Problem-solving? No, problem-opening!

A method to reframe and teach data ethics as a transdisciplinary endeavour

Calzati, Stefano; Ploeger, Hendrik

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Problem-solving? No, problem-opening! A method to reframe and teach data ethics as a transdisciplinary endeavour

Stefano Calzati[1] 🆔 and Hendrik Ploeger[1] 🆔

Abstract

Starting from the recognition of the limits of today's common essentialist and axiological understandings of data and ethics, in this article we make the case for an ecosystemic understanding of data ethics (for the city) that accounts for the inherent value-laden entanglements and unintended (both positive and negative) consequences of the development, implementation, and use of data-driven technologies in real-life contexts. To operationalize our view, we conceived and taught a master course titled 'Ethics for the data-driven city' delivered within the Department of Urbanism at the Delft University of Technology. By endorsing a definition of data as a sociotechnical process, of ethics as a collective practice, and of the city as a complex system, the course enacts a transdisciplinary approach and problem-opening method that compel students to recognize and tackle the unavoidable multifacetedness of all ethical stances, as well as the intrinsic open-endedness of all tech solutions, thus seeking a fair balance for the whole data-driven urban environment. The article discusses the results of the teaching experience, which took the form of a research-and-design workshop, alongside the students' feedback and further pedagogical developments.

Keywords

Data ethics, city, teaching, transdisciplinarity, ecosystem, sociotechnical

Today, the field of ethics has become central to the discussion on the development, implementation, and use of data-driven technologies.[1] Acknowledging the disruptive impact that these technologies can have on people as both individuals and collectives, in fields as diverse as (but not limited to) medicine, mobility, law, social behavior, literacies and education, governance, urban development (Bigman and Grey, 2018; Calzati and van Loenen 2023; Calzati and van Loenen 2023b; Fotopoulou, 2021; Gillespie et al., 2023; Taylor, 2020; Viljoen, 2021), ethics has been increasingly regarded as the discipline to look at to identify the boundaries as well as the compass for defining what is morally acceptable when it comes to the development, implementation, and use of these technologies (Floridi et al., 2021; Mittelstadt et al., 2016; Murillo et al., 2023).

Any literature review that would aim to cover scholarly publications over the last 15 years linking ethics to data-driven technologies (henceforth, 'data ethics')[2] can hardly overestimate the breadth of the topic, from ethics and computer science (Himma and Tavani, 2008; Horton et al., 2022; Quinn, 2006) to ethics and AI (Coeckelbergh, 2020; Jobin et al., 2019; Siau and Wang, 2020), passing through an ethics-infused governance for/through these technologies

(Floridi, 2018; Kasapoglu et al., 2021; van Maanen, 2022). Here we do not aim to conduct such comprehensive literature review, but rather to unravel the roots of the link between data and ethics because the way in which these terms are widely coupled in literature rests on a (more or less unacknowledged) essentialist and axiomatic ground which ultimately oversimplifies the issue at stake – how can we do good with/through data-driven technologies? – and hinders an effective tackling of the moral dilemmas that data-driven technologies pose when adopted in complex real-life scenarios.

Notably, we do so by presenting and reflecting upon the teaching experience of an elective course on 'Ethics for the data-driven city' within the Master program in Geomatics at Delft University of Technology (TU Delft). In this way, the

[1]Urban Data Science Group, Department of Urbanism, Delft University of Technology, Delft, Netherlands

Corresponding author:
Stefano Calzati, Urban Data Science Group, Department of Urbanism, Delft University of Technology, Delft, Netherlands
Email: s.calzati@tudelft.nl

article outlines the contours of a sociotechnical, non-axiological and ecosystemic approach to data, ethics, and the city, which substantially departs from the mainstream understanding of these terms, opening the way to a radical shift in the way of teaching data ethics, especially in/for the urban environment.

The article is organized as follows: Section "Data ethics: an overview" provides an overview on ethics in connection with today's data-driven technologies and the digital society at large; Section "The teaching of data ethics: limits and prospects" zooms in on the teaching of data ethics, unveiling the limits of the current understanding of both ethics and data-driven technologies; Section "Teaching of data ethics in/for the urban environment" advances an ecosystemic approach to data and ethics in the context of the urban environment; Section "Designing the course 'ethics for the data-driven city'" presents how such conceptualization found an enactment in the research-and-design elective course 'Ethics for the data-driven city', organized and taught conjointly by the authors of this article. Section "Results and discussion" discusses the results of the teaching experience for the years 2022/2023, including feedback from the students. Section "Conclusion" reflects upon the limitations and possible further developments for the teaching of data ethics, as well as links the course to the theoretical part of the article.

Data ethics: an overview

A good starting point for the present discussion is Floridi and Taddeo's (2016: 3) seminal work on data ethics, in which the authors define data ethics as follows:

> the branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values).

On the one hand, this definition has the merit to be tech-agnostic, encompassing a broad set of innovations, from data life cycle to algorithms and AI, as well as diverse (technical and non-technical) practices. In this respect, the definition acknowledges the heterogeneity that the broad term 'data' is meant to embrace – and this is also how we intend to treat it as well. Being at the basis of today's algorithm-based tech innovation, data are the pillars of any digital technology. On the other hand, however, in the last part this same definition reveals an axiological (good vs. bad) and normative (do's vs. don'ts) standpoint towards ethics, which (a) assumes the very possibility of

identifying morally good solutions and (b) endorses, if not prescribes, the adoption of these solutions. This standpoint towards ethics, then, is somewhat detached from a cognizant social contextualization. To be sure, such detachment is subservient to an operationalization of data ethics in the form, for instance, of guidelines, recommendations, codes, or best lessons. Indeed, numerous initiatives by a variety of stakeholders – e.g., research centers, national and international bodies, and the private sector – have blossomed in recent years with the aim to release ethically robust frameworks and practices within and through which to make sure that emerging data-driven technologies are developed and used responsibly (e.g. European Commission, 2019; IEEE, 2017; Montreal Declaration, 2017). Yet, criticism has been raised both against the quantitative proliferation of these initiatives, which creates a fragmented and partially irreconcilable scenario (Floridi et al., 2021) and produces forms of tokenization of ethics to the benefit of the industry (Metcalf et al., 2019), as well as against the inefficacy of the released documents, which often translate into forms of 'ethics washing' (Floridi, 2019) lacking a concrete political impact (van Maanen, 2022).

More radically, the axiomatic and normative polarization to which ethics is subjected risks triggering an oversimplification of the scenario at stake that does not do justice to the complex non-zero-sum moral dilemmas[3] arising at the intersection of ethics and data. The assumption that it *is* possible to identify and endorse morally good solutions in objective terms rests on a positivist framing that trivializes not only ethics, but also the thick sociocultural fabric in which data-driven technologies operate. Far from being reducible to 'good' and 'bad' labels, ethics identifies, rather, a diffused ethos (Mittelstadt, 2019). As such, it shall be regarded as a practice and method of enquiry instead of a toolbox: as Markham et al. (2018: 2) note, indeed, 'although ethics is often considered a philosophical stance that precedes and grounds action, it is a value-rationality that is actually produced, reinforced, or resisted through practice'. This also means that it is not sufficient to recognize 'ethical pluralism' (Ess, 2006) in connection with the implementation and use of data-driven technologies in context, that is, the different ethical stances and values[4] held by different subjects whenever a new technology is adopted. Ethical pluralism, indeed, assumes a *positionalist* standpoint according to which it becomes eventually possible to harmonize these different stances and values once they are explicitly accounted for. And yet, to map these positions might not be that straightforward nor sufficient. Whenever data-driven technologies are implemented and used in a given context, to arise are value-laden entanglements that is difficult if not impossible to disentangle, insofar as the same subject might well embed mutually exclusive values at once (cf. Pink et al., 2024). In this respect, the emerging field of behavioral

ethics has shown the multifaceted nature of people's decision-making processes whenever they are confronted with moral dilemmas: not only 'people of good character may do bad things because they are subject to psychological shortcomings or overwhelmed by social pressures' (Prentice, 2014: 325), but most actions fall in a gray area that resists a clear-cut discernment in terms of 'good vs. bad'.

