

COMMENTARY

# The virtuous smart city: Bridging the gap between ethical principles and practices of data-driven innovation

Viivi Lähteenoja<sup>1</sup>  and Kimmo Karhu<sup>2</sup>

<sup>1</sup>Faculty of Social Sciences, University of Helsinki, Helsinki, Finland

<sup>2</sup>Department of Industrial Engineering and Management, Aalto University, Espoo, Finland

**Corresponding author:** Viivi Lähteenoja; Email: [viivi.lahteenoja@helsinki.fi](mailto:viivi.lahteenoja@helsinki.fi)

**Received:** 11 February 2022; **Revised:** 03 March 2023; **Accepted:** 14 March 2023

**Keywords:** AI ethics; data-driven innovation; data policy; smart city; virtue ethics

## Abstract

For smart cities, data-driven innovation promises societal benefits and increased well-being for residents and visitors. At the same time, the deployment of data-driven innovation poses significant ethical challenges. Although cities and other public-sector actors have increasingly adopted ethical principles, employing them in practice remains challenging. In this commentary, we use a virtue-based approach that bridges the gap between abstract principles and the daily work of practitioners who engage in and with data-driven innovation processes. Inspired by Aristotle, we describe practices of data-driven innovation in a smart city applying the concepts of virtue and *phronêsis*, meaning good judgment of and sensitivity to ethical issues. We use a dialogic case-study approach to study two cases of data-driven innovation in the city of Helsinki. We then describe as an illustration of how our approach can help bridge the gap between concrete practices of data-driven innovation and high-level principles. Overall, we advance a theoretically grounded, virtue-based approach, which is practice oriented and linked to the daily work of data scientists and other practitioners of data-driven innovation. Further, this approach helps understand the need for and importance of individual application of *phronêsis*, which is particularly important in public-sector organizations that can experience gaps between principle and practice. This importance is further intensified in cases of data-driven innovation in which, by definition, novel and unknown contexts are explored.

## Policy Significance Statement

This work's policy implications are for decision-makers and practitioners of data-driven innovation in the public sector committed to ethical practice. We join those who argue for flexible and reasonable policymaking that considers the context. We suggest that a virtue-based approach can support their efforts to connect abstract and unchanging principles with contextual and shifting demands of day-to-day practice.

## 1. Introduction

Researchers have long recognized the potential and realized risks involved with data-driven innovation (e.g., Shattuck, 1984). To address these concerns, administrations and governments have been developing policies and tools to address the social effects of these innovative systems.<sup>1</sup> Based on a study by the Ada

<sup>1</sup> For a collection of relevant policy documents from 2006–2019, see <https://www.pt-ai.org/TG-ELS/policy/>

Lovelace Institute (2021), overwhelmingly the most common type of policy instrument chosen for this purpose is “Principles and guidelines.”

However, the definition of principles is not sufficient for artificial intelligence (AI), or data-driven innovation in general, to be ethically sustainable (for criticism of principlism in general, see, e.g., Ten Have, 2001; Page, 2012; Groves, 2015). Some reasons identified are the lack of tools and methods to translate principles into practice (Mittelstadt, 2019; Vakkuri et al., 2019), oversimplifying complex social constructs into idealized measures (Hagendorff, 2022), and ethical guidelines being prone to manipulation by industry actors (Rességuier and Rodrigues, 2020).

In this paper, we propose and illustrate a method for conducting this translation, which has been inspired by our reading of Aristotelian virtue ethics. We examine two instances of existing processes of data-driven innovation in the city of Helsinki and describe them through concepts of virtue and *phronêsis*, or practical reason. Using a dialogic case-study approach (Rule and John, 2015), we describe how data scientists, as practitioners of data-driven innovation, must and do engage their *phronêsis* in order to make ethically charged decisions. We give illustrative examples of four “bridging” data-driven innovation virtues that help connect the principles of the organization and the practical decisions made by innovation practitioners, which support this engagement.

We situate ourselves in an emerging field of practicing ethics in research and innovation. This body of literature was comprehensively reviewed recently by Reijers et al. (2018). As they write, “the complexity and ambivalence of ethical issues emerging from the design and outcomes of contemporary R&I [research and innovation] call for the development of comprehensive methods that can be used by ethicists, researchers, policymakers and various other stakeholders (technology users, companies, etc.)” (p. 1439).

Following the typology suggested by Reijers et al. (2018), our focus is primarily on contributing to the development of “intra” (as opposed to “ex ante” or “ex post”) methods of practicing ethics in innovation processes. That is, we focus on practices that come into play in the design and testing stages of the data-driven innovation process. Common challenges for these “intra” methods have been identified in the literature as, first, their inability to (help) integrate ethics into the work of the practitioners themselves (Borning and Muller, 2012; for critiques along these lines, see, e.g., Brey, 2000; Le Dantec et al., 2009). A second common challenge for “intra” methods is a lack of theoretical grounding in practices designed to embed values into the innovation process (for critiques along these lines, see, e.g., Albrechtslund, 2007; Manders-Huits, 2011; van de Poel, 2013).

