Stephen Downes

Ethics, analytics, and the duty of care

(doi: 10.53227/108467)

Rivista di Digital Politics (ISSN 2785-0072)
Fascicolo 2, maggio-agosto 2023

Ente di afferenza:
()

Stephen Downes

Ethics, analytics, and the duty of care

ETHICS, ANALYTICS, AND THE DUTY OF CARE

Artificial intelligence (Ai) and Learning analytics have raised a host of ethical issues and a renewed attention to matters such as fairness, justice, and benevolence. This paper offers a comprehensive analysis of these topics, surveying the applications of Ai and analytics, with a focus on learning technology, and listing the ethical issues that have been raised. This is followed by an analysis of the ethical decision points that arise in the design of Ai and analytics, a study of relevant ethical codes for related professions, and an overview of the theories of ethics underlying those codes, leading to a contemporary analysis based in a philosophy of care ethics, and concluding with a discussion of ethical practices.

KEYWORDS *Analytics, Ai, Ethics, Care, Community, Culture.*

1. The joy of ethics

Ethics should make us joyful, not afraid. Ethics is not about what's wrong, but what's right. It speaks to us of the possibility of living our best life, of having aspirations that are noble and good, and gives us the means and tools to help realise that possibility. We spend so much more effort trying to prevent what's bad and wrong when we should be trying to create something that is good and right.

Similarly, in learning analytics, the best outcome is achieved not by preventing harm, but rather by creating good. Technology can represent the best of us, embodying our hopes and dreams and aspirations. That is the reason for its existence. Yet, «classical philosophers of technology have painted an excessively gloomy picture of the role of technology in contemporary culture», writes Verbeek (2005, 4). What is it we put into technology and what do we expect when we use it? In analytics, we see this in sharp focus.

Stephen Downes, National Research Council Canada, Ottawa, Canada, email: Stephen.Downes@nrc-cnrc.gc.ca, orcid: 0000-0001-6797-9012.

ISSN 2785-0072

Ethics, at first glance, appears to be about «right» and «wrong», perhaps as discovered (Pojman 1990), perhaps as invented (Mackie 1983). The nature of right and wrong might be found in biology, rights, fairness, religion, or any number of other sources, depending on who is asked. Or instead, ethics may be based in virtue and character, as described by Aristotle (350-2003, I.III) in ancient Greece. Either way, ethics is *generally* thought of as speaking to what actions we «should» or «ought» to take (or «should not» or «ought not» take).

In this paper, however, I argue that ethics is based on perception, not principle. It springs from that warm and rewarding sensation that follows when we have done something good in the world. It reflects our feelings of compassion, of justice, of goodness. It is something that comes from inside, not something that results from a good argument or a stern talking-to. We spend so much effort drafting arguments and principles as though we could convince someone to be ethical, but the ethical person does not need them, and if a person is unethical, reason will not sway them.

We see the same effect in analytics. Today's artificial intelligence engines are not based on cognitive rules or principles; they are trained using a mass of contextually relevant data. This makes them ethically agnostic, but they defy simple statements of what they ought not do. And so the literature of ethics in analytics express the fears of alienation and subjugation *common* to traditional philosophy of technology. And we lose sight, not only of the good that analytics might produce, but also of the best means for preventing harm.

What, then, do we learn when we bring these considerations together? That is the topic of this essay. Analytics is a brand-new field, coming into being only in the last few decades. Yet it wrestles with questions that have occupied philosophers for centuries. When we ask what is right and wrong, we ask also how we come to know what is right and wrong, how we come to learn the distinction, and to apply it in our daily lives. This is as true for the analytics engine as it is for the person using it.

2. The benefits of Ai

The focus of this paper is the use of analytics as applied to learning and education (typically called «learning analytics»). Learning analytics is typically defined in terms of its objective, which is to improve the chance of student success (Gasevic *et al.* 2015). Accordingly, when founding the Society for learning analytics (SoLAR) George Siemens defined learning analytics as «the measurement, collection, analysis and reporting of data about learners

and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs» (Siemens 2012).

From the perspective of ethics in analytics, it may be wisest to adopt a broad definition of «learning analytics». After all, as Griffiths *et al.* (2016) argue, the Jisc Code of Practice for Learning Analytics uses a wider definition of using data about students and their activities «to help institutions understand and improve educational processes and provide better support to learners» (Sclater and Bailey 2015; 2023). Hence, we will frequently refer to «Ai» or «Ai and analytics» as a form of shorthand to capture this broad perspective.

Despite its sudden popularity in the fall of 2022 with the release of generative Large language models (Llm) Artificial intelligence (Ai) has been with us for many years now, beginning with Alan Turing's (1936) conceptual advances, Newell and Simon's (1959) General problem solver, and Bill Rosemblatt's (1958) Perceptron. None of these, of course, ever achieved General artificial intelligence (Gai). The term Ai references the ambition, not the outcome, of the various technologies collected under it. Hence, when we refer to Llm as Gai, we are not claiming that these have achieved Gai, but rather, that they are part of the larger research program leading to that goal.

Tentatively at first, but with increasing momentum up to the flood of innovation we see today, the Ai research program has produced algorithms and models that have been integrated into many of our tools and processes, from adaptive cruise control in cars, to fault detection systems in pipelines, to automated translation services offered by Google. It is beyond the scope of this paper to produce a comprehensive list of those benefits, but they should be acknowledged at the outset, and can be broadly categorised under five major themes.

Built on these basic capabilities are four widely used categories (Brodsky *et al.* 2015; Boyer and Bonnin 2017) to which we add additional fifth and sixth categories, generative analytics and deontic analytics:

- descriptive analytics, answering the question «what happened?»;
- diagnostic analytics, answering the question «why did it happen?»;
- predictive analytics, answering the question «what will happen?»;
- prescriptive analytics, answering the question «how can we make it happen?»;
- generative analytics, which use data to create new things, and
- deontic analytics, answering the question «what should happen?».

Descriptive analytics

Descriptive analytics include analytics focused on description, detection and reporting, including mechanisms to pull data from multiple sources, filter it, and combine it. The output of descriptive analytics includes visualisations such as pie charts, tables, bar charts or line graphs. Descriptive analytics can be used to define key metrics, identify data needs, define data management practices, prepare data for analysis, and present data to a viewer (Vesset 2018).

Diagnostic analytics

Diagnostic analytics look more deeply into data in order to detect patterns and trends. Such a system could be thought of as being used to draw an inference about a piece of data based on the patterns detected in sample or training data, for example, to perform recognition, classification or categorization tasks, for example, detecting Ai-generated faces (Li and Lyu 2019), sentiment analysis (Rientes and Jones 2019, 114), and automated grading (Lu 2019).

While some have witnessed the emergence of Ai only in 2022, those who have worked in neural networks and related technologies has seen challenge after challenge fall over the years, limited only by the limits of computer chips, data storage, and human input. Nothing in the current deployment of Ai suggests that these advances will slow anytime soon.