A step towards the mitigation of the axiomatic and normative standpoint on ethics comes from value-sensitive design, an approach to the development of technological artefacts concerned with keeping ethics in the loop. The idea of designing technology through ethics considers ethics not as something that can be given a priori, but as a continuous process: since values are inevitably engrained into technology, value-sensitive design is concerned with shaping the building process of technology by keeping ethics as a compass all along (d'Aquin et al., 2018; Shilton, 2012). Through this *instrumentalist* standpoint, the link between data and ethics is turned on its head and the focus shifts from technology's implementation and use to its development. At stake is no longer how to do good with/through technology, but how ethical values can guide the development of, and can be incorporated into, technology.

For one thing, value-sensitive design has the merit of opening up ethics as a process; yet, on a closer look this approach rests on a mirroring flaw to the axiomatic and normative understanding of ethics, endorsing an essentialist understanding of technology which remains blind over the complexity of technological artefacts. Stahl (2022: 72) summarizes this nicely arguing that an over-stress on 'a particular piece of technology' overlooks the fact that 'ethical, social, human rights and other issues never arise from a technology per se, but result from the use of technologies by humans in societal, organizational and other setting'. For instance, when value-sensitive approaches (Brey, 2000; Tavani, 2016) attempt to identify a practice or a technological feature that is controversial from a moral perspective, technology gets implicitly reified as something that can be isolated in terms of either features or practices. Instead, as Hasselbalch (2019: 2) writes, 'we should perceive data ethics initiatives as open-ended spaces of negotiation embedded in complex socio-technical dynamics, which respond to multifaceted challenges extended over time'. The solutionist idea of fixing technology – e.g. its biases or uses – while isolating it from its own genesis and context is doomed to produce unaccounted chains of externalities against individuals, collectives, or the environment, which do bear a core ethical dimension. Hence, it becomes crucial to move towards an approach to data ethics that is able to account for the socio-technical entanglements between people and technology. As we will see, to set the goal of designing socially beneficial data-driven technologies is not enough, because it is the very possibility of designing such technologies and

identifying in abstracted decontextualized terms what is beneficial, that gets challenged.

To be sure, literature in this direction, especially in Science and Technology Studies (STS) and Critical Data Studies, is already present and it is along this line that the present article follows. For instance, Tallacchini (2009: 282) unveiled the process of institutionalization of ethics as an 'isolated item' coopted 'in the production of "valid ethical knowledge" as a datum no further exposed to debate'. Along the same line, Casiraghi (2023) displayed the continuities and discontinuities between the challenges historically faced by ethics in the life sciences with those currently faced against the emergence of AI. Ultimately, Casiraghi's review criticizes today's 'exclusion, from ethical debates, of people and groups interested in discussing *ends* rather than how to develop ethical frameworks and principles to achieve, ex-post, pre-defined ones' (11). Leonelli (2019), in turn, endorsed a sociotechnical perspective to defend the inherent contextuality and interpretative openness of all data. In this sense, she speaks of a 'relational' view, whereby 'the meaning assigned to data depends on their provenance, their physical features and what these features are taken to represent' (5). This relational view resonates with the standpoint maintained here according to which any assessment of the (supposed) goodness and/or badness of data-driven technologies is an issue of framing, that is, of the perspective from which to look at the human-technology entanglement (cf. Meng, 2021). Most importantly, we move away from a positionalist stance – at stake is more than the mapping of data journeys across sciences – to foreground, instead, the collective-level dimension of data-driven technologies' development, use, and impact, with the goal to unveil the inevitable value-laden tradeoffs at stake and un/intended consequences. In this respect, as we will explain, we endorse an ecosystemic standpoint and a 'republican' (Calzati and van Loenen, 2023a) governance model for the fair (i.e. balanced) management of data-driven initiatives in/for the urban environment.

The teaching of data ethics: limits and prospects

The essentialist-axiomatic understanding of data ethics finds a reflection in the literature concerned with the teaching of this topic.

The need for ethics education within engineering-related disciplines, such as computer science, geomatics, mechanical engineering, has a long track record (Johnson, 1985). Currently, however, a proper integration of ethics into hard-science curricula remains scarce (Oliver and McNeil, 2021). One major challenge when it comes to the teaching of ethics in connection with data is the necessity to strike the right interdisciplinary blend between philosophical and sociological standpoints, on the one hand, and engineering-related concerns, on the other. Data ethics compels the humanities and

hard sciences to meet halfway, requiring both teachers and students to attune to and familiarize with concepts, approaches, and methodologies that, more often than not, fall out of their scholarly comfort zone. Not rarely, this challenge translates into tailored courses that attempt to incorporate ethics as an element of a broader engineering landscape. As Furey and Martin (2019: 14) write, these courses 'help students recognize the structure of ethical problems and their solutions [through] key ethical problem-solving tool [and] highly idealized hypothetical cases designed to test ethical theories and to isolate relevant moral variables'. To expose students to the ethical dilemmas embedded in their (future) work is certainly relevant (Johnson, 2017). Technology is as much a technical as a societal affair. Yet, most courses for these disciplines and students betray an axiomatic understanding of ethics which, as seen, falls short of accounting for the value-laden entanglements at stake (cf. Borenstein and Howard, 2021). Just to remain focused on the quote above, words such as 'solutions', 'problem-solving', and 'moral variables' entail a mechanicist conceptualization of the issues to be taught which tends to overlook the insolvability of most moral dilemmas and the unavoidable externalities connected with the implementation and use of data-driven technologies. In other words, in these courses ethics gets operationalized into a set of (value- and principle-) tools to be handled when in need, often relegating ethics to stand-alone classes (Fiesler et al., 2020). Discussing the teaching of data literacies, Kilhoffer et al. (2023: 13) note that the tendency is to reduce ethics to what can be objectively assessed, in this way sanitizing the collective political dimension of ethics, well-illustrated by their observation: 'while some teachers are required to teach cyber hygiene, they often treat it as a box-ticking exercise'.

If, on the one hand, ethics tends to be instrumentalized, on the other hand it is technology to be essentialized, through courses that aim to teach how to design ethically robust technologies. For instance, concerning the development of AI, it is common to encounter attempts that either provide a mathematical definition of fairness, based on metrics against which algorithms can be tested on a pass/fail basis (cf. Kleinberg et al., 2016), and/or that apply fairness to this or that technology or data process, so that these can be objectively evaluated on a contextual basis (Lee et al., 2021). Either case, students are taught that it is possible to develop technology to make it behave ethically. From this premise, it does not surprise that the proliferation of courses on data ethics focus on aspects such as fairness or accountability (Mittelstandt et al., 2016) which, having a legal dimension attached to them, seem to be more easily operationalizable. As Hagendorff (2020: 103) writes 'most frequently mentioned aspects are those for which technical fixes can be or have already been developed'. In this way, the axiomatic understanding of ethics and the essentialist understanding of technology enter a self-reinforcing loop: ethics is reduced to a binary affair which, in turn, technology can help fix by being boosted

according to a good versus bad logic of functioning. Most importantly, and problematically, this loop eschews a critical discussion on the non-neutral impact of technology and the non-zero-sum ethical effects it triggers (Hagendorff, 2020). This demands more than ever an approach to the teaching of data ethics that is non-essentialist and non-axiomatic in its foundation as well as ecosystemic in its outlook.

Teaching data ethics in/for the urban environment

In 2022, the two authors of this article designed a 5ECTS elective course titled 'Ethics for the data-data driven city' within the Faculty of Architecture and the Built Environment at TU Delft. Although open to master students from all the other university's faculties, the course was embedded in the master program in Geomatics, within the department of Urbanism. To guide the conception of the course was the necessity to rework the mainstream essentialist and axiomatic framing of data ethics, foregrounding, by contrast, an ecosystemic understanding of data, ethics, and the city. Below, we discuss the conceptual foundations of the course.