With this commentary, we want to advance a theoretically grounded, virtue-based approach, which is practice-oriented and can be linked to the daily work of data scientists and other practitioners of data-driven innovation, especially in the context of the public sector. Such an approach allows us to describe how practitioners engage their *phronêsis* in ethical decision-making in a context-sensitive and dynamic environment, where the appropriate application of abstract principles will be different from case to case. This flexibility is particularly important in the context of innovation, which always entails novelty.

The structure of the paper is as follows. First, the section “Virtue and the Smart City” briefly reviews the existing literature and describes our interpretation of virtue ethics in the context of data-driven innovation in the smart city. Second, the section “Dialogic Case Study” briefly describes our methodological approach and the cases studied. Third, the section “Phronêsis for the Data Scientist” analyzes the two cases and proposes four examples of how thinking through virtues can help bridge the gap between abstract ethical principles and algorithmic choices made by the data scientist. Finally, we provide concluding remarks in the “Discussion” section.

## 2. Virtue and the Smart City

We next briefly describe our understanding of two key concepts of Aristotle’s ethics,<sup>2</sup> virtue and *phronêsis*, which will help us sketch out a theoretical frame to translate ethical principles into more practical terms.

<sup>2</sup> It should be noted that there is no one single correct way to understand Aristotle’s ethics. Centuries of prolific scholarship has produced a variety of interpretations of and elaborations on “Aristotelian ethics”. We focus on the *Nicomachean Ethics* as our source. Aristotle (1984).

Our choice to apply Aristotelian language and concepts to the ethics of innovation and technology is by no means unprecedented. Virtue ethical approaches to these domains have been explored and convincingly defended in recent years, especially by Vallor (2016) and others (such as Blok et al., 2015; Sand, 2018; Barford, 2019; Costello, 2019; Reijers, 2020; Frigo et al., 2021; Astola et al., 2022). As Barford summarizes: “the human qualities studied by virtue ethics—e.g., personal character and habits of acting and choosing with practical wisdom so as to promote human flourishing and not only one’s personal goals or those of one’s employer, to judge and to prioritize—are required successfully to engage in values-based ethical design [of ICT systems],” p. 1.

Having said that, there is a frequent criticism leveled at virtue ethical theories regarding their alleged inability to effectively guide action (see, e.g., Loudon, 1997). However, as Sand (2018) responds to this criticism, virtue ethics can successfully contribute to addressing problems of applied ethics by generating moral knowledge. Further, as Sand also points out, effective action guidance is a more general problem for all normative theories when applied to wicked problems and not specific to virtue theories. Therefore we, alongside Annas (2011) and Vallor (2016), find it useful to apply analysis inspired by virtue ethics in the context of technological innovation.

In Aristotelian thinking, virtue is a deliberate practice—a way of acting—that enables its possessor consistently to excel in their function. While virtues are more than mere “simple dispositions to engage in certain behaviors stereotypical of the virtues” (Russell, 2014, p. 203), it is critical to note that they do come about as the result of habituation (*êthos*) or “the repeated performance and practice of the actions typical of the virtues” (Basilio, 2021, p. 531). Further, there is a distinction between virtues of character and intellectual virtues. The exercise of specifically virtues of character is what enables one to perform one’s function well, while the exercise of intellectual virtue enables one to arrive at the truth. It is important to stress here that this notion of virtue is one crucially involving *activity* and not of only *being* (explored by, e.g., Broadie, 1993). Merely possessing a virtue is not sufficient, rather it must be accompanied by activity.

Being virtuous in one’s practice is a deliberate balance. For Aristotle, virtue of character is a mean state that is destroyed by the extremes of deficiency or excess. For example, courage is the virtue (mean) that lies at the right distance from foolhardiness (excess) and cowardice (deficiency). It refers to the right balance of drive and caution *appropriate for the situation at hand*. Virtue does not necessarily mean being equidistant from the extremes, and the appropriate position shifts from situation to situation. It is a matter of judgement, or what Aristotle refers to as the intellectual virtue of *phronêsis*, practical reason, to determine the appropriate balance or mean between deficiency and excess for one’s specific situation.

A city’s ethical principles for data or AI use can be understood as representations of the (aspirational) virtues of the smart city. The successful exercise of *phronêsis* can then be understood as finding the right balance and a way to apply the abstract ethical principles in the concrete practices and processes. So understood, it becomes imperative for our aim—which is to bridge the gap between principle and practice—to consider in more detail the exercise of *phronêsis*, to which task we turn next through employing a dialogic case study.