Predictive analytics

As the name suggests, predictive analytics uses data to extrapolate to future events. Predictive analytics can support resource planning and event response (Drew 2016), learning design (Rientes and Jones 2019, 116), and academic advising (O'Brien 2020). In contemporary analytics predictive algorithms do not rely on general rules or principles specific to the prediction, but rather, are based on models taking into account many the full range of environmental data, which may result in predictions made as a result of the interaction of thousands of variables.

Prescriptive analytics

Prescriptive analytics recommends a course of action. An oft-cited application is the potential of learning analytics to make content recommendations, either as a starting point, or as part of a wider learning analytics-supported learning path. For example, the Personalised adaptive study success (Pass) system supports personalisation for students at Open universities australia (Oua)

(Sclater *et al.* 2016). Other applications include adaptive group formation (Zawacki-Richter *et al.* 2019, 4), hiring (Metz 2020), and decision-making (Parkes 2019).

Generative analytics

Generative analytics is different from the previous four categories in the sense that it is not limited to answering questions like «what happened» or «how can we make it happen», but instead uses the data to create something that is genuinely new. In a sense, it is like predictive and prescriptive analytics in that it extrapolates beyond the data provided, but while in the former two we rely on human agency to act on the analytics, in the case of generative analytics the analytics engine takes this action on its own. Generative analytics came to the fore in 2022 with the release of products such as Stable Diffusion and ChatGpt.

Deontic analytics

There is an additional question that needs to be answered, and has been increasingly entrusted to analytics: «what ought to happen?». Recently the question has been asked with respect to self-driving vehicles in the context of Philippa Foot's «trolley problem» (Foot [1967] 1978). In a nutshell, this problem forces the reader to decide whether to take an action to save six and kill one, or to desist from action to save one, allowing (by inaction) six to be killed. It is argued that automated vehicles will face similar problems. Accordingly, researchers from Microsoft have developed the Defining issues test (Dit) to evaluate the moral reasoning capabilities of recent Ai systems such as Gpt-3 and ChatGpt (Tanmay *et al.* 2023, 1).

It may be argued that these outcomes are defined ahead of time by human programmers. For example, cars made for rich people have their ethical priorities preset: they will protect passengers, not bystanders (Morris 2016). But not all ethical outcomes will be preprogrammed; arguably, an Ai's ethical stance will often emerge as a byproduct of other priorities and activities. The very nature and existence of Ai will drive significant and social changes. As Liu *et al.* (2020, 2) write, «Ai technology raises fundamental questions of power and control across society, and has been anticipated to challenge almost every sphere of human activity, as society transitions into a «digital lifeworld». Uses and concerns range across diverse sectors, from legal decision-making and policing to health-care, transport, and military uses, to name but a few».

3. Issues related to Ai

In the previous section we hope that we have established that there is a wide range of uses for learning analytics and Ai in education, from tools institutions can use to manage resources and optimise offering through to tools individuals can use to learn more effectively and quickly. If there were no benefits to be had from analytics, then there would be no ethical issues. But in part because there are benefits, there are ethical issues. No tool that is used for anything is immune from ethical implications.

The ethics of analytics is particularly complex because issues arise both when it works, and when it doesn't. Consequently, in an approach we will follow, Narayan (2019) classifies these issues under three headings: issues that arise when analytics works, issues that arise because analytics are not yet reliable, and issues that arise in cases where the use of analytics seems fundamentally wrong. To these three sets of issues we will add a fourth describing wider social and cultural issues that arise with the use of analytics and Ai.

In this section we collected all mentions of issues related to Ai found in the literature and popular media during the study period, resulting in a relatively complete listing of expressed concerns. The criterion for inclusion in this list was only that the concern was expressed, and not whether it was in some way established or proven.

When analytics works

As Mark Liberman (2019) observes, «Modern Ai (almost) works because of machine learning techniques that find patterns in training data, rather than relying on human programming of explicit rules». This is in sharp contrast to earlier rule-based approaches that «*generally* never even got off the ground at all». As we have seen, analytics can be used for a wide range of tasks, some involving simple recognition, some involving *deeper* diagnostics, some making predictions, and some even generating new forms of content and even making determinations about what should or ought to be done. In such cases, it is the accuracy of analytics that raises ethical issues. In many cases there is a virtue in not knowing something or not being able to do something that is challenged when analytics reveals everything. The following comprise a few examples.

- Surveillance – once surveillance becomes normal – so normal it's in your street lights – it can have an impact on rights and freedoms (Shaw 2017).

- Tracking – as Cavoukian (2013, 23) writes, «it is one thing to be seen in public. It is another to be tracked by the state».
- Anonymity. It is widely argued that «anonymity helps support the fundamental rights of privacy and freedom of expression» (Bodle 2013) yet the Online Disinhibition Effect (Suler 2004), which helps students feel safe and secure and helps them «come out of their shell», has also attributed as factor in online bullying and abuse (O'Leary and Murphy 2019).
- Facial recognition. Mark Andrejevic and Neil Selwyn (2019) point to the dehumanising nature of facially focused schooling and the foregrounding of students' gender and race, among other concerns;
- Privacy issues.«The collection or aggregation of data, informed consent, de-identification of data, transparency, data security, interpretation of data, as well as data classification and management» (Griffiths *et al.* 2016, 6).
- Assessment issues. Students have mixed feelings about such systems, preferring «comments from teachers or peers rather than computers» (Roscoe *et al.* 2017).
- Lack of discretion. «Organisational actors establish and re-negotiate trust under messy and uncertain analytic conditions» (Passi and Jackson 2018, 1).
- Lack of appeal. «Will the prestige and trust placed in machines, often assumed to be «neutral» and fail-proof, tempt us to hand over to machines the burden of responsibility, judgement and decision-making?» (Demiaux and Abdallah 2017, 5).
- Content manipulation. Arguably technologies like Deepfakes are «a looming challenge for privacy, democracy and national security» (Chesney and Citron 2018, 1760).
- User manipulation. Ai «could start chatting with you – actually, experimenting on you – to test what content will elicit the strongest reactions (and) could easily prey on wide swaths of the public for years to come» (Paul and Posard 2020).