Data and technology: a sociotechnical approach

Following a normative understanding, data are considered as the unrefined version of information. This is a realist stance, that is, one that regards data as true and factual by default that leads to consider information as the meaningful organization of data, assuming that the two are inherently isomorphic and the passage from one to the other can occur smoothly. Such a stance defines, de facto, how information and data are understood within information science and applied fields, with the data-information-knowledge-wisdom (DIKW) pyramid (Figure 1) as a trademark (Frické, 2009).
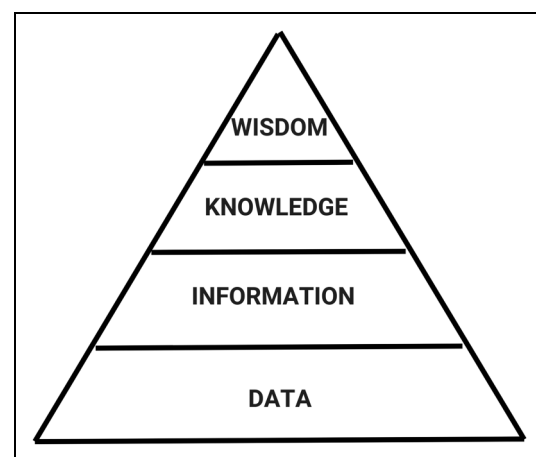

Figure 1. The data, information, knowledge, wisdom pyramid.

However, not only is it possible to subvert such relations – with information preceding data – but argue in favour of the non-isomorphism between the two. This allows to deconstruct the 'naturalization' of data and to foreground, by contrast, their sociotechnical fabric. Landauer (1996: 188) nails this down affirming that 'information is not a disembodied abstract entity; it is always tied to a physical representation'. Data, in other words, do not pre-exist in nature; instead, they are an enframing of information that comes into being under certain sociotechnical conditions. On the one hand, this implies a paradigmatic shift in the way we think about data, passing from data as a 'thing' to data as a sociotechnical process: data are always created through certain means, used for certain purposes, in certain contexts, by certain people, and with certain results. On the other hand, data inevitably foreground certain aspects and hide others (e.g. 'dark data', cf. Hand 2020): this means that a data-based model can never be the reality itself, but is just a representation of reality, no matter how much data one collects. Hence it becomes crucial, as a preliminary step, to explore the sociotechnical conditions – actors, factors, and discourses – through which data are being created.

By extension, technology shall not be simply reduced to a set of tools, but regarded more broadly as an apparatus, embracing 'a thoroughly heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral and philanthropic propositions' (Foucault, 1980: 194). This might sound as an all-encompassing definition of technology, but the point to make is that technology is always already imbricated with/in the socio-cultural fabric from which it is informed and which, in turn, it contributes to inform. Notably, as Hess and Ostrom (2007: 10) explain, technology has a pivotal role in 'captur[ing] the previously uncapturable, with the resource being converted from a nonrivalrous, nonexclusive public good into a resource that needs to be managed, monitored, and protected, to ensure sustainability and preservation'. This means that technology is an apparatus responsible for reifying certain aspects of reality and turning them into something that can be seized, distributed, and managed. This is why, as Kranzberg (1986) correctly affirms, in his first law: 'technology is neither good nor bad; *nor is it neutral* [italics added]'. Subverting the refrain according to which the supposed goodness or badness of technology would simply depend on the use one makes of it, Kranzberg's first law implies any technology is always good *and* bad at once, insofar as its implementation and use, far from being a neutral affair, produce, at all times, value-laden entanglements that demand to be assessed from different perspectives simultaneously. The recognition of technology as an apparatus moves away from an essentialist stance towards an ecosystemic one able to account for the ethical unintended and intended effects that data-driven technologies produce, leading also to a critique of risk-based approaches as insufficient to tackle public ethical concerns (cf. Wynne, 2001).

Beyond good versus bad: A non-axiomatic approach to ethics

In the same fashion as a normative and essentialist conception of data and technology risks oversimplifying their sociotechnical coming into being, an axiomatic understanding of ethics risks polarizing the discussion over their impact in society. Wilk (2019: 10) acknowledges that 'ethics does not always provide a right answer to moral problems. For many ethical issues there is not a "right" answer'. More radically, it is not only the discernment of good and bad to get contested, but the very possibility of finding (ethically robust) answers and solutions. This means to endorse a non-axiomatic understanding of ethics which regards ethics not as a toolbox, but a diffused dimension requiring constant contextualization and ongoing negotiation. In other words, what is to be challenged is the very possibility of defining once and for all (in absolute and abstract terms) good and bad technologies (including their use). More specifically, we define ethics as a *systematic reflection* on what a community *collectively* justifies as good or bad, how, and why. This definition, which bridges relational ethics (Zigon, 2019) to utilitarian positions, has two advantages: (1) it regards ethics as a practice that cannot be framed once and for all, nor be abstracted from the context; (2) it regards ethics has bearing a collective connotation by default, meaning that it is not possible to either reduce ethics to an individual affair only (cf. 'virtue ethics'), nor to instrumentalist positions. The concept of community is not unproblematic, especially concerning how to delimit it (cf. Tannier and Thomas, 2013). In this regard, we advocate the importance of applying both a spatial and temporal dimensions to the community, thus considering it as a dimension where, on the one hand, the whole is something more and different than the sum of its part and, on the other, whose thriving is enabled by securing balance over time. This, in turn, entails also the need to explore the power asymmetries among the actors forming the community and to design check-and-balance mechanisms that enable these asymmetries to be redressed in a rolling basis. Aligning to the complexity to which behavioural ethics hints, ethics is a fundamentally uncertain practice, insofar as it provides a temporary synthesis, among different stances, of what a given collective considers as good. This resonates with idea of 'care', whereby the value of ethics resides in the 'relational, contextualised, embodied and realised through practices rather than residing in stand-alone principles' (Atenas et al., 2023: 19). From here, when coupled with data-driven technologies,

an ethics-as-practice that maintains a collective dimension by default leads to explore value-laden entanglements[5] in the development and implementation of data-driven technologies, as well as unintended consequences (both positive and negative) of their use in context.[6]

The city as a complex system

Cities are socio-techno-physical agglomerations that cannot be approached as a machine (Mattern, 2021), that is, as something to be mapped, broken into smaller parts, and then processed and recombined. Rather, they are 'hybrid complex systems' (Portugali, 2011) composed of biotic and artificial elements, whose codependency characterizes the coming into being of an urban ecosystem in which the whole is something more and different from the sum of its parts. As such, cities cannot be studied nor developed by isolating either elements or their interactions but require to be studied in their entirety, insofar as they manifest emergent behaviours that are very difficult, if not impossible, to predict (Grieves and Vickers, 2017).

On this point, Bettencourt (2015: 220) contends that 'the challenge for a modern science of cities is to define urban issues in their own right and to seek integrated solutions that play to the natural dynamics of cites'. The idea of resorting to integrated solutions characterizes de facto an ecosystemic approach whereby urban interventions – such as the implementation and use of data-driven technologies – are explored based on a comprehensive view of the city. Further on, Bettencourt (2015: 231) specifies that self-organizing practices – that is, practices able to give agency to the needs of all actors involved – represent the best response to tackle city's complexity, insofar as self-organization 'places emphasis on creativity and on effective social organizations, capable of coordinating their knowledge and action'. It is social self-organization that is key to enable an integrated approach to the city which applies, at once, to its complex 'messiness' (Jacobs, 1969), to the socio-technical design, implementation, and use of data-driven technologies, as well as to a non-axiomatic unpacking of the value-laden entanglements arising at the intersection of the physical and the social. In this respect the imaginary city of the future 'Urbidata' (Ploeger and van Loenen, 2018) represents an ideal ethical space of enquiry at the convergence of material infrastructures, informational fluxes, institutional actors, and grassroots collective initiatives.