### 3. Dialogic Case Study Approach

Our goal in this paper is to illustrate how a virtue-based approach can bridge the gap between high-level principles for the city as a whole with the hands-on work of data scientists and others as they engage in the practices of data-driven innovation. In what follows, we aim to illustrate how virtues operate on a practical level, namely with regard to the individuals—in our cases, data scientists—whose actions make up parts of the practices that are data-driven innovation. To study this, we explore two data-driven innovation cases being developed by the data and analytics team in the city of Helsinki.

To study these cases, we use a bidirectional dialogic case study approach (Rule and John, 2015). Our purpose here is to comment on the theory using the two concrete example cases and more generally to show how to bridge the gap between theory and practice. With this approach, the cases and theory engage

in two ways: by both building the theory from the cases and by testing it with the cases—enabling a two-way commentary.

The first case involves advanced data analytics developed to provide better quality services for the long-term unemployed. The context for this case is a large pilot project in Finland in which the responsibility of employment services is being transitioned from a national agency to the level of individual municipal governments.<sup>3</sup> This case involves personal data collected about individuals and their use of employment and other services provided by the city. The purpose of the algorithm developed is to find a personalized path for job-seekers towards employment or other occupation. The innovative aspect in this case is in aiming to include, in addition to simple employment services use data, also data from complementary sources that the city controls. This combination would allow for more personalized and predictive service provision, with the stated goals including better served city residents.

The second case involves developing a machine learning algorithm to enable third-party event organizers more effectively to connect with residents and visitors of the city. This case involves event description data and the purpose of the algorithm in this case is to automatically and meaningfully classify and categorize the events. The innovative aspect here is that, instead of a more typical big data approach where access to resources largely determines success, the aim is more equitable and non-discriminatory treatment of both the two groups, event organizers, and potential audiences. For example, on the organizer side the goal includes providing equal visibility to different parties regardless of their characteristics such as size of marketing budget. Meanwhile, on the audience side, the goal includes deliberate exposure to events—search results and recommendations—that is, accurately tailored to an individual's specific interests given that those interests may appear as negligible niche outliers in a mass data analysis approach and, as a result, get filtered out of view.

For the purposes of this theoretical commentary, we relied on a single informant in the city of Helsinki, a data scientist representing the project coordination and data science perspectives, with whom we had three, approximately 90-min interview sessions. We also used project documentation and publicly available online materials to gather additional perspectives. In addition, the authors are former employees of the city of Helsinki with first-hand experience and knowledge of the data governance, management, strategy, and data policy perspectives both in general and regarding these specific cases.<sup>4</sup>

To analyze the case materials, we used an iterative approach. In between each interview session, we re-examined our case materials and iteratively developed an updated synthesis bridging the practices evidenced in cases and theories. In particular, we analyzed the examples of algorithmic choices described by the informant and mapped them using Vallor's (2016) technomoral virtues. We then engaged again in a dialogue with the informant to gather additional information and missing perspectives. After three iterations, our analysis saturated and resulted in the identification of the four common themes as presented in Table 1.

It is important to note here that cities are typically rather conservative organizations and are also highly regulated by national and/or European Union legislation. This is especially the case with regard to the use of (personal) data and its combination from distinct registers. While our chosen cases may at first sound like fairly simple or ordinary data science projects, they in reality require novel and innovative approaches involving technologies, service design, legal interpretation, recruitment strategy, procurement standards, overall policy, and management in the city government.

#### 4. *Phronêsis* for the Data Scientist

We next turn into the results of our study. Overall, we find that data scientists, as practitioners of data-driven innovation, must engage their practical reason, *phronêsis*, to make ethically charged decisions or,

<sup>3</sup> <https://tem.fi/en/local-government-pilots-on-employment>

<sup>4</sup> However, at the time of publication, we are no longer employed by the city government. We mention our former experience for the sake of transparency to the reader and assert that our past professional experience has not compromised our scientific objectivity and has in fact contributed to the overall quality of our research.

**Table 1.** Four data-driven innovation virtues linking technomoral virtues and technological choices

Technomoral virtues	Example virtues of data-driven innovation	Some examples of technological choices
<ul style="list-style-type: none"> <li>• Perspective (holding on to the moral whole)</li> <li>• Empathy (compassionate concern for others)</li> <li>• Care (loving service to others)</li> </ul>	Responsible use of personal data	For example, the level and kind of anonymization
<ul style="list-style-type: none"> <li>• Honesty (respecting truth)</li> <li>• Humility (knowing what we do not know)</li> </ul>	Evidence-based decisions	For example, fit and generalizability of the model
<ul style="list-style-type: none"> <li>• Justice (upholding rightness)</li> <li>• Empathy (compassionate concern for others)</li> </ul>	Fair treatment of all actors	For example, skewness of training data
<ul style="list-style-type: none"> <li>• Perspective (holding on to the moral whole)</li> <li>• Flexibility (skillful adaptation to change)</li> </ul>	Service to individuals in society	Case and context specific measures

in the language of virtue, to find the appropriate balance between deficiency and excess, in the context of those decisions.