When analytics fails

As Mark Liberman (2019) comments, Ai is brittle. When the data are limited or unrepresentative, it can fail to respond to contextual factors our outlier events. It can contain and replicate errors, be unreliable, be misrepresented, or even defrauded. In the case of learning analytics, the results can range

from poor performance, bad pedagogy, untrustworthy recommendations, or (perhaps worst of all) nothing at all. Some examples:

- Error. «Questions arise about who is responsible for the consequences of an error, which may include ineffective or misdirected educational interventions» (Griffiths *et al.* 2016, 4).
- Unreliable data. Analytics requires reliable data, «as distinguished from suspicion, rumour, gossip, or other unreliable evidence» (Emory university libraries 2019) but this is often not the case.
- Consistency failure. Many analytics systems operate over distributed networks. As such, there may be cases where part of the network fails. This creates the possibility of a consistency failure (Gilbert and Lynch 2002, 51).
- Bias. «Machine learning algorithms are picking up deeply ingrained race and gender prejudices concealed within the patterns of language use, scientists say» (Devlin 2017).
- Misinterpretation. An Ai misinterprets laptop placement as «paying attention» (Metz 2020).
- Misrepresentation. For example, the Scientific content analysis (Scan), the creator of which says the tool can identify deception. However, a scientific review of the system found the opposite (Armstrong and Sheckler 2019; Brandon *et al.* 2019).
- Distortion. For example, recommendation engines that lead to radicalization. «It seems as if you are never «hard core» enough for YouTube's recommendation algorithm (Tufekci 2018).
- Bad pedagogy. There is a risk, writes Ilkka Tuomi (2018), «that Ai might be used to scale up bad pedagogical practice». For example, badly constructed analytics may lead to evaluation errors. «Evaluation can be ineffective and even harmful if naively done "by rule" rather than "by thought"» (Dringus 2012, 89).

Bad actors

Bad actors are people or organisations that attempt to subvert analytics systems. They may be acting for their own benefit or to the detriment of the analytics organisations or their sponsors. The prototypical bad actor is the hacker, a person who uses software and infiltration techniques to intrude into computer systems. Bad actors create ethical issues for analytics because they demonstrate the potential to leverage these systems to cause harm.

- Conspiracy theorists. These often replicate analytical methods and dissemination, and sometimes subject existing analytics for their own purposes (Yeung 2023).
- Stalkers. The use of facial recognition systems such as Clearview, could be used to track people down. «Making these things for public consumption puts survivors at risk, and we need to think about the unintended consequences» (Shwayder 2020).
- Collusion. Analytics engines working in concert can become bad actors in their own right. For example, Calvano *et al.* (2020, 3267) showed that «algorithms powered by Artificial intelligence (Q-learning) in a workhorse oligopoly model of repeated price competition».

When analytics is fundamentally dubious

Narayan (2019) describes the following «fundamentally dubious» uses of learning analytics: predicting criminal recidivism, policing, terrorist risk, at-risk kids, and predicting job performance. «These are all about predicting social outcomes», he says, «so Ai is especially ill-suited for this». There are good examples of cases where analytics fail in such cases; Narayan cites a study by that shows «commercial software that is widely used to predict recidivism is no more accurate or fair than the predictions of people with little to no criminal justice expertise» (Dressel and Farid 2018, 3). Even if analytics gets it right, there is an argument to be made that it should not be applied in such cases or applied in this way.

- Predictive policing. «Police officers could also inadvertently use their perceptions of students who appear on the list to make decisions about how to adjudicate a crime that takes place» (Lieberman 2020).
- Racial profiling. «Ai systems used to evaluate potential tenants rely on court records and other datasets that have their own built-in biases that reflect systemic racism, sexism, and ableism» (Akselrod 2021).
- Identity graphs. Hamel (2016) asks «Is it legal and ethical for 3rd parties to build consumer profiles from your social and online presence, merge it with their own internal data, credit scores and any other data sources they can find?».
- Autonomous weapons, not just in warfare. Ai-enabled learning management systems, for example, to prohibit cheating (Otavec 2022), enforce copyright regulations, or regulate unauthorised access to le-

arning materials using automated public disclosure of information or retaliatory measures such as malware or viruses.

Social and cultural issues

This is a class of issues that addresses the social and cultural infrastructure that builds up around analytics. These are not issues with analytics itself, but with the way analytics changes our society, our culture, and the way we learn.

- Opacity. Failing to adhere to a principle of «notification when an Ai system makes a decision about an individual» allowing individuals to «experience the advantages of Ai, as well as to opt out of using such products should they have concerns» (Fjeld *et al.* 2020, 45).
- Alienation. For example, «the difficulty in reaching a real person» (Guillaud 2020). Or for example, demeaning assessment processes – «You are not willing to even have someone at your firm look at my résumé» (Keppler 2020).
- Non-explanability. If an Ai system has a «substantial impact on an individual's life» and cannot provide «full and satisfactory explanation» for its decisions, then the system should not be deployed» (Fjeld *et al.* 2020, 43).
- Lack of accountability. Rieke *et al.* (2018) write, «advocates, policymakers, and technologists have begun demanding that these automated decisions be explained, justified, and audited».
- Social cohesion and filter bubbles. A cycle that augments and reinforces these patterns, putting people in filter bubbles (Pariser 2012) whereby over time they see only content from a point of view consistent with their own.
- Feedback effects. Ai prediction of an event makes the event more likely to occur.
- Indifference. For example, Emily Ackerman (2019) reports of having been in a wheelchair and blocked from exiting an intersection by a delivery robot waiting on the ramp.
- Lack of consent. Google revealed its «Project Nightingale» after being accused of secretly gathering personal health records (Griggs 2019); Google also offers a «Classroom» application and questions have been raised about its data collection practices on that platform (Singer 2017).
- Surveillance culture. Focusing on one particular identification method misconstrues the nature of the surveillance society we're in the

process of building. Ubiquitous mass surveillance is increasingly the norm» (Schneier 2020).

- Loss of power and control for example, over one's own work. «Scholarship –both the content and the structure – is reduced to data, to a raw material used to produce a product sold back to the very institutions where scholars teach and learn» (Watters 2019).
- Loss of a sense of right and wrong. Ambarish Mitra (2018) «with enough inputs, we could utilise Ai to analyse these massive data sets – a monumental, if not Herculean, task – and drive ourselves toward a better system of morality».
- Loss of ownership. «Could humans essentially be blocked out of content creation by the pace of Ai text generation and the resulting claims of copyright for every possible meaningful text combination?» (Carpenter 2020).
- Loss of responsibility. Autonomous self-organizing systems may operate independently of the intent of the designer (Ieee 2016, 196) As Bostrom and Yubkowsky (2014) write, «the local, specific behavior of the Ai may not be predictable apart from its safety,even if the programmers do everything right».
- Winner takes all. This concern has been raised by the Electronic frontier foundation (Eff) (Eckersley *et al.* 2017). They ask, «how can the data-based monopolies of some large corporations, and the "winner-takes-all" economies associated with them, be addressed? ».
- Environmental impact. Some technologies, such as blockchain, are already known to have a potentially significant impact on the environment (Hotchkiss 2019). Analytics and Ai could have a similarly detrimental impact (Meinecke 2018).
- Safety. Analytics can be hacked in ways that are difficult to detect. For example, hackers were able «to fool the neural networks that guide autonomous vehicles into misclassifying a stop sign as a merge sign» (Danzig 2020).