An ecosystemic vision: data ethics for the city

An ecosystem is characterized by homeostasis, that is, the balanced interaction among its constitutive elements, as well as between them and the environment. While largely related to the natural world, the notion of ecosystem has also been applied to other settings, such as the urban environment (Jarke et al., 2019) and the data landscape

(van Loenen et al., 2021). It is in this spirit that, by intersecting data-driven technologies and the city, Sennett (2018) speaks of the need to design and enact an open-ended urban dimension which is able to accommodate conflict and dissonance from within and synthesize a plurality of stances and values. Easterling (2021: 62) summarizes this well arguing that 'it is the architecture of interplay and entanglement that is the real innovation (…) Value begins with physical arrangement, location, community, diversity'. This, in turn, requires not only to investigate this or that location and value-laden arrangement, but to guide the development, implementation, and use of data-driven technologies by striking a systemic balance among people, values, and apparatuses (Figure 2). From an ethical point of view, the socio-techno-physical urban ecosystem is in equilibrium as long as it behaves fairly, understanding fairness as a meta-principle that can be disentangled on a rolling and contextual basis as a balancing exercise involving both individual and collective rights, interests, and values (cf. Rochel, 2021).

Designing the course "ethics for the data-driven city"

Transdisciplinary method: From problem-solving to problem-opening

Our goal was to turn this sociotechnical, non-axiological and ecosystemic understanding of the data/technology, ethics, and the city into a course on 'data ethics for the city'. Notably, the idea was to foster a transdisciplinary approach embracing our standpoints and which could enable students not only to grasp the complexity of the
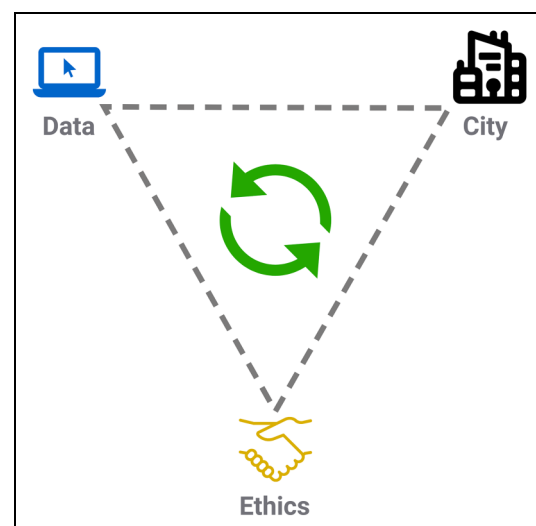
Figure 2. The ecosystem connecting data as sociotechnical processes, the city as a complex system, and ethics as a collective practice.

ethical dilemmas that data-driven technologies in/for the city bring with themselves, but also to critically operationalize such understanding through the realization of their final assignments. If we speak of transdisciplinary approach, it is because, differently from multidisciplinarity and interdisciplinarity, transdisciplinarity posits by default the unitary nature of knowledge: as Nicolescu (2005) writes, transdisciplinarity 'concerns the dynamics engendered by the action of several levels of reality at once'. This is key to enact a sociotechnical, non-axiomatic and ecosystemic understanding of data ethics for the city. Indeed, in a data environment where all answers are accessible and assembled on demand, students shall be especially encouraged to cultivate doubt, intended as an adaptive stance stemming from the awareness of the intrinsic uncertainty and value-laden entanglements of our own being and acting in the world. From here, we envisioned the exploration of the double-sidedness of tech-related concepts and principles, as well as of the unavoidable emergence of unintended consequences connected with the implementation and use of data-driven technologies in the urban context.

This led us to develop a pedagogical method to data ethics that is not problem-solving, but *problem-opening*, that is, a method that recognizes and constantly problematizes the ethical multifacetedness and inherent open-endedness of all ethical stances and tech 'solutions'. Just to give some examples, we compelled students to critically engage with values such as 'transparency', 'openness', 'inclusivity', 'trust', or 'privacy', which are repeatedly attached to the functioning of today's data-driven technologies and services.[7] The critical point is that any value cannot be taken as one-dimensional or in isolation: any one of them always presupposes its own opposite. Thus, there cannot be, for instance, openness (e.g. of data) and transparency, without defining, acknowledging, and accounting for closure and opacity. Or also, a data-driven service designed to promote inclusiveness might achieve that for certain people and not for others; or it might be inclusive for certain people under certain conditions, but then result exclusive for these same people under other conditions. Similarly, Duenas-Cid and Calzati (2023) show that trust shall be best explored as an entangled concept – 'dis/trust' – which accounts for the duplicitous co-presence of its own two opposites. The same goes for personal data: Purtova (2017: 192) rightly claims that 'just as light sometimes acts as a particle and sometimes as a wave, data sometimes act as personal data and at other times as non-personal data'. At stake is the fundamental awareness that there is no clear-cut way to discern once and for all whether a certain set of data contains personal data or not; these are two complementary features. Last, speaking of Open Government Data, Bates (2014) notes that 'the ends to which openness is being driven by different social actors have become more complex and contested. For some advocates this emerging complexity has been framed in terms of the "unintended consequences" of OGD'. From a pedagogical point of view, it is precisely these 'unintended consequences' (both negative as positive) that need to become the focus of attention: they are not happening 'by chance', they are systemic and require transdisciplinary tackling.

Unpacking the course: research-and-design workshop

The elective course 'Ethics for the data-driven city' was delivered in the presence in the fourth quarter (April–June) of the academic years 2021–2022, 2022–2023, and 2023–2024. Although iteration plays a role in the (ongoing) refinement of the course, for clarity, we focus on the year 2022–2023. The course spanned over eight weeks, comprising of 13 classes in total, 2 h each, held twice a week. Eight students from different departments and faculties, as well as with different socio-cultural backgrounds (i.e. Europe and Asia), enrolled in the course. Master students at TU Delft were a particularly fitting target for our course because they mostly have a technical background; hence, these students are well-equipped to develop tomorrow's data-driven technologies, but also usually exposed, due to how their own curricula are designed, to a deficiency in ethical reasoning.

The first three classes were introductory ones, with each of them touching upon one of the three pillars of the course: data, ethics, and the city. As a compass to guide the classes, compulsory readings were provided (cf. syllabus). This allowed us to set the scene for the course helping students familiarize with our sociotechnical, non-axiological, and ecosystemic understanding of data ethics for the city.[8] Moreover, these three classes were enriched with real-life hands-on experiences, including a visit to the TU Delft living lab The Green Village,[9] a guest lecture from digital researcher and artist Winnie Soon,[10] and a visit to an exhibition by Dutch digital artist Richard Vijgen at Het Nieuwe Instituut,[11] a museum for architecture, design, and digital culture in Rotterdam.

From class 4 onwards the course was conceived as a research-and-design workshop, aligning to the technical, practice-based orientation of TU Delft (on teaching ethics through design cf. Metcalf et al., 2015). Due to the nature of the course, attendance was compulsory, and a key pedagogical cornerstone was in-class discussion not only to acknowledge the presence of multiple ethical standpoints (often culture-dependent), but also to recognize the insolvability of many value-laden scenarios affected by the use of data-driven technologies. To concretize this, we followed suggestions in existing literature according to which courses in data ethics are most effective when striking a balance between theory and practice (Haws, 2001) facilitating the ethical framing of technical issues through the presentation of real-life case studies (Burton et al., 2017). To this, we added references to non-technical and non-

academic works, such as literary texts (e.g. Isaac Asimov and Luis Borges), arts (e.g. Maurits Escher and René Magritte), and TV shows (e.g. 'Black Mirror').