Merely exercising one's *phronêsis* does not, of course, guarantee optimal or even good results, nor is it intended to do so. Further, at times individuals may be constrained by their institutional environment in a way that hinders acting upon what one deems appropriate. While *phronêsis* is helpful when speaking about the ethics of individuals, more work is needed to understand the ethics of organizations that comprise more than just individuals (for a similar conclusion, see Mittelstadi, 2019, p. 555.)

Analysis of the two cases studied led to the identification of four common themes, described as examples of translational or bridging virtues of data-driven innovation (see the middle column of Table 1). These virtues can be considered as specific instances of one or more of Vallor's more general technomoral virtues, which are mentioned in the leftmost column. Vallor's still fairly generic virtues turned out to be most useful as an approximation and a stand-in for a set of data ethics principles at the city. Finally, via these more specific, bridging virtues of data-driven innovation, we are able to make much more direct and immediate connections to the particular technical and algorithmic choices that the data scientist faces on a day-to-day basis (visually the middle column is the bridge between the two ends). Examples of these choices present in our cases are given in the right-most column.

It bears emphasizing again here that each row of this table represents an illustrative example of the kind of bridge we seek, and each cell in the middle column likewise is an example that we happened upon in the course of this particular research effort. Neither the set of rows or the set of what we call "virtues of data-driven innovation" in the middle column are to be considered exhaustive or comprehensive—nor indeed are the virtues mentioned not in need of (further) moral justification. Our purpose here is only to visualize the missing layer between principles and practices and to show how the space between the leftmost and rightmost columns might be filled out in the instance of this specific data scientist and these two specific cases. Next, we briefly discuss each of the four virtues of data-driven innovation mentioned in the middle column.

#### 4.1. Responsible use of personal data

This virtue concerns practically all cases involving personal data and, due to the requirements of the EU GDPR, it is one of the best-known topics in data-driven innovation in Europe. Personal data has

considerable value and utility potential while being subject to significant privacy protection requirements to safeguard against abuse (see, e.g., Politou et al., 2018). Further, responsibility and the closely linked notion of accountability are key aspects of nearly all existing sets of AI ethics principles.

In terms of the general technomoral virtues, this virtue relates to holding on to the moral whole (perspective) in that focus must not narrow too much to either maximizing value (data abuse, or excessive exploitative data practices) or minimizing potential harm (denial of data, or refusal to process personal data at all) and to being compassionate for others (empathy) in that it must always be kept in mind that the data are intimately connected with real, flesh-and-blood people.

The technical process of adjusting the level of anonymization, for which a plurality of techniques pose additional challenges, is an example of a data scientist's exercise of *phronêsis*. By choosing an appropriate level and type of anonymization the data scientist can balance the information requirements by the algorithm in question while preserving privacy of data subjects.

#### 4.2. Evidence-based decisions

Data-driven management in a public-sector organization involves dealing with heterogeneous, incomplete, and/or poor-quality data, as well as approximations and probabilities. The practitioners responsible for choosing, processing, and visualizing the relevant data as evidence for decision-makers or developing automated models must therefore exercise this virtue to be successful in delivering the best possible results, that is, the evidence as well as description of its limitations.

In terms of the general technomoral virtues, this relates to respecting truth (honesty), in that evidence is sought and given due weight in decision-making processes, and to knowing what we do not know (humility), in that limitations of any data presented as evidence are included in that presentation.

An example of a method at the data scientist's disposal in preparing data and developing models are the adjustments to the generalizability of the model used to generate evidence. Excess in this case is hyper-precision by overfitting the model. Overfitting makes the model too specific and hinders its generalizability and so limits its utility for decision-makers. Deficiency in this regard is underfitting, which fails to capture the essential trends in the data and so produces results that are so general as to be meaningless. *Phronêsis* is needed here for finding the appropriate fit.

#### 4.3. Fair treatment of all actors

An inherent challenge with data science, and its application in data-driven innovation, is that biased data leads to biased results. Bias, while inevitable, cannot be allowed to result in discriminatory, unequitable, or unfair practices or policies. So one of the challenges facing the data scientist is recognizing bias where it appears—in the data and its results—and responding appropriately in order for the bias not to cause unwanted treatment of the people involved. More broadly, ensuring both individual and group fairness at the same time is challenging (Dwork et al., 2012) but not necessarily conflicting and may often in fact be a case of mere misconception that can be solved with a nuanced consideration of the sources of unfairness (Binns, 2020).