The scope of ethics in analytics

In the work above we've identified some areas that lie outside most traditional accounts of analytics and ethics. We found we needed to widen the taxonomy of learning analytics to include deontic analytics, in which our systems determine what ought to be done. And we have to extend our description of ethical issues in analytics to include social and cultural issues, which speak to how analytics are used and the impact they have on society.

And it is precisely in these wider accounts of analytics that our relatively narrow statements of ethical principles are lacking. It is possible to apply analytics correctly and yet still reach a conclusion that would violate our moral sense. And it is possible to use analytics correctly and still do social and cultural harm. An understanding of ethics and analytics may begin with ethical principles, but it is far from ending there.

There are some studies, such as Fjeld *et al.* (2020), that suggest that we have reached a consensus on ethics and analytics. I would argue that this is far from the case. The appearance of «consensus» is misleading. For example, in the Fjeld *et al.* (2020) survey, though 97% of the studies cite «privacy» as a principle, consensus is much smaller if we look at it in detail (*ibidem*, 21). The same if we look at the others, e.g. accountability (*ibidem*, 28).

And these are just studies strictly within the domain of Artificial intelligence. When we look outside the field (and outside the background assumptions of the technology industry) much wider conceptions of ethics appear it should be clear that our response to the different types of issues should vary according to circumstance. Where Ai fails for one reason or another, our response should be (presumably) to seek to avoid failure. Where Ai is misused by bad actors, our response ought to focus on legal and legislative remedies. Not all issues raised by critics are ethical issues of Ai, specifically, though of course ignoring them would have ethical implications.

4. The decisions we make

When we talk about whether technology can produce good things like integrity, care and trust, we shouldn't be thinking about whether an Ai or analytic system, all by itself, can produce the feelings necessary in order to produce these good kinds of ethical results. No, it is the wider system of data input designers, business models and all the rest of it that can produce care.

Now we know we can produce hate using technology, so it's not clear that we can't produce care. But the question here isn't going to be whether we can have algorithms that produce care, the question is, what would a technology that produces care look like and how would we approach designing one? That's the motivation for this section.

Our technology is not the result of one or even several decisions. It is important to consider all the different processes that go into analytics and Ai to examine the sorts of decisions that we make as we design these systems and as we use these systems to see where the care would go into it, just as we already know hate and prejudice goes into it, and even to see what we would think

counts as an ethical approach to these design and delivery decisions. We know pretty much how to use these to produce hate, whether through an absence of care, whether through a deliberate manipulation of their technology, but we haven't spent nearly as much time on all of the mechanisms that produce ethical machines.

So the purpose of this module as a whole is to be clear that, you know, by an analysis of the actual mechanisms of producing analytics in Ai, the actual workflows, the actual decisions that we take. We can see what the ethical import of our contribution to that is and how that can produce the result.

The learning context

Ai does not operate in a vacuum; it is applied in a specific context, one in this case informed by «the existing body of research knowledge about learning and teaching» (Gašević *et al.* 2015). This background defines what we are trying to do, what we are trying to measure or predict, and who is involved. These could be defined, say, with the framework provided by Greller and Drachsler (2012, 44) describing a pedagogical model containing six dimensions: competences, constraints, method, objectives (distinguishing between reflection and prediction), data, and stakeholders» (Seufert *et al.* 2019). Each of these will have a direct bearing on the decisions we make.

In Ai decision-making all stakeholders need to be involved in the decision-making, otherwise, the prioritisation of, say, institutional stakeholders may lead to undesirable outcomes (Jaschik 2016). Thus, Ai decision-making involves «encouraging or requiring that designers and users of Ai systems consult relevant stakeholder groups while developing and managing the use of Ai applications» (Fjeld *et al.* 2020, 58). Stakeholders include learners themselves, instructors, researchers and educational institutions (Khalil and Ebner 2015). Stakeholders are further divided into «data subjects» (what the analytics are *about*) and «data clients» (who manages or *uses* the analytics) (Jambekar 2017).

Objectives define what stakeholders will do with the analytics. Beyond such vaguely worded ambition as «improving learning», stakeholders may wish to increase efficiency of learning systems, improve system performance, increase transparency of the learning process, and improve student achievement (Buckingham Shum and Crick 2012). These in turn are based on a range of metrics, including actual costs, learner engagement, course completion rate, pass rate, and more. Objectives may also point to the wider benefits of learning, including economic benefits or public good (Drew 2018).

How Ai works

Contemporary Ai is based on artificial neural networks. These networks are first «trained» and then deployed into an application (Fig. 1).

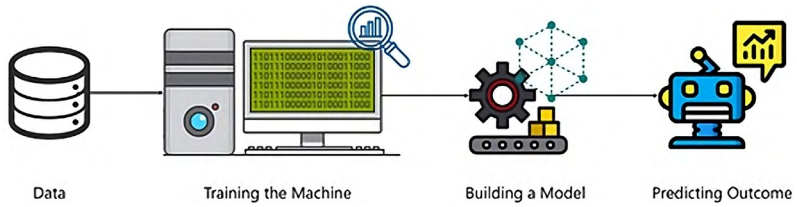

Fig. 1. How Ai works.
Source: Edureka (2023)[1].

Depending on the data available and desired output, developers may choose between:

- Supervised learning, in which inputs and desired outputs are known;
- unsupervised learning, in which the algorithm trains itself;
- reinforcement learning, in which the algorithm decides for itself, but can be corrected using a feedback mechanism.

Networks can be trained to perform a variety of tasks, including regression, feature detection, clustering, and prediction. How the software performs, and what exactly is clustered or predicted depends a *lot* on the learning style.

The learning function in a neural network defines how each individual neuron receives input from one or more other neurons, applies a statistical function to the sum of that input, and on that basis either sends or does not send a signal to other neurons in turn. Learning involves changing the variables governing that function (Banoula 2023), either directly, or through learning algorithms. These variables include:

- Threshold: determines whether or not an input value will trigger an output value.
- Bias: a negative number applied to the input that controls sensitivity.
- Activation value: a number generated from the input to the neuron; an activation function is the algorithm a neuron uses to generate its activation value from the input.
- Weights: a multiplier altering how much influence each input value has on the neuron.

[1] https://www.edureka.co/blog/artificial-intelligence-with-python/.

In a neural network, neurons are organised into layers. In «deep» neural networks, one or more layers is placed between an input layer and an output layer. As we will see below, neural networks of different topographies perform different functions. A neural network is «trained» by using data to alter connection weights. A backpropagation network, for example, will use feedback from its output to increase or decrease the weights of specific connections based on a «cost function» measuring the difference between a given and desired outcome.

The final combination of neurons and weights obtained after training is called a *model*. Ai is applied by providing fresh input data to a model and observing the output. Numerous decisions are taken in the creation of a model, most with no obvious ethical dimension at all, and yet all of which have a bearing on the outcome.

