no problem.

Second, when policy-makers, both in political and in business contexts, wonder why we should engage in moral evaluation when legal compliance is already available (this is a recurring topic in the discussion of the GDPR), the answer should be clear: compliance is necessary but insufficient to steer society in the right direction. Because digital regulation indicates what the legal and illegal moves in the game are, so to speak, but it says nothing about what the *good* and *best* moves could be to win the game—that is, to have a better society. This is the task of both digital ethics, on the side of moral values and preferences,

and of digital governance, on the side of best management. This is why, for example, the European Data Protection Supervisor (the EU's independent data protection authority) established the Ethics Advisory Group in 2015, in order to analyse the new ethical challenges posed by digital developments and current legislation, especially in relation to the GDPR. The report we published (EDPS Ethics Advisory Group 2018) should be read as a contribution to, and a stepping stone towards, a normative governance of the infosphere in the EU.

Third, digital ethics may be understood now in two ways, as *hard* and *soft ethics*. Hard ethics is what we usually have in mind when discussing values, rights, duties and responsibilities—or, more broadly, what is morally right or wrong, and what ought or ought not to be done—in the course of formulating new regulations or challenging existing ones. In short, hard ethics is what makes or shapes the law. Thus, in the best scenario, lobbying in favour of some good legislation or to improve that which already exists can be a case of hard ethics. For example, hard ethics helped to dismantle apartheid in South Africa and supported the approval of legislation in Iceland that requires public and private businesses to prove that they offer equal pay to employees, irrespective of their gender (the gender pay gap continues to be a scandal in most countries).

Soft ethics covers the same normative ground as hard ethics, but it does so by considering what ought and ought not to be done *over and above* the existing regulation, not against it, or despite its scope, or to change it, or to by-pass it (e.g. in terms of self-regulation). In other words, soft ethics is *post-compliance ethics*: in this case, "ought implies may". Now, both hard and soft ethics usually presuppose *feasibility* or, in more Kantian terms, assume that "ought implies can", given that an agent has a moral obligation to perform an action only if this is possible in the first place. It follows that soft ethics also assumes a *post-feasibility* approach. Add that any ethical approach, at least in the EU, accepts, as its minimal starting point, the implementation of the Universal Declaration of Human Rights (UDHR) and The Charter of Fundamental Rights of the European Union (within the UDHR, there are rights that may assume greater relevance than others, such as article 3 on security, article 7 on non-discrimination, article 12 on privacy, article 19 on freedom of opinion and expression, and article 20 on freedom of association). And the result is that the space of soft ethics is both partially bounded, and yet unlimited. To see why, it is easy to visualise it in the shape of a trapezoid (see Figure 2), with the lower side representing a feasibility base that is ever-expanding through time—we can do more and more things—the two constraining sides representing legal compliance and human rights, and the open

upper side representing the space where what is morally good may happen in general and, in the context of this article, may happen in terms of shaping and guiding the development of our mature information societies.
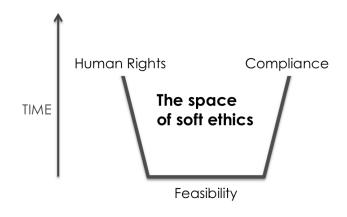

Figure 2 The space of soft ethics

Soft digital ethics can be rightly exercised in places of the world where digital regulation is already on the good side of the moral vs. immoral divide. But it would be a mistake to argue for a soft ethics approach to establish a normative framework when agents (especially governments and companies) are operating in contexts where human rights are disregarded, e.g. in China, North Korea or Russia, or in contexts where hard ethics is precisely what is needed to change the current regulation, e.g. in the United States and, *a fortiori*, in the three countries already mentioned. It is really within the European Union (EU) that soft ethics can rightly be exercised, to help companies, governments and other organisations to take more and better advantage, morally speaking, of the opportunities offered by digital innovation. Because even in the EU, legislation is necessary but insufficient. It does not cover everything (nor should it), and agents should leverage digital ethics in order to assess and decide what role they wish to play in the infosphere, when regulations provide no simple or straightforward answer, when competing values and interests need to be balanced (or indeed when regulations provide no guidance), and when there is more that can be done over and above what the law strictly requires. This is why

it is in the EU that a good use of soft ethics could lead to "good corporate citizenship" within a mature information society.

Fourth, given the open future addressed by digital ethics, it is obvious that *ethics foresight analysis*—based on data analytics applied strategically to the ethical impact assessment of digital technologies, goods, services, and practices (see Figure 3)—must become a priority (Floridi 2014). For the task of digital ethics is not simply to "look into the [digital] seeds of time / And say which grain will grow and which will not" (*Macbeth*, I.3, 159-162), it also seeks to determine which ones *should* grow and which should not.