In terms of the general technomoral virtues, this relates to upholding rightness (justice), in that the goal is to prevent injustice and discriminatory effects, and having compassionate concern for others (empathy), in that the social and other context of the people involved will require adjusting for.

To practice this virtue, the data scientist, for example, has to understand the possible skewness of training data and use normalization, over-sampling, or other means to balance class representation. To find the right corrective measures in the service of the vulnerable in particular is key here. This virtue is a prime example of how important the individual exercise *phronêsis* is in the realization of the larger goals of data-driven innovation, as one cannot rely on one exact mathematical measure to take.

#### 4.4. Service to individuals in society

When designing and developing public services, one needs to weigh the interests of the individual against those of society or larger community. For a public-sector body, resources are limited and individually

```

# define validation loss
def km_loss(delta, N, k, a = 1):
    """
    delta: train/test loss metric, dissimilarity measure
    between data point and cluster centroid
    N: number of data points
    k: number of clusters
    a: regularization coefficient
    """
    return (delta + a*k**2)/N

```

**Figure 1.** Code excerpt for balancing between community and individual utility.

tailored care and services are not often possible within the bounds of those resources. The task in these cases is to ensure that most people are adequately served while no one, regardless of the complexity of their needs, is left behind.

In terms of the general technomoral virtues, this relates to holding on to the moral whole (perspective), in that one must retain the big picture of societal well-being while caring for specific people and their needs, and skillful adaptation to change (flexibility), in that the right approach to any given case will have to be dynamic and react to changes in the environment and society as well as emerging individual needs.

Excess in this virtue is a kind of erasure of individual experience and difference, while deficiency is prioritizing the individual above all other concerns. For this virtue, which could be described as the right consideration for the collective with respect to the individual and vice versa, we did not observe a single common mechanism for the data scientist to deploy in exercising their *phronêsis* as was the case for the three preceding virtues. We did, however, observe two particular actions taken.

The first was to parametrize the loss function<sup>5</sup> in a clustering algorithm (see the code excerpt from our case in Figure 1). By choosing different values for the parameters, a data scientist can balance between a preference for larger or smaller clusters. In practice, preferring larger clusters means providing a more generic outcome that approximates the need of more people at once and preferring smaller clusters produces more precisely targetable groups.

The second action we observed was in response to the observation that algorithms often tend to favor the general masses and that, as a result, rare and exceptional instances are not found. This is a fundamental problem in all reasoning based on statistical analysis, but it takes various forms in different contexts and so there are multiple approaches possible to address it. The action we observed was to add randomness. By increasing randomness, a data scientist can balance toward niche categories by equally suggesting some unexpected and rare instances.

To conclude this section, some summary remarks. In the course of our work, we saw that our informant was continuously exercising their *phronêsis* in addition to their technical skills and knowledge to be able to do their work. We conceptualize this as their activity of attempting to find the appropriate balance between deficiency and excess, that is, navigating the landscape of virtue with *phronêsis*. In the first case study, we noticed that a balance was being sought between sufficient anonymization of personal data of the people using employment services and the usefulness of the data in terms of information richness and utility. In the second case, the balance to be struck involved the generalizability of the ML-based classification model and how well the model finds specific solutions.

## 5. Discussion

In this commentary, we have described theoretical and practical ways in which we can use virtue ethics to understand and to inform practices of data-driven innovation in the public sector. We have described a

<sup>5</sup> A loss function is a method of evaluating how well an algorithm (e.g., a machine learning model) predicts or models a featured data set.

theoretical grounding for our approach that is inspired by our reading of Aristotle, and an illustrative application of that approach.

Our motivation in developing our approach has been to bridge the gap between abstract ethical principles, which are increasingly common for organizations practicing data-driven innovation, and the practical choices that data scientists and other practitioners make every day in order to do their jobs. We have done so by using a dialogic case study approach (Rule and John, 2015) to apply conceptual tools from virtue ethics in our study of two concrete cases of data-driven innovation at the city of Helsinki. Through this approach, we identified four themes and described four corresponding, illustrative “virtues of data-driven innovation.” In doing so, we showed how our approach might handle bridging the gap between general principles and concrete practices of civil servants as they engage their *phronêsis* to produce virtuous outcomes.