Data-related issues

Neural networks are not programmed (beyond what has been described in the previous section), they are «trained» using data. Many of the ethical issues associated with Ai are related to data, therefore, data management plays an important ethical role in Ai. Following are some of the key issues related to data in Ai (Feast 2019):

- Data integration: data is linked to other data, and can reveal more than was intended (Cohen *et al.* 2014).
- Bias: an incomplete or skewed training dataset.
- Labels: humans label training data in order to teach the model how to behave, and humans create these labels.
- Features and modelling techniques: the measurements used as inputs for machine-learning models.
- Subjectivity: there is no such thing as context-free data; data cannot manifest the kind of perfect objectivity that is sometimes imagined (Radan 2019, 16).
- Risks: stale and outdated data (Cohen *et al.* 2014), limitation on scope of data (Hand 2018).

The responsibility for identifying and mitigating these issues is distributed across the full range of stakeholders. Loshin (2002) identifies a list of parties laying a potential claim to data, including creators, consumers, packagers, funders, decoders, subjects, and more. Each of these will have different interests and objectives, and may adhere to varying ethical standards.

Organizing data

Due in part to the needs of the Ai algorithms described above, and due in part to data-related issues, data is processed in a number of ways before being applied. Following are some of the mechanisms forming important parts of this workflow.

- Data cleaning: «the process of identifying, deleting, and/or replacing inconsistent or incorrect information from the database» (Kowalewski 2020).
- Data quality: not an attribute of the source data so much as it is an output of data cleaning, consisting of such factors as accuracy, completeness, consistency, relevance, timeliness, validity and uniformity.
- Classification and naming: in supervised learning, these include operations performed as a part of labelling in Ai, that is, using human-readable signs that interpret a specific piece of data. These may be based on classifications, taxonomies, ontologies or natural kinds (van Rees 2008, 432-433), which may be created prior to data cleaning, or machine generated on previously cleaned data.

In all of this a wide range of standards could be applied and there are numerous mechanisms available to data workers. We could ask, is there a «right» way to label data? Do we all agree on what kinds of things there are in the world? It is arguable that we do not; the perspective, point of view, or «frame» we use determines how we will describe the data. Data can also be classified algorithmically, however, there are numerous classification algorithms, for example, Logistic regression, Naive bayes, K-Nearest neighbors, Decision tree, and Support vector machines, each with their own implications on data classification (Kumar 2021).

Algorithms and topologies

As mentioned above, algorithms are trained using input data. They vary according to how they are trained. Here are some examples:

- Hebbian learning: often summarized as «cells that fire together wire together», or «any two cells or systems of cells that are repeatedly active at the same time will tend to become «associated», so that activity in one facilitates activity in the other» (Hebb 1949, 70).
- Backpropagation: as described above, errors are measured and correction sent back through the network (Rumelhart *et al.* 1986, 533).
- Group method of data handling (Gmdh) develops neurons for all possible combinations of two inputs to the layer. It then continues

to choose only those neurons that supply the best possible Mean squared error (Mse) (Pandya 2005).
- Competitive learning: nodes compete for the right to respond to a subset of the input data, and in so doing become «feature detectors» for different classes of input patterns (Hassoun 1995, 3-4).
- Neuroevolution: various approaches whereby algorithms generate neural networks, parameters, topology and rules (Miikkulainen 2011).

Neural networks also vary according to how the layers and connections between neurons are organised, resulting in different network «topologies».
- Feedforward: in, for example, the perceptron (Rosenblatt 1958) and multi-layer perceptron, data flows from input to output (ie., it feeds forward) (Upadhyay 2019).
- Radial basis function network: formulated by Broomhead and Lowe (1988), these are non-linear classifiers (ie., they draw circles in data).
- Convolutional neural network (Cnn). Samples different parts of the input data, usually followed with a pooling layer, which reduces the overall size of the matrix.
- Recurrent neural networks (Rnn). The output from a neuron also becomes part of the input for that neuron (Donges 2019).
- Long short-term memory (Lstm). Process data sequentially and keep its hidden state through time, which allows it to process sequences of data (Hochreiter and Schmidhuber 1997).
- Hopfield networks. Hopfield (1982): memories could be energy minima of a neural net. «The purpose of a Hopfield net is to store 1 or more patterns and to recall the full patterns based on partial input».
- Attractor networks. «An attractor network is a type of recurrent dynamical network that evolves toward a stable pattern over time».

Why is this important? Algorithms and topologies in and of themselves have no particular ethical standing; they just are. Developers cannot just ask, «What do I need to do to fix my algorithm?» to align it with ethical values. And yet the choice of algorithm (or, rather, the many choices involved in designing or selecting one or more algorithms) will significantly impact what an Ai system can do and what it produces as output. So developers must rather ask: «How does my algorithm interact with society at large, and as it currently is, including its structural inequalities?» (Zimmerman *et al.* 2020).

Models and interpretations

The application of Ai involves the selection of pretrained models and applying them to particular cases. This selection can have a significant impact on outcomes. «So what if I chose the wrong algorithm to predict which instructors teach programming? But what if I had instead been creating a model to predict which patients should receive extra care? Then using the wrong algorithm could be a significant problem» (Young 2020).

To this point we have been representing Ai in terms of data, algorithms and topologies. It should be clear that, for the most part, a description of these will not be sufficient to offer an *explanation* of why an Ai responded as it did. An explanation is an *interpretation* of a model, and «it's one thing to detect a new cluster of words and phrases, and something else to assign an interpretation» (Liberman 2020).

On what basis do we assign an interpretation? «The use of black-box models makes it difficult for us to determine why decisions are being made» (Dhuri 2020). It's all just numbers and statistics. «This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear» (Anderson 2008). «Do numbers speak for themselves? We believe the answer is «no». Significantly, Anderson's sweeping dismissal of all other theories and disciplines is a tell: it reveals an arrogant undercurrent in many Big Data debates where other forms of analysis are too easily sidelined» (Boyd and Crawford 2012, 666).

Defining a model inherently means asking a question, and the choice of question is critical (Seufert *et al.* 2019): What problems are high priorities? How will the outcome be used? How will we respond to adverse outcomes (esp. in statistical cases) How will the outcomes be measured? Models are «trained», yes, but the training is the result of extensive programming: Are rigorous programming standards used? Is the program open source?

Perhaps we can ask the question slightly differently. We can ask, how does an Ai see the world? What interpretation does it place on all that data. Interpretation is a type of skill. Jonna Vance (2021): perceptual expertise as «an enhanced capacity for perceptual recognition or discrimination with respect to some feature or category». For example, «one can be a perceptually expert recognizer or discriminator of bird species, cars, or tumours depicted in X-rays». We can think of an Ai, then, as an expert recognizer. Is perceptual expertise always virtuous? Vance writes, «there is no guarantee that perceptual expertise will have a net positive contribution to the proportion of true beliefs or knowledge». So, «are privileged epistemic agents subject to different epistemic obligations than marginalised or oppressed epistemic agents are?».