Concerning the workshop-style of the course, each class revolved around a given topic, and before each class students had to read two compulsory articles, spanning across critical data studies, philosophy of technology, digital cultures, technology governance, and political ecology. In the first part of each class, a student presented and discussed one of the two articles, making sure that the presentation was linked to a current real-life example and ended with a few open questions for the whole class. Then, through moderated debate, the second article was also brought to the fore, creating a dialogue between the readings and the students' own backgrounds, and leaving to us the role of class coordinators, rather than lecturers (this was easier especially in the second half of the course when students became familiar with each other and the class setup).

The second part of each class was dedicated to discussing students' advancement towards the realization of their final assignment via either individual or collective presentations. The final (individual-based) assignment relates more closely to the research-and-design part of the course, and was conceived as an ongoing mentored process, rather than an output to be submitted at the end of the course (facilitated by the small cohort of students). This is how we proceeded. Since the first class, students were asked to identify a possible case study linking data processing – i.e., the collection, storage, distribution, (re)use of either analogue or digital data to inform a service, such as an app; a process, such as enrollments or subscriptions; an initiative or a technology testing, such as smart bike lanes – to a given urban setting, taking TU Delft campus as our 'city'. This narrowing down of the breadth of the spatial dimension under exam (which builds upon the experience from the course in its first year, 2021/2022) had the goal to facilitate the choice of the students (due to the limited duration of the course) and to anchor such choice to their own experience, since they were already familiar with the campus.

Once the case study was identified, the research part comprised of three steps: (1) *framing* (e.g. how the case study is presented; how it works; who leads it/who is involved; its function, targets, and goals); (2) *data-value entanglements* (e.g. based on the framing, what are the values mobilized and foregrounded, as well as their own opposites; which data are collected and which ones are overlooked); (3) *open-ended scenarios* (e.g. mapping expected and/or obtained results; possible unintended – both positive and negative – consequences).

Concerning the design part, this consisted of two outputs: (a) project journal; (b) (final) artefact. Since we were interested in the learning process as much as in the final result, we asked students to keep track from the end of class 1, by means of a project journal, of their own reflections about the research part of their work, jotting down ideas and advancements towards the identification of their case study, as well as the realization of their final artefacts. The journal had to be in writing, but we let students free to decide if it was handwritten or typed, as well as to add other media formats (drawings, photos, images, videos, animations, vocal messages, etc.). The journal then worked as a sort of self-reflective support revealing the ups and downs of the learning process.

By artefact we meant the realization of a (simple) physical or digital object – e.g. a model, a video, an interactive chart or platform, an app mockup, etc. (with the exclusion of posters, based on the first year experience, as these tend to represent a last-minute choice) – that exposes and/ or redresses the ethical tensions – in the form of data-value entanglements and open-ended scenarios – embedded in the chosen case study. Complementary to it, we asked students to write a short 'author's statement' (500 words max) describing the artefact and addressing the following four points: (1) what they focused on; (2) why; (3) which ethical tensions emerged; (4) how these tensions were explored through the artefact (i.e. design choices).

Last, the students had the chance to discuss their own work (journal and artefact) during a 15-min oral exam. In fact, the exam was not aimed to test their knowledge, but their ability to motivate, further elaborate, and defend their own work.

Results and discussion

In this section, we present the final assignments of two students to exemplify the results coming out of the course. Furthermore, we introduce the evaluating feedback for the whole course received from the entire class in order to further reflect on its strengths and weaknesses.

Assignment 1: the outdoor mobility digital twin

One student focused his final assignment on the Outdoor Mobility Digital Twin (OMDt) of TU Delft campus, part of the Mobility Innovation Centre Delft (MICD), which was launched officially in July 2020. According to the initiative's website,[12] the OMDt project aims to:

> monitor and visualize all traffic on our campus, including pedestrians and cyclists. (…) Sensors collect visual data and instantly convert these data into privacy-proof traffic information, showing pedestrians and cyclists as moving dots. This information is combined with publicly available data such as the opening times of bridges and real-time locations of trains, buses and trams. (…) The system is ultimately intended to make *predictions* about movements and crowding on campus, both short-term (15 minutes ahead) and long-term (days, weeks and months).

The student, who was also able to get in contact with the director of MICD, wrote in his author's statement that 'the case study caught my attention because I was unaware that we even had a sensor network on campus. The fact that many other students also did not know this intrigued me'. From here, the students identified two relevant ethical tensions related to the case study, especially concerning its deployment and design.

To begin with, an issue of transparency is at stake as far as the informed collection of data is concerned. The student put this down as follows: 'Currently, most observed people do not know they are observed, nor what happens to these observations. More transparency and communication could be considered regarding the data collection, processing, and use of the model'. As a matter of fact, it is the politics surrounding the whole data life cycle to be under scrutiny, beyond the declared privacy-proof collection and handling of data. This is even more relevant considering that while TU Delft campus's roads are a public space and open to the general public, the development and implementation of the OMDt sees the involvement of private stakeholders that are not bound to public scrutiny even when their tech solutions integrate and build upon open government data.

Second, the predictive mobility model gives rise to potential forms of exclusion in that the road users observed and tracked do not exhaust the diversity of people moving around the campus. Notably, the targeted road users are pedestrians, cyclists, cars, and public transports (i.e. buses). From here, the aggregated data are classified based on parameters such as velocity, acceleration, and location. This – in the students' own words – may lead to ethical tensions, for instance 'when considering people accompanied by animals or people with disabilities'. The model indeed remains blind over these other road users, with the consequence that adjustments and predictions based on it may not only be inaccurate, but more problematically 'overlook the specific requirements of individuals with unique mobility needs'.

Based on this analysis, the student created a physical artefact in the form of a black box showcasing in its interior a traffic scene that could be looked at from different perspectives, each producing a different effect on the unfolding of the perceived scene. Figure 3 shows the final artefact and its coming into being – from conceptualization to design – through the student's journal.

In the author's statement, the student described the rationale of the artefact, titled 'Pathways of Inclusion', as follows:

> I have decided to opt for a physical artefact because I believe that in the current digital and data driven world the physical is too often forgotten or ignored. (…) Stylewise, I have chosen to make the outside of the artefact black, signaling and symbolizing the "black box" that this
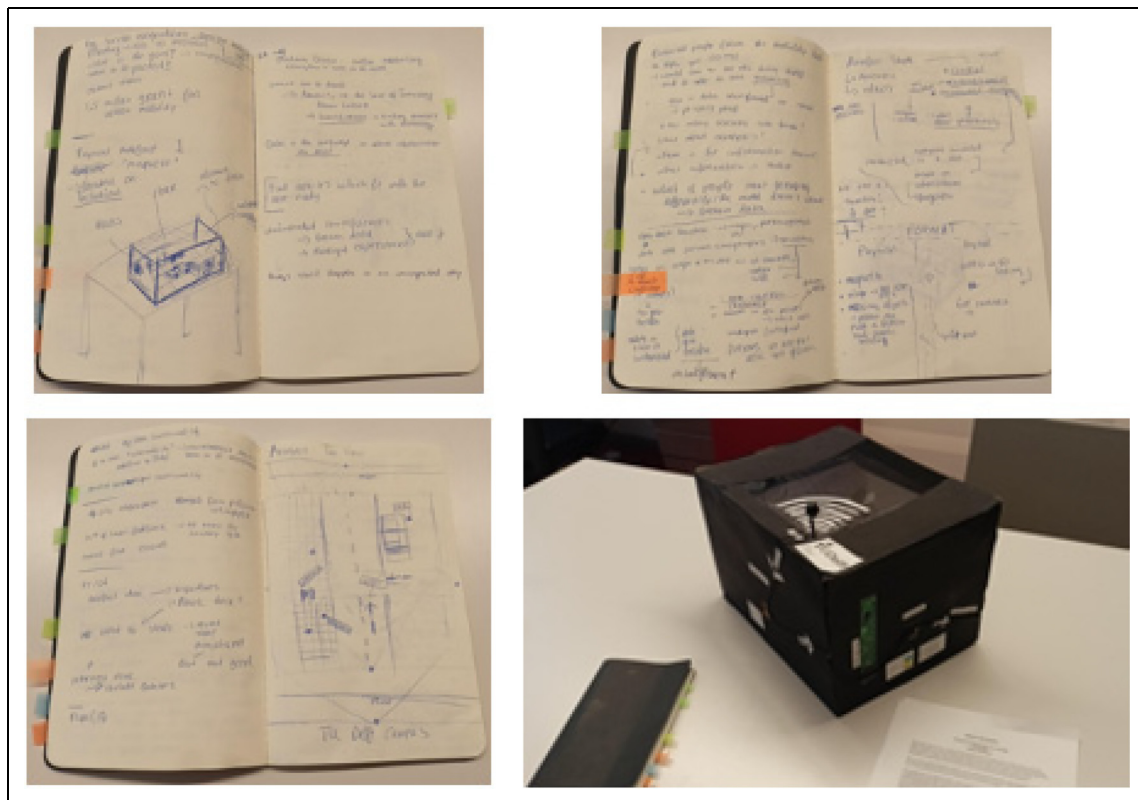