A major insight gleaned from exposing our theory to our practitioner’s first-hand experience was the deepened appreciation of how essential the individual practitioners’ moral sensitivity and judgement, *phronêsis*, is especially in contexts where the explicit and acknowledged structures and processes for translating values and principles into practice are largely absent, as is most often the case in organizations today. We believe that increasing visibility of, verbalizing, and adding transparency to processes of data-driven innovation will help not only understand but also to govern them better. This importance is further intensified in cases of data-driven innovation in which by definition novel and unknown contexts are explored.

Our work contributes within the field of practicing ethics in research and innovation and it addresses the two commonly cited issues facing methods that attempt to introduce or strengthen ethical thinking and doing in the phases of innovation design and testing (so-called intra methods): their lack of theoretical grounding and their inability to incorporate ethics into the day-to-day work of innovation practitioners (Reijers et al., 2018).

Further, our work relates to the ethical discussion around innovation that focuses specifically on the notion of responsibility (e.g., Coeckelbergh, 2006; Swierstra and Jelsma, 2006; Grinbaum and Groves, 2013; Simakova and Coenen, 2013; Von Schomberg, 2013; Ferrari and Marin, 2014; Grunwald, 2014; Pavie, 2014; Sand, 2016). While we have refrained from framing our work explicitly around the term ‘responsibility’, understood as “the implementation of ethical values, transparency, reflexivity, participation and more generally bottom-up governance of innovation” (Sand, 2018, p. 92), we feel the notion is integral to our work which uses the language of virtue.

Finally, we have observed a specific interest in the relevant literature to focus on individual innovators and pioneering developers of new technologies (e.g., Gardner, 1993; Sand, 2018). Our research differs from this approach in that the practitioner of innovation in our focus is not the once-in-a-generation genius but rather the scientifically skilled civil servant. Our discussion of individual virtues, therefore, looks very different and so has a different scope of applicability.

Our work comes with limitations. First, as the main data source, we have chosen to focus on an individual person with a specific role in these innovation processes. We recognize the fact that processes and practices of innovation, perhaps especially in the context of the public sector, are highly collaborative, complex, and collective. This limitation also provides an opportunity for future research. We suggest that our method can equally be applied to understand also any other actors playing different roles in these complex systems, such as chief information officers, data architects, policy experts, data protection lawyers, procurement specialists, service designers, or elected political leaders. Additionally, it is possible to explore applications of this approach to collective processes in addition to different individuals participating in those processes.

Second, the four data-driven innovation virtues we discussed are by no means an exhaustive or a comprehensive list. Some of the relevant literature is motivated in part by the identification and even classification of specific virtues that are especially relevant for innovation, such as creativity (Sand, 2018; Astola et al., 2022). In contrast, rather than focusing on specific virtues, we want to point out that there is a need for identifying the layer in between the abstract and general ethical principles and the concrete and practical algorithmic choices.

Further, the examples we presented are not meant to be generalizable, abstractable, or universally applicable: that would be sliding back into principlism. Instead, they are the results of applying the language and theoretical tools provided by virtue ethics to the actual practice of, for example, writing code by a data scientist. This application is provided to demonstrate that virtue theoretical tools can fruitfully be used to render visible and graspable what is usually considered opaque or “black box” about innovation processes, namely what goes on between having a principle (like “responsibility”) on the one hand, and the individual data scientist’s actual practice (like the decision to opt for a straightforward substitution technique for data masking rather than adopting a full-blown encryption approach).

The policy implications of this work are for both decision-makers and practitioners of data-driven innovation in the public-sector context who are committed to ethical practice. We join Hemerly (2013) and Meyer (2015), both of whom argue for flexible and reasonable policymaking that takes the context into account, and suggest that a virtue-based approach can support their efforts to connect abstract and unchanging principles with contextual and shifting demands of day-to-day practice.

**Acknowledgments.** The authors are grateful for the case studies generously described to us by Nuutti Kytö (né Sten).

**Funding statement.** The authors declare none.

**Competing interest.** Both authors have been previously employed by the city of Helsinki.

**Author contribution.** Conceptualization: V.L., K.K.; Methodology: V.L., K.K.; Writing original draft: V.L., K.K. All authors approved the final submitted draft.

**Data availability statement.** Data availability is not applicable to this article as no new data were created or analyzed in this study.