Testing, application and evaluation

Software testing is a large field but in general the objective is to determine that the program will produce the output you expect it will, given the circumstances. Each state of the process is tested: from the original request (to ensure that the correct data is being collected, the request to the Ai is correct, the request is properly sent, etc.) to data testing (for validity, reliability, variety, consistency, etc., as mentioned above), the application (for security, performance, usability and failover). Ultimately, though, any given model must be tested in real-world applications «with adequate protections and precautions» (Cohen *et al.* 2014).

The application of an Ai application, like any process, device or approach, requires a period of introduction and assimilation among the population intended to benefit from its use. The recent pandemic shows the importance of this; the rejection of vaccines as a treatment shows that the knowledge behind their development was not sufficiently integrated into people's values and belief systems. This is the subject of «knowledge translation», a term coined by the Canadian institutes of health research (Cihr) in 2000 to describe «the exchange, synthesis and ethically-sound application of knowledge». A more recent characterization, «knowledge mobilisation», describes the «activities relating to the production and use of research results» as «an umbrella term encompassing a wide range of activities relating to the production and use of research results, including knowledge synthesis, dissemination, transfer, exchange, and co-creation or co-production by researchers and knowledge users» (Wilsdon 2015).

The application of Ai is not without its decision points. James Clay (2020) writes, «we must not forget the human element of data and analytics. It's not enough to deliver accurate analysis, predictions, and visualisations. Staff and students in universities and colleges need to be data literate to enable them to understand and act on that data. Appropriate and effective interventions will only be possible if staff and students are able to understand what is being presented to them and know what and how they could act as a result».

Finally, evaluation. We don't mean testing to determine whether the Ai or analytics application works, but rather, whether the use of Ai produces satisfactory results. But what counts as a satisfactory result is very much in the eyes of the beholder. Evaluation in this sense takes into account a much wider context. Factors that have nothing to do with the design and development of Ai come into play. We evaluate learning analytics, for example, not simply based on the question of whether they are «improving learning», but on whether

they (say) «promote autonomy», «support the UN sustainable development goals», «promote organisational efficiency», or «enhance shareholder value».

5. Ethical codes

A *common* response to the ethical issues raised by Ai has been to develop an ethical code. These codes characteristically identify a set of ethical issues as problematic, and identify a set of responses that address those issues. Adherence to the ethical code then comes to define ethical conduct with respect to Ai (or, in what is possibly a shorthand expression, ethical Ai).

Principles

One of the major characteristics of these ethical codes is that there is a set of ethical principles that proponents hold in *common*, or ought to be *generally* applied across the Ai domain. This claim is often implicit, though there are numerous occasions when it is stated explicitly. It is worth noting (but need not be argued at this point) that the ethical values felt to be *common* and prevalent are historically liberal democratic values such as human rights and freedoms, non-maleficence, justice and fairness.

As a case in point, consider the analysis offered by Floridi and Cowls (2019) (Fig. 2). «Our analysis finds a high degree of overlap among the sets of principles we analyze», they write, arguing that they can «identify an overarching framework consisting of five core principles for ethical Ai» as illustrated in figure one.

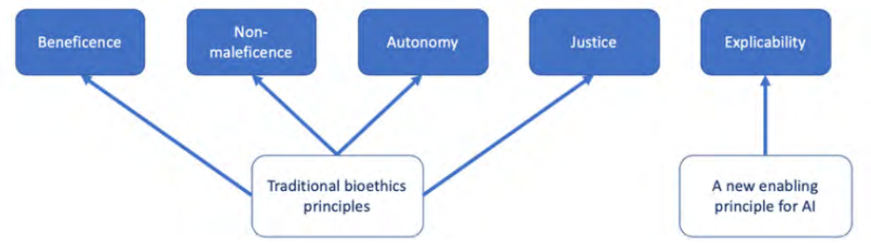

Fig. 2. An ethical framework of the five overarching principles for Ai which emerged from the analysis
Source: Floridi and Cowls (2019)[2].

[2] https://www.researchgate.net/figure/An-ethical-framework-of-the-five-overarching-principles-for-AI-which-emerged-from-the_fig2_355882962.

The sets of principles analysed by Floridi *et al.* (2018), however, are remarkable for the homogeneity of the sets of authors. They include «global thought leaders», «attendees of the high-level Asilomar conference», members of the European Commission, members of the British House of Lords, the Ieee, and «stakeholders». They are unified by their interest in Ai, legislation and policy, and therefore, their perceived need for a «code of ethics» to govern the use of Ai.

Their focus is understandable. It draws from a tradition of articulating codes of ethics for professional practice in general. And these, too, are widely help to express a *common* set of values. For example, consider Metcalf (2014): «There are several principles that can be found at the core of contemporary ethics codes across many domains:

- respect for persons (autonomy, privacy, informed consent),
- balancing of risk to individuals with benefit to society,
- careful selection of participants,
- independent review of research proposals,
- self-regulating communities of professionals,
- funding dependent on adherence to ethical standards».

It ought to be recognized that the Metcalf (2014) set of principles is quite distinct from the Floridi and Cowls set of principles. A *broader* analysis of various sets of principles across various professional domains shows *no* commonality across codes and disciplines. A *deeper* analysis shows that even in areas where there appears to be broad consensus, there is significant disagreement in the details.

We see this, for example, in the set of codes analysed by Fjeld *et al.* (2021) in which again we see an assertion that there are principles held in *common* across them all. But while there is is broad agreement (of about 70 % of documents) around «transparency and explainability», this agreement breaks down when pressed for details, yielding a general non-consensus: «28% open source data and algorithms; 11% right to information; 25% notification when Interacting with Ai; 3% open government procurement; 19% notification when Ai makes a decision about an individual; 17% regular reporting»[3] (Fjeld *et al.* 2021, 41). And this is *still* with a relatively homogeneous set of contributors.

Randall Cunningham's (2008) XKCD expresses the situation quite well[4]: The same is true of ethical codes.

[3] See https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf.

[4] https://xkcd.com/927/.

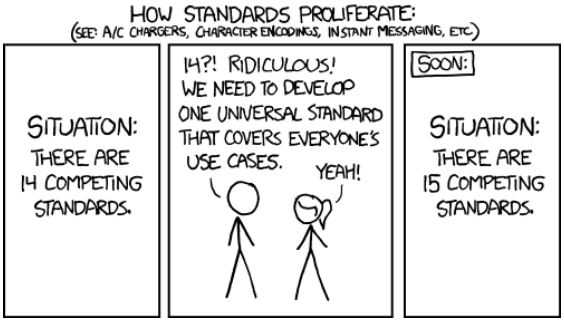

Fig. 3 How Standards Proliferate.
Source: Cunningham (2008).