Figure 3 Ethics foresight analysis cycle

Or to use a metaphor already introduced above, the best way to catch the technology train is not to chase it, but to already be at the next station. In other words, we need to anticipate and steer the ethical development of technological innovation. And we can do this by looking at what is actually feasible, privileging, within this, what is environmentally sustainable, then what is socially acceptable and then, ideally, choose what is socially preferable (see Figure 4).
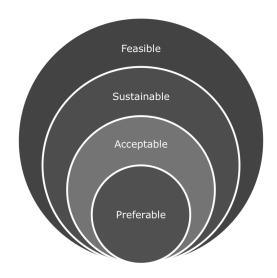
Figure 4 Digital ethics impact assessment

We do not yet have an infosphere equivalent for the concept of *sustainability* of the biosphere, so our current equation is incomplete (see Figure 5).

$$\text{biosphere} : \text{sustainability} = \text{infosphere} : x$$

Figure 5 A difficult equation to balance

In Figure 4, I suggested that we interpret the x as social "preferability" but I am aware it is just a place holder for a better idea to come. This may take a while, given that "the tragedy of the commons" was published in 1968 but the expression "sustainable development" was only coined by the Brundtland Report almost twenty years later, in 1987 (Brundtland 1987). Yet the lack of conceptual terminology does not make the governance of the digital a mere utopian effort. In particular, digital ethics, with its values, principles, choices, recommendations and constrains, already influences the world of technology much more than any other force. This because the evaluation of what is morally good, right, or necessary shapes public opinion—hence the socially acceptable or preferable and the politically feasible, and so, ultimately, the legally enforceable, and what agents may or may not do. In the long run, people (as users, consumers, citizens, patients, etc.) are constrained in what they can or cannot do by organisations, e.g. businesses, which are constrained by law, but the latter is shaped and constrained by ethics, which is where

people decide in what kind of society they want to live (see Figure 6). Unfortunately, such a normative cascade becomes obvious mainly when backlash happens, i.e. mostly in negative contexts, when the public rejects some solutions, even when they may be good solutions. A normative cascade should instead be used constructively, to pursue the construction of a mature information society of which we can be proud.
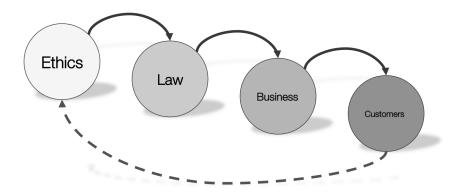

Figure 6 Example of a normative cascade, with business as agent and people as customers. Business could be replaced by government and people by citizens.

From all this it follows that ethics in general and digital ethics in particular, cannot be a mere add-on, an afterthought, a late-comer, an owl of Minerva that takes its flight only when the shades of night are gathering—once digital innovation has taken place, and possibly bad solutions have been implemented, less good alternatives have been chosen, or mistakes have been made. Nor can it be a mere exercise of questioning. The building of critical awareness is important but it is also only one of the four tasks of a proper ethical approach to the design and governance of the digital. The other three are signalling that ethical problems matter, engaging with stakeholders affected by such ethical problems, and, above all, providing sharable solutions. Any ethical exercise that in the end fails to provide some acceptable recommendations is only a preamble. So digital ethics must inform strategies for the development and use of digital technologies from the very beginning, when changing the course of action is easier and less costly, in terms of resources and impact. It must sit at the table of policy-making and decision-taking procedures from day one. For we must not only think twice but, most importantly, we must think *before* taking important steps. This is particularly relevant in the EU, where I have argued that soft ethics can be properly exercised and where SETI (science,

engineering, technology and innovation) developments are crucial. If soft digital ethics can be a priority anywhere, this is certainly in Europe. We should adopt it as soon as possible.

References

British Academy, The Royal Society. 2017. Data management and use: Governance in the 21st century - A joint report by the British Academy and the Royal Society.

Brundtland, Gro Harlem. 1987. The Brundtland Report, World Commission on Environment and Development. Oxford University Press.

Cabinet Office, Government Digital Service. 2016. Data Science Ethcal Framework.

EDPS Ethics Advisory Group. 2018. Towards a digital ehics.

Floridi, Luciano. 2014. "Technoscience and Ethics Foresight." *Philosophy & Technology* 27 (4):499-501.

Floridi, Luciano. 2016. "Mature Information Societies—a Matter of Expectations." *Philosophy & Technology* 29 (1):1-4.

Floridi, Luciano, and Mariarosaria Taddeo. 2016. "What is data ethics?" *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2083).

Nietzsche, Friedrich Wilhelm. 2008. *Twilight of the idols, or, How to philosophize with a hammer.* Oxford: Oxford University Press.