## References

- Ada Lovelace Institute** (2021) AI Now Institute, Open Government, and Partnership. ‘Algorithmic Accountability for the Public Sector’. Accessed: 2021-10-29. Available at <https://ainowinstitute.org/pages/algorithmic-accountability-for-the-public-sector-report.html>.
- Albrechtslund A** (2007) Ethics and technology design. *Ethics and Information Technology* 9(1), 63–72. <https://doi.org/10.1007/s10676-006-9129-8>
- Annas J** (2011) *Intelligent Virtue*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199228782.001.0001>
- Aristotle** (1984) The complete works of Aristotle. In Barnes J (ed.), *The Revised Oxford Translation*, Vol. 2. Princeton, NJ: Princeton University Press.
- Astola M, Bombaerts G, Spahn A and Royakkers L** (2022) Can creativity be a collective virtue? Insights for the ethics of innovation. *Journal of Business Ethics* 179(3), 907–918. <https://doi.org/10.1007/s10551-021-04833-0>
- Barford L** (2019) Contemporary virtue ethics and the engineers of autonomous systems. In *2019 IEEE International Symposium on Technology and Society (ISTAS)*. Medford, MA, USA: IEEE, pp. 1–7. <https://doi.org/10.1109/ISTAS48451.2019.8937855>
- Basilio GD** (2021) Habituation in Aristotle’s ethics: The Eudemian ethics, the common books, the Nicomachean ethics. *Journal of the History of Philosophy* 59(4), 531–557. Available at <https://www.proquest.com/docview/2590215750/abstract/F1B5D5F0EB714B18PQ/3>.
- Binns R** (2020) On the apparent conflict between individual and group fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York, NY: Association for Computing Machinery, pp. 514–524. <https://doi.org/10.1145/3351095.3372864>
- Blok V, Gremmen B and Wesselink R** (2015) Dealing with the wicked problem of sustainability: The role of individual virtuous competence. *Business & Professional Ethics Journal* 34(3), 297–327. Available at <http://www.jstor.org/stable/44074859>.
- Borning A and Muller M** (2012) Next steps for value sensitive design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, pp. 1125–1134. <https://doi.org/10.1145/2207676.2208560>
- Brey P** (2000) Disclosive computer ethics. *ACM SIGCAS Computers and Society* 30(4), 10–16. <https://doi.org/10.1145/572260.572264>
- Broadie S** (1993) *Ethics with Aristotle*. New York, USA: Oxford University Press.
- Coeckelbergh M** (2006) Regulation or responsibility? Autonomy, moral imagination, and engineering. *Science, Technology, & Human Values* 31(3), 237–260. <https://doi.org/10.1177/0162243905285839>
- Costello GJ** (2019) The philosophy of innovation in management education: A study utilising Aristotle’s concept of Phronesis. *Philosophy of Management* 18(3), 215–230. <https://doi.org/10.1007/s40926-018-00104-7>