Figure 3. Some pages from the student's journal and the final artefact.

model and technology is for an outsider. The main function of the artefact is enabling the observer to see the same urban mobility scene through different lenses, both showcasing the data acquisition process, and the ethical perspectives. The fact that one has to hold the artefact close to the eyes to be able to see through the lenses also illustrates the problem of transparency; one has to put effort to actually see and understand what happens in the black box.

Overall, the artefact does not only expose the ethical tensions identified by the student, but it does so by creating an experience for the viewer/user that forces a critical reflection through embodied affection. The peeping into the black box from different angles and through different filters produces forms of opacity and exclusions in the same way as the OMDt might do. The artefact, indeed, puts the viewer in the role of data processor subjected to inevitable blind spots as the model is. We found this a very good concretization of the non-axiomatic framing of ethics in connection with the exploration of the inevitable tensions that the implementation and use of data-driven technologies in an urban setting bring to the fore.

Assignment 2: privacy use ICT facilities

The second example comes from a student who decided to focus her assignment on the email service of the TU Delft and, more specifically, on the ethical tensions arising whenever access to a student or employee's email box is granted to third parties, based on the identification of 'weighty situations'. This is how the policy is described on the University's website:

> The basic principle is that the contents of your e-mail box or personal data folder may only be viewed with your permission. It is, however, conceivable that an exception to this principle must be made on the basis of weighty reasons. For example, conflict situations. In such a case, your immediate manager can submit a request to gain access to your e-mail box or personal data folder, without your permission. Here, your immediate manager must have obtained permission from the Head of Department and the Head of HR.

In terms of institutional communication this is all the student was able to obtain, as no further specifications were provided about what these 'weighty' and 'conflict' situations consist of and/or how they are assessed or by whom. In her author's statement, the student explained that she arrived at this topic after she realized that the University had discontinued the automatic forward service allowing to have one's own university's emails automatically forwarded to a preferred email service provider. Looking more into the issue, she found that the institutional motivation to discontinue this service was the tightening of the security measures concerning the sharing of personal data and private information to third parties, insofar as it was not possible for the University to determine whether external email providers take sufficient security measures to protect students and employees' data.

Based on this consideration, the student found puzzling the lack of clear specifications concerning the criteria, rules, and modalities for allowing third parties to access one's own university email box in some circumstances: 'it was intriguing to critically examine the ethical dimensions of these weighty situations, recognizing the need to address the potential unintended consequences arising from such access', she noted in her journal.

The exploration of the topic led the student to seek clarification from various institutional actors at different levels within the University (e.g. IT, human resources, data stewards), without however getting any specific and accurate response. This unexpected result forced the student to shift the focus of her ethical enquiry away from an evaluation of the fairness of the decisional process surrounding 'weighty situations' towards the exposure of the lack of transparency on the responsibilities and modalities of such process. In other words, the infrastructural-commercial aspect of the email service was sidelined in favour of the procedural-policy aspect, foregrounding the sociotechnical nature of the service under exam. This what the student wrote in her (typed) journal:

> As I tried getting answers from the relative parties of TU Delft on this matter, I received unresponsiveness mostly. Their lack of answers and engagement intrigued me, leading me to symbolize this in my artifact as the "black, mysterious box." This box represents the unknown aspects related to granting access or storing emails, particularly in weighty situations.

In fact, the final artefact of the student is composed of two complementary parts, a physical one (Figure 4) and a digital one in the form of a short, animated clip that is meant to supply the narrative for the physical artefact, 'represent[ing] the process of an email being sent to a TU Delft webmail account'. Concerning the physical artefact, the student explained in the author's statement that,

> the black mystery box signifies the unknown aspects related to granting access or storing emails, particularly in weighty situations. To further illustrate the intricacies of the topic, I have combined printed email conversations with relevant [University] parties [anonymized]. These conversations revolve around inquiries regarding the procedure for granting access, conflict situation definitions, documentation requirements, relevant regulations, involvement with the service provider, and time limitations on access in weighty situations. The main idea behind was to highlight the lack of responses and answers gained in return from the relevant TU Delft parties.

Figure 4. The physical artefact showcased by the student during the oral exam.

Overall, based on the (lack of) evidence she was confronted with, the student was able to twist her ethical enquiry, no longer questioning the possible value-laden entanglements of the decisional process and actors surrounding the assessment of weighty situations, but targeting the unavoidable ethical implications connected with the lack of clear specifications concerning these same situations. This is especially relevant in light of the University's attentiveness towards the securization of students and employees' personal data and information. It is this very discrepancy that led the student to physically represent the sociotechnical process of emails' handling as a black box, with the aim 'to foster increased awareness surrounding the ethical challenges inherent in the access of emails by third parties'. We found this assignment particularly relevant because it materialized the absence of (requested) information while building a proper narrative contextualization to the artefact through the shot animated video.

Feedback

Upon the completion of the course, a brief anonymous survey was submitted to the students asking them to evaluate the whole learning experience as well as to provide suggestions for possible improvements. Concerning the course, students were asked to grade on a scale from 1 to 5 (with 1 as 'very bad' and 5 as 'very good') the following questions: 'How did you like the course?'; 'How did you find the structure of the course?'; 'How did you find the contents of the classes?'; 'How did you find the communication form the teachers?'; 'How did you find the guidelines for the assignments?' Overall, in 2022/2023 the course received an average score of 4.52, with the highest score (5) for the 'communication from the teachers' and the lowest (3.7) for the 'guidelines for the assignments'. Concerning the topics treated, the most popular were the 'visit to The Green Village' (4.8), 'smart citizenship and data governance' (4.7), 'data, city & inequalities' (4.7), and 'data & environmental sustainability' (4.7); while 'health data' (3.7) and 'platformization' (3.8) scored the least. Beyond that, students particularly enjoyed visits and references to non-technical, non-academic works that easily exemplify ethical dilemmas and theoretical cornerstones concerning the value-laden entanglement and unintended consequences of the use of data-driven technologies in today's society.

Among the suggestions for possible improvements were (a) the definition of a proper rubric for the assessment of the final assignments (we did provide five criteria, without however a defined grading scale); and (b) identification of case studies related to the students' study curricula (which might imply to let students identify and discuss relevant real-life examples based on the readings and their own background).

Conclusion

Limitations and further developments

Despite the positive feedback, we are aware that the course faced some limitations. First, being an elective course, the number of students who enrolled was small. This reduced, at least in part, the richness and diversity of the

in-class discussion. Moreover, the small cohort of students compelled us to design the final assignment as an individual task, preventing students from working in pairs or in small groups, an option that could have led to more fruitful discussions and multi-perspectival critical insights towards the realization of their artefacts. Second, the research-and-design we created would benefit from a longer timeframe for its realization, ideally spanning the whole semester instead of only a quarter. In this respect, the fact that two Geomatics' students asked us to supervise their master theses along the lines of the course signals their keenness to further explore ethics in connection within the development and use of data-driven technologies.