Values

What justifies the specific recommendations made in ethical codes? We might say that «An individual's professional obligations are derived from the profession and its code, tradition, society's expectations, contracts, laws, and rules of ordinary morality» (Weil 2008), but this raises as many questions as it answers. When we analyse the documentation of ethical codes we find a wide range of underlying sets of values or justifications.

For example, many codes reference universality as a justification for moral and ethical principles. For example, the Universal declaration of ethical principles for psychologists asserts, «the Universal declaration describes those ethical principles that are based on shared human values» (Iupsys 2008). Related to universality, but not the same, is the doctrine of fundamental rights. The High-level expert group on artificial intelligence (Ai Hleg), for example, cites four ethical principles, «rooted in fundamental rights, which must be respected in order to ensure that Ai systems are developed, deployed and used in a trustworthy manner» (Ai Hleg 2019).

But more pragmatic considerations may also apply. Discussions of Ai ethics often appeal to a balance of risks and rewards. For example, the Ai-4People declaration states «an ethical framework for Ai must be designed to maximise these opportunities and minimise the related risks» (Floridi *et al.* 2018, 7). This is a broadly consequentialist approach and therefore results in a different calculation in each application. It also requires an understanding of what the consequences actually are. Or perhaps ethics isn't really a case of balancing competing interests, but rather, based on a maximisation of benefits. The Information and Privacy Commissioner in Ontario, for example (Cavoukian 2013) asserts that «a positive-sum approach to designing a regulatory fra-

mework governing state surveillance can avoid false dichotomies and unnecessary trade-offs».

These differences are reflective of the fact that over the 3,000 or so years humans have discussed such matters, no agreement on the basis, nature or principle of ethics has been found.

6. Approaches to ethics

Numerous approaches to ethics exist, which may be classified roughly as follows.

Virtue and character

From this perspective, ethics is in the first instance the study of virtue in a person. This may reference virtue as a Platonic ideal, or virtue may be revealed by a person's actions how they conduct themselves in society. The nature of virtue is usually characterised as consisting of virtuous traits or dispositions such as honesty, frugality, piety, humility, caring, courage, and the like. Aristotle lists twelve such virtues. In Confucianism, we could identify such virtues as benevolence, righteousness, propriety, wisdom and fidelity (Wahing 2021, 3). Virtue, though, is not defined by these traits, it is defined as the character or essence that unites them. A virtuous person is a person who acts virtuously, while virtue itself is the moral characteristic such a person needs to live virtuously. Thus, for example, the stoic person, as defined by Stoicism, can be described as virtuous.

The vagueness of virtue theory is at once a strength but also a fundamental flaw. Friedrich Nietzsche, for example, asks what sort of virtue a «superman» (*ubermensch*) might have (e.g. Nietzsche 1999, ch. 4). Such a man would face no limitations on conduct. Would a «superman» really dedicate his life to fighting crime? Apply the principle to contemporary politics and the world of Donald Trump, in which lying is a virtue, or in which stealing is a virtue, if you can get away with it. In a Donald Trump world, it is virtuous to take what you can. Politics isn't the art of negotiation and compromise, it is the art of power and leverage. Viewed from a certain perspective, Donald Trump is the most virtuous of us because he understands this and acts on it. It calls to mind the Wiccan ethos, «do what thou wilt, is the whole of the law».

Duty

For Immanuel Kant and those who followed, ethics is based on duty. Kant's system of ethics is based on two major principles: first, the «categorical imperative» is the maxim that ethical principles ought to be universally applicable, that is, we should act according to the maxim that we could make it the case that this type of act would become a universal law of nature. We ought to ask, «what if everybody did that?» Second, the idea that we should treat people as valuable in and of themselves. In other words, we should treat people as «ends», not «means». People are not objects to be used, but as rational, sentient, and most importantly, *ethical* beings, have inherent worth and standing.

These principles combined create on us a *duty* to act in such a way that supports the well-being of society and the individuals in society. It's intuitively appealing; «with great power comes great responsibility». It's not enough to simply *be* virtuous, we must *act* according to our virtue. However: what counts as universalizable? *How*, exactly, does one treat another person as a means or an end? Kantian ethics is often appealed to as a defence of naturalism, that is, the idea that we should avoid unnatural acts. But arguably, anything the human body can do is natural, which leaves slim grounds for objecting to something on the basis that it's not natural. Indeed, characterised correctly, *any* act can be thought of as universalizable. «Any person sitting in Stephen's office at 4:00 p.m. September 9, 2023, may take whatever they want».

Consequentialism

«Consequentialism» is a catch-all phrase for a host of ethical theories ranging from «do no harm» to «the ends justify the means». A «consequence» is the result, effect or importance of an action or condition. Thus consequentialism is the idea that an ethical act, or ethical principle, is evaluated according to its consequences.

Different consequentialist theories vary in their account of what count as desirable consequences. For individuals, happiness is desirable, which may be described as the presence of pleasure and the absence of pain. The emphasis here may vary; hedonists seek physical pleasure exclusively, for example, or we might, as John Stuart Mill (1879, chapter 2) recommends, seek the «high pleasures» of knowledge and enlightenment. An Epicurean, meanwhile, might define pleasure as the absence of suffering, teaching that all humans should seek to attain the state of *ataraxia*, meaning untroubledness. Avoiding pain may be, as a Buddhist might argue, a matter of attitude; we experience unbearable suffering because of the tight grip of our grasping itself, it is in wanting

permanence in a world that is forever changing that results in suffering. Consequences may be individual or collective; Mozi (1929), describing a social consequentialism (Harris 2017), wrote, «the sage-kings of old appreciated what Heaven and the spirits desire and avoided what they abominate, in order to increase benefits and to avoid calamities in the world».

Similarly, different societies have different understandings of desirable consequences. In the United States, it is «life, liberty and the pursuit of happiness», while the French may seek «liberty, equality, fraternity», while a Canadian may value «peace, order and good government».

Social contract

This concept of the social contract is usually represented as a form of political organisation, but here, the core idea of a social contract is the idea that ethics results from an agreement within a community. The major components of social contract ethics are: first, the process or method by which agreement is reached; second, the determination of the contents of the resulting agreement; and third, the motivation is to abide by the agreement.

Different approaches to social contract ethics described these three components differently. For example, in *A Theory of Justice*, John Rawls (1999, 104) describes a hypothetical «original position» in which participants, screened by a «veil of ignorance» (*ibidem*, 11), negotiate the social contract, the result of which, he argues, is a theory of justice as fairness (Rawls 1999), and specifically, that each person has the same claim to equal basic liberties such that, first, they create conditions of fair equality of opportunity, and second, they are to be to the greatest benefit of the least-advantaged members of society (Rawls 2002, 42–43).