- Dwork C, Hardt M, Pitassi T, Reingold O and Zemel R** (2012) Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. New York, NY: Association for Computing Machinery, pp. 214–226. <https://doi.org/10.1145/2090236.2090255>
- Ferrari A and Marin F** (2014) Responsibility and visions in the new and emerging technologies. In Arnaldi S, Ferrari A, Magaudda P and Marin F (eds), *Responsibility in Nanotechnology Development (The International Library of Ethics, Law and Technology)*. Dordrecht, Netherlands: Springer, pp. 21–36. [https://doi.org/10.1007/978-94-017-9103-8\\_2](https://doi.org/10.1007/978-94-017-9103-8_2)
- Frigo G, Marthaler F, Albers A, Ott S and Hillerbrand R** (2021) Training responsible engineers. Phronesis and the role of virtues in teaching engineering ethics. *Australasian Journal of Engineering Education* 26(1), 25–37. <https://doi.org/10.1080/22054952.2021.1889086>
- Gardner H** (1993) *Creating Minds: An Anatomy of Creativity Seen Through the Lives of Freud, Einstein, Picasso, Stravinsky, Eliot, Graham and Gandhi*. New York: Basic Books.
- Grinbaum A and Groves C** (2013) What is “responsible” about responsible innovation? Understanding the ethical issues. In *Responsible Innovation*. Chichester, UK: John Wiley & Sons, Ltd, pp. 119–142.
- Groves C** (2015) Logic of choice or logic of care? Uncertainty, technological mediation and responsible innovation. *NanoEthics* 9(3), 321–333. <https://doi.org/10.1007/s11569-015-0238-x>
- Grunwald A** (2014) The hermeneutic side of responsible research and innovation. *Journal of Responsible Innovation* 1(3), 274–291. <https://doi.org/10.1080/23299460.2014.968437>
- Hagedorff T** (2022) Blind spots in AI ethics. *AI and Ethics* 2(4), 851–867. <https://doi.org/10.1007/s43681-021-00122-8>
- Hemerly J** (2013) Public policy considerations for data-driven innovation. *Computer* 46(6), 25–31. <https://doi.org/10.1109/MC.2013.186>
- Le Dantec CA, Poole ES and Wyche SP** (2009) Values as lived experience: Evolving value sensitive design in support of value discovery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, pp. 1141–1150. <https://doi.org/10.1145/1518701.1518875>
- Louden RB** (1997) On some vices of virtue ethics. In Crisp R and Slote M (eds), *Virtue Ethics*. Oxford: Oxford University Press.
- Manders-Huits N** (2011) What values in design? The challenge of incorporating moral values into design. *Science and Engineering Ethics* 17(2), 271–287. <https://doi.org/10.1007/s11948-010-9198-2>
- Meyer M** (2015) Two cheers for corporate experimentation: The a/B illusion and the virtues of data-driven innovation. *Colorado Technology Law Journal* 13, 273.
- Mittelstadt B** (2019) Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Page K** (2012) The four principles: Can they be measured and do they predict ethical decision making? *BMC Medical Ethics* 13(1), 10. <https://doi.org/10.1186/1472-6939-13-10>
- Pavie X** (2014) The importance of responsible innovation and the necessity of “innovation-care”. *Philosophy of Management* 13(1), 21–42. <https://doi.org/10.5840/pom20141313>
- Politou E, Alepis E and Patsakis C** (2018) Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity* 4(1), ty001. <https://doi.org/10.1093/cybsec/ty001>
- Reijers W** (2020) Responsible innovation between virtue and governance: Revisiting Arendt’s notion of work as action. *Journal of Responsible Innovation* 7(3), 471–489. <https://doi.org/10.1080/23299460.2020.1806524>
- Reijers W, Wright D, Brey P, Weber K, Rodrigues R, O’Sullivan D and Gordijn B** (2018) Methods for practising ethics in research and innovation: A literature review, critical analysis and recommendations. *Science and Engineering Ethics* 24(5), 1437–1481. <https://doi.org/10.1007/s11948-017-9961-8>
- Rességuier A and Rodrigues R** (2020) AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society* 7(2), 2053951720942541. <https://doi.org/10.1177/2053951720942541>
- Rule P and John VM** (2015) A necessary dialogue: Theory in case study research. *International Journal of Qualitative Methods* 14(4), 1609406915611575. <https://doi.org/10.1177/1609406915611575>
- Russell DC** (2014) Phronesis and the virtues (NE vi 12–13). In Polansky R (ed.), *The Cambridge Companion to Aristotle’s Nicomachean Ethics*. Cambridge: Cambridge University Press, pp. 203–220. <https://doi.org/10.1017/CCO9781139022484.010>
- Sand M** (2016) Responsibility and visioning—Opening Pandora’s box. *NanoEthics* 10(1), 75–86. <https://doi.org/10.1007/s11569-016-0252-7>
- Sand M** (2018) The virtues and vices of innovators. *Philosophy of Management* 17(1), 79–95. <https://doi.org/10.1007/s40926-017-0055-0>
- Shattuck J** (1984) Computer matching is a serious threat to individual rights. *Communications of the ACM* 27(6), 538–541. <https://doi.org/10.1145/358080.358090>
- Simakova E and Coenen C** (2013) Visions, hype, and expectations: A place for responsibility. In *Responsible Innovation*. New York, NY: John Wiley & Sons, Ltd, pp. 241–267. <https://doi.org/10.1002/9781118551424.ch13>
- Swierstra T and Jelsma J** (2006) Responsibility without moralism in technoscientific design practice. *Science, Technology, & Human Values* 31(3), 309–332. <https://doi.org/10.1177/0162243905285844>
- Ten Have HAMJ** (2001) Theoretical models and approaches to ethics. In ten Have H and Gordijn B (eds) *Bioethics in a European Perspective. International Library of Ethics, Law, and the New Medicine*. Dordrecht, Netherlands: Springer, pp. 51–82. [https://doi.org/10.1007/978-94-015-9706-7\\_3](https://doi.org/10.1007/978-94-015-9706-7_3)

- Vakkuri V, Kemell K-K, Kultanen J, Siponen M and Abrahamsson P** (2019) Ethically Aligned Design of Autonomous Systems: Industry Viewpoint and an Empirical Study. arXiv:1906.07946 [cs.CY] <https://doi.org/10.48550/arXiv.1906.07946>
- Vallor S** (2016) *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford, USA: Oxford University Press.
- van de Poel I** (2013) Translating values into design requirements. In Michelfelder DP, McCarthy N and Goldberg DE (eds), *Philosophy and Engineering: Reflections on Practice, Principles and Process*. *Philosophy of Engineering and Technology*. Dordrecht, Netherlands: Springer, pp. 253–266. [https://doi.org/10.1007/978-94-007-7762-0\\_20](https://doi.org/10.1007/978-94-007-7762-0_20)
- Von Schomberg R** (2013) A vision of responsible research and innovation. In *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. London: Wiley, pp. 51–74. <https://doi.org/10.1002/9781118551424.ch3>