As a proof of concept, we believe the course highlighted the extent to which the imbrications between the 'technical' and the 'social' – data and people – is de facto a codependency which requires exploring new ways to conceptualize and teach what it means to do good with/through data-driven technologies in complex real-life scenarios. We also believe that the positive feedback we got stands as evidence of a demand for such courses and approaches. Hopefully this teaching experience can contribute to ongoing pedagogical debates in connection with critical data ethics, forcing an ecosystemic transdisciplinary reframing of the field.

Final remarks

By linking the teaching experience to the theoretical part of this article, we can outline some final remarks. First of all, the course actualized the extent to which 'good' and 'bad' (uses of) technologies are a matter of framing, not an essentialist one. This means that these concepts depend on the context and the given timeframe, as well as on the people involved and their values over time, rather than being ethical labels to be attached to a specific technology (or, even worse, person). This entails, in turn, to move beyond an instrumentalist ethical stance and to look, instead, at the inevitable tech-value entanglements and open-ended scenarios that any (ever-neutral) technology brings with itself. In this respect, the course compelled students to cultivate a different way of thinking critically about (the definition of) problems, by shifting the ground on which their approach (often in the form of engineering problem-solving) rests. Most importantly, we tried to achieve this not as an end in itself but as an ongoing process that could find a materialization in the students' final work. This is because we believe and advocate that *to think is to act*: the former is not something abstract and detached from 'the real world'; it is always already a way of configuring reality.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Stefano Calzati ![ORCID] https://orcid.org/0000-0002-4590-6709
Hendrik Ploeger ![ORCID] https://orcid.org/0000-0002-7740-9184

Supplemental material

Supplemental material for this article is available online.

Notes

1. In this context the term 'data-driven technology' encompasses a broad spectrum of digital technologies and services that make use of big data to function. In this respect, the authors would like to stress that the use of the term does not endorse a data- or tech-centric standpoint; it simply aims to identify a diverse set of algorithmic technologies.
2. For the present discussion, the term 'data' in 'data ethics' (Floridi and Taddeo, 2016) encompasses a broad and varied set of technological innovations from data-driven technologies and AI to large-language models and robotics, starting from the consideration that data are at the core of all these innovations.
3. While the framing in terms of dilemmas might already constitute a form of ethical boxing, what matters here is the idea of ethical dilemmas as unsolvable win-win situations, that is, of scenarios inevitably open to tradeoffs among values, with both positive and negative un/intended consequences.
4. Here we understand 'value' following Wilk's (2019) definition: 'Values are abstract concepts of what a society regards as most worthwhile, for example, the value of equality between human beings'.
5. One example discussed in our course involves the failed use of a popular navigating app in an emergency situation: https://eu.usatoday.com/story/tech/news/2017/12/07/california-fires-navigation-apps-like-waze-sent-commuters-into-flames-drivers/930904001/
6. One example discussed in our course is the notorious case of a Chicago citizen targeted by a policing algorithm, which triggered a chain of dramatic self-fulfilling events: https://www.theverge.com/c/22444020/chicago-pd-predictive-policing-heat-list
7. As describe further below, we did not provide any a priori list of values to students as they had to explore the chosen case study to understand how it was value framed.
8. We should specify that while issues of privacy and data protection returned from time to time in the course, we expressly stated at the beginning that these issues constituted our baseline, rather than the horizon of the whole course.
9. https://www.thegreenvillage.org/en/: 'At The Green Village, knowledge- and educational institutions, entrepreneurs, government bodies and civilians can research, experiment, validate and demonstrate their sustainable innovations'.
10. https://researchers.arts.ac.uk/2466-winnie-soon: 'Winnie Soon's research and practice intersects with art and technology in the areas of Software Studies and Computational

Cultures (…) to understand the potential and implications of technology in a wider cultural and societal context'.

11. https://nieuweinstituut.nl/en.
12. https://micd.tudelftcampus.nl/projects/outdoor-mobility-digital-twin/

References

Atenas J, Havemann L and Timmermann C (2023) Reframing data ethics in research methods education: A pathway to critical data literacy. *International Journal of Educational Technology in Higher Education* 20(1): 11. Available at: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2731 (accessed 30 October 2023).

Bates J (2014) Open government data and the neoliberal state. Available at: https://blogs.lse.ac.uk/impactofsocialsciences/2014/10/02/open-governmentdata-and-the-neoliberal-state/ (accessed 30 October 2023).

Bettencourt LM (2015) Cities as complex systems. In: Furtado BA, Sakowski PAME and Tóvolli MHE (eds) *Modelling Complex Systems for Public Policies*. Brasilia: Ipea, 217–236.

Bigman YE and Gray K (2018) People are averse to machines making moral decisions. *Cognition* 181: 21–34.

Borenstein J and Howard A (2021) Emerging challenges in AI and the need for AI ethics education. *AI and Ethics* 1: 61–65.

Brey P (2000) Disclosive computer ethics. *ACM SIGCAS Computers and Society* 30(4): 10–16.

Burton E, Goldsmith J, Koenig S, et al. (2017) Ethical considerations in artificial intelligence courses.

Calzati S and van Loenen B (2023) A fourth way to the digital transformation: The data republic as a fair data ecosystem. *Data & Policy* 5(21). https://doi.org/10.1017/dap.2023.18.

Calzati S and van Loenen B (2023b) Towards a citizen- and citizenry-centric digitalization of the urban environment: urban digital twinning as commoning. *Digital Society* 2(3).

Casiraghi S (2023) Anything new under the sun? Insights from a history of institutionalized AI ethics. *Ethics and Information Technology* 25: 28. https://doi.org/10.1007/s44206-023-00064-0.

Coeckelbergh M (2020) *AI Ethics*. Cambridge: MIT Press.

d'Aquin M, Troullinou P, O'Connor NE, et al. (2018) Towards an "ethics by design" methodology for AI research projects. *In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*: 54–59. https://doi.org/ff10.1145/3278721.3278765.

Duenas-Cid D and Calzati S (2023) Dis/trust and data driven technologies. *Internet Policy Review* 12. https://doi.org/10.14763/2023.4.1727.

Easterling K (2021) *Medium Design. Knowing How to Work on the World*. London: Verso Books.

Ess C (2006) Ethical pluralism and global information ethics. *Ethics and Information Technology* 8(4): 215–226.

European Commission (2019) Ethics guidelines for trustworthy AI. Available at https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 30 October 2023).

Fiesler C, Garrett N and Beard N (2020) What do we teach when we teach tech ethics? A syllabi analysis. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*: 289–295. https://doi.org/10.1145/3328778.3366825.

Floridi L (2018) Soft ethics and the governance of the digital. *Philosophy & Technology* 31: 1–8.

Floridi L (2019) Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology* 32: 185–193.

Floridi L, Cowls J, Beltrametti M, et al. (2021) An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. In: Floridi L (ed) *Ethics, Governance, and Policies in Artificial Intelligence*. Cham: Springer, 19–39.

Floridi L and Mariarosaria T (2016) What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2083): 20160360.

Fotopoulou A (2021) Conceptualising critical data literacies for civil society organisations: Agency, care, and social responsibility. *Information, Communication & Society* 24(11): 1640–1657.

Foucault M (1980) *Power/knowledge: Selected interviews and other writings, 1972–1977*. New York: Pantheon Books.

Frické M (2009) The knowledge pyramid: A critique of the DIKW hierarchy. *Journal of Information Science* 35(2): 131–142.

Furey H and Martin F (2019) AI Education matters: A modular approach to AI ethics education. *AI Matters* 4(4): 13–15.

Gillespie T, Aufderheide P, Carmi E, et al. (2023) Expanding the debate about content moderation: scholarly research agendas for the coming policy debates. *Internet Policy Review* 9(4). https://doi.org/10.14763/2020.4.1512.