What would motivate someone to accept a social contract? The alternative might be that much worse; and we would live lives that are «solitary, poore, nasty, brutish and short», as suggested by Hobbes (1994, XIII.9). Or perhaps we might recognize such rights as innate, as does Rousseau in saying «Man is born free» (2004, 1.1). Or perhaps it is the recognition that we are inherently social, and thus require the conditions to support that sociality? The motivations may be as many as the variety of resulting social contracts that have appeared through history, whether as religious creeds, political constitutions, manifestos, or compacts.

Metaethics

The four great theoretical approaches to ethics – virtue, duty, utility, agreement – do not exhaust the domain of moral discourse, but the disputes with and among them make clear the need for *broader* discussion of the foundation of any possibility of ethics at all. What makes an ethical statement true?

Various possibilities have been suggested. In our review of Kant, we came across the idea of universality, that is, the idea that an ethical principle must be universally applicable, akin to a law of nature. Alternatively, others have suggested that nature itself forms the foundation of ethics. What, though, is more natural than one's own body, one's senses and feelings? Perhaps morality is the highest expression of our cognitive capacities, of reason and enlightenment. Or perhaps it is non-cognitive, more like sensations. Perhaps ethics is like science, where we discover right and wrong, or instead, perhaps right and wrong are invented to serve some other purpose. Are ethics even something we use when we make a decision, or something we come up with after the fact?

The end of ethics

It is at this juncture we must part with the idea of traditional ethics. We have ethics – that much is evident – but we don't know *why* we have ethics. In this way, human cognition is similar to machine learning, in the sense that it is inscrutable. Of these, David Weinberger (2021b) writes, «In the cry "We don't know how machine learning works!" we hear that these models do indeed work... Our encounter with MLMs doesn't deny that there are generalisations, laws or principles. It denies that they are sufficient for understanding what happens in a universe as complex as ours».

Simple principles are not sufficient to address complex thoughts, ideas or problems. Jones (2011) writes, we «must deal with interdependent problems, navigating nonlinear and often unpredictable change processes, involving a diverse range of stakeholders». First, «the capacities to tackle complex problems are often distributed among actors». Second, «complex problems are difficult to predict: many social, political and economic problems are not amenable to detailed forecasting». And third, «complex problems often involve conflicting goals».

Ethics are not abstract. Though we can derive universal principles based on abstract values or calculations, we shouldn't. «The only real weapon against the fearful vision of a cold siber-Cyberia is joy. Appreciation of the space gives the surfer his bearings and balance in Cyberia» (Rushkoff 1994, 180-182).

7. The duty of care

It is often assumed that ethics is a matter of argument; if we provide the right reasons, people will see what the ethical approach is to any given dilemma, and follow that approach. The argumentative approach is based on what Robert Nozick (1981, 4) called «coercive philosophy»: «arguments are powerful, and best when they are knockdown, arguments force you to a conclusion».

Often, however, the opposite is the fact. The argument doesn't force anyone to a conclusion. If anything, it forces people who disagree to retrench. «The moral principles people endorse relate to their life experiences, family roles, and position in society. For instance, exposure to war or abusive/dysfunctional family relations impedes moral reasoning. More *generally*, many studies have shown that the moral judgments people make depend on their age, gender, parental status, education, multicultural experiences, war experiences, family experiences, or religious status» (Ellmers *et al.* 2019, 351-352).

People are often not swayed by debate; if anything, debating with them legitimises their position, at least from their perspective. Rather, it could be argued that for many people, moral reasoning and moral judgements are the subject of relationships with distinct and particular individuals to whom we owe some responsibility of care, for example, the caring of a mother for a child (Weinberger 2021a).

This concept forms the basis of what has come to be called «the duty of care».

Care as a legal concept

The origin of a duty of care as a legal concept is arguably the case heard by Lord Atkin of a decomposing snail in a bottle of ginger beer. The question before the courts was whether the people who created and bottled the ginger beer have any responsibility toward the person who eventually consumed it, and especially for the ill effects they suffered by the person. The court decided in a split decision that they did. There is a legal obligation, which is imposed on an individual, requiring adherence to a standard of reasonable care while performing any acts that could foreseeably harm others.

Of course this example is quite different from the concept of care with mother and child. But though we need to keep these two distinct senses in mind, there's a lot of overlap between them. The relation between the carer and the cared for is asymmetrical, and this asymmetry creates a responsibility of one toward the other. Sometimes this responsibility is codified in law, other times it appears to be more of a biological imperative.

In the legal sense, the idea is that there's a responsibility or legal obligation of a person to avoid acts or omissions that could reasonably be expected to cause harm to others. This duty becomes more specific and more urgent when a special relationship exists, as in, for example, a legal or medical professional, or a teacher.

As a standard of *ethics*, however, a definition of a duty of care as a depersonalised functional, definition of what counts as being responsible in these professions is obviously inadequate. From the perspective of new technologies with new affordances, like Artificial intelligence and analytics, thinking of the duty of care only as a legal concept will have some important limitations, especially as these legal principles abstract away from the very real and very important relations that teachers have with students that students have with each other and that all of us have with each other within society.

The ethical concept of care

So what is the ethical concept of care? Wikipedia states that the ethics of care is a normative ethical theory that holds that moral action centres on interpersonal relationships and care or benevolence as a virtue. Now that definition clearly is coming out of the terminology and the taxonomy of the other ethical theories, and not from within the perspective of care itself. But what's important here is that it draws out the concept of impersonal interpersonal relationships as core to care. «Care, caring, carefulness and being cared for are embedded, multidimensional, empowering, fraught and temporally multifarious, rather than unitary and static» (Motta and Bennett 2020, 640).

The concept of care itself predates, of course, the ethics of care. It's represented as a process, a way of relating to someone that involves development, just as for example in the same way that friendship can only emerge in time through mutual trust and a deepening and qualitative transformation of the relationship. Milton airoff suggests that to care for another person is in the most significant sense to help him grow and actualize himself. The idea here is of mutual growth, of mutual development, of the element of relationship that exists in care, even with even before it becomes a feminist philosophy. Caring has a way of ordering his other values and activities around it.

The ethical concept of care has its origins in feminist epistemology and feminist ethics, and it's important not to take it out of context. A white North American man such as myself may be tempted to present it as another theory and to present the authors as if they were each presenting each a distinct and independent theory and arguing with each other about what that should be. But this would be, I suggest, a misreading. That is not to say that the authors

don't have their own perspectives. Absolutely, they do, and these perspectives really shine through clearly. But it would be a more appropriate treatment to think of the ethics of care as an undifferentiated whole and of these authors as highlighting different aspects of it rather than as arguing among each other about what it should be.

Care and relationships

The core assumption of care as a theory is that persons are understood to have varying degrees of dependence and interdependence on one another. Other individuals are affected by the consequences of one's choices. It's not sufficient to treat the vulnerable person as a passive recipient of care. Rather, care is based on relationships. Care is based on a mutual exchange between