Grieves M and Vickers J (2017) Digital twin: Mitigating unpredictable, undesirable emergent behaviour in complex systems. In: Kahlen FJ, Flumerfet S and Alves A (eds) *Transdisciplinary Perspectives on Complex Systems*. Cham: Springer, 85–113.

Hagendorff T (2020) The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines* 30: 99–120.

Hand DJ (2020) *Dark Data: Why What You Don't Know Matters*. Princeton: Princeton University Press.

Hasselbalch G (2019) Making sense of data ethics. The powers behind the data ethics debate in European policymaking. *Internet Policy Review* 8(2): 1–19.

Haws DR (2001) Ethics instruction in engineering education: A (mini) meta-analysis. *Journal of Engineering Education* 90(2): 223–229.

Hess C and Ostrom E (2007) *A Framework for Analysing the Knowledge Commons*. Cambridge: MIT Press.

Himma KE and Tavani HT (2008) *The Handbook of Information and Computer Ethics*. Hoboken: Wiley.

Horton D, McIlraith SA, Wang N, et al. (2022) Embedding ethics in computer science courses: Does it work? *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*: 481–487. https://doi.org/10.1145/3478431.3499407.

IEEE (2017) Initiative on ethics of autonomous and intelligent systems. Available at: https://ethicsinaction.ieee.org (accessed 30 October 2023).

Jacobs J (1969) *The Economy of Cities*. New York: Vintage.

Jarke M, Otto B and Ram S (2019) Data sovereignty and data space ecosystems. *Business & Information Systems Engineering* 61(5): 549–550.

Jobin A, Ienca M and Vayena E (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9): 389–399.

Johnson D (1985) *Computer Ethics*. New Jersey: Prentice-Hall.

Johnson D (2017) Can engineering ethics be taught? The Bridge: *Journal of the National Academy of Engineering* 1(47): 59–64.

Kasapoglu T, Masso A and Calzati S (2021) Unpacking algorithms as technologies of power: Syrian refugees and data experts on algorithmic governance. *Digital Geography and Society* 2: 1–9.

Kilhoffer Z, Zhou Z, Wang F, et al. (2023) How technical do you get? I'm an English teacher": Teaching and learning cybersecurity and AI ethics in high school. *In 2023 IEEE Symposium on Security and Privacy*: 1–18. https://doi.org/10.1109/SP46215.2023.10179333.

Kleinberg J, Mullainathan S and Raghavan M (2016) Inherent trade-offs in the fair determination of risk scores. Available at: https://arxiv.org/pdf/1609.05807 (accessed 30 October 2023).

Kranzberg M (1986) Technology and history: Kranzberg's laws. *Technology and Culture* 27(3): 544–560.

Landauer R (1996) The physical nature of information. *Physics Letters A* 217(4-5): 188–193.

Lee MSA and Singh J (2021) The landscape and gaps in open source fairness toolkits. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*: 1–13. https://doi.org/10.1145/3411764.3445261.

Leonelli S (2019) Data governance is key to interpretation: Reconceptualizing data in data science. *Harvard Data Science Review* 1(1). https://doi.org/10.1162/99608f92.17405bb6.

Markham AN, Tiidenberg K and Herman A (2018) Ethics as methods: Doing ethics in the era of big data research – Introduction. *Social Media and Society* 4(3): 2–9.

Mattern S (2021) *A City is Not a Computer: Other Urban Intelligences*. Princeton: Princeton University Press.

Meng X-L (2021) What are the values of data, data science, or data scientists? *Harvard Data Science Review* 3(1). https://doi.org/10.1162/99608f92.ee717cf7.

Metcalf J, Crawford K and Keller E (2015) *Pedagogical Approaches to Data Ethics. Draft Report for the Council for Big Data, Ethics, and Society*. New York: Data & Society Research Institute.

Metcalf J, Moss E and Boyd D (2019) Owning ethics: Corporate logics, Silicon Valley, and the institutionalization of ethics. *Social Research: An International Quarterly* 86(2): 449–476.

Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1: 501–507.

Mittelstadt B, Allo P, Taddeo M, et al. (2016) The ethics of algorithms: Mapping the debate. *Big Data & Society* 3(2): 205395171667967.

Montreal Declaration for a Responsible Development of Artificial Intelligence (2017) Forum on the Socially Responsible Development of AI. Available at: https://www.montrealdeclaration-responsibleai.com/thedeclaration (accessed 30 October 2023).

Murillo LFR, Wylie C and Bourne P (2023) Critical data ethics pedagogies: Three (non-rival) approaches. *Big Data & Society* 10(2).

Nicolescu (2005) Towards transdisciplinary education. *The Journal for Transdisciplinary Research in Southern Africa* 1(1): 5–16.

Oliver JC and McNeil T (2021) Undergraduate data science degrees emphasize computer science and statistics but fall short in ethics training and domain-specific context. *PeerJ Computer Science* 7: e441.

Pink S, Korsmeyer H, Dahlgren K, et al. (2024) Automation in electric vehicle futures. *Mobilities*: 1–19. https://doi.org/10.1080/17450101.2024.2325385.

Ploeger H and van Loenen B (2018) 2050: The story of Urbidata. In: van Loenen B, Vancauwenberghe G and Crompvoets J (eds) *Open Data Exposed*. the Netherlands: T.M.C. Asser Press, The Hague, 269–288.

Portugali J (2011) *Complexity, Cognition and the City*. Berlin, Heidelberg: Springer Science & Business Media.

Prentice P (2014) Teaching behavioral ethics. *Journal of Legal Studies Education* 31(2): 325–365.

Purtova N (2017) Health data for common good: Defining the boundaries and social dilemmas of data commons. In: Adams S, Purtova N and Leenes R (eds) *Under Observation: The Interplay Between eHealth and Surveillance*. Berlin: Springer, 177–210.

Quinn M (2006) On teaching computer ethics within a computer science department. *Science and Engineering Ethics* 12: 335–343.

Rochel J (2021) Ethics in the GDPR: A blueprint for applied legal theory. *International Data Privacy Law* 11(2): 209–223.

Sennett R (2018) *Building and Dwelling: Ethics for the City*. New York: Farrar, Straus and Giroux.

Shilton K (2012) Values levers: Building ethics into design. *Science, Technology & Human Values* 38(3): 374–397.

Siau K and Wang W (2020) Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management (JDM)* 31(2): 74–87.

Stahl BC (2022) From computer ethics and the ethics of AI towards an ethics of digital ecosystems. *AI and Ethics* 2: 65–77.

Tallacchini M (2009) Governing by values. EU ethics: Soft tool, hard effects. *Minerva* 47: 281–306.

Tannier C and Thomas I (2013) Defining and characterizing urban boundaries: A fractal analysis of theoretical cities and Belgian cities. *Computers, Environment and Urban Systems* 41: 234–248.

Tavani HT (2016) *Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing*. London: Wiley.

Taylor L (2020) The price of certainty: How the politics of pandemic data demand an ethics of care. *Big Data & Society* 7(2): 205395172094253.

van Loenen B, Zuiderwijk A, Vancauwenberghe G, et al. (2021) Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. *JeDEM-eJournal of EDemocracy and Open Government* 13(2): 1–27.

van Maanen G (2022) AI ethics, ethics washing, and the need to politicize data ethics. *Digital Society* 1(2): 9.

Viljoen S (2021) A relational theory of data governance. *Yale LJ* 131: 573.

Wilk A (2019) Teaching AI, ethics, law and policy. Available at: https://arxiv.org/pdf/1904.12470.pdf (accessed 30October 2023).

Wynne B (2001) Creating public alienation: Expert cultures of risk and ethics on GMOs. *Science as Culture* 10(4): 445–481.

Zigon J (2019) Can machines be ethical? On the necessity of relational ethics and empathic attunement for data-centric technologies. *Social Research: An International Quarterly* 86(4): 1001–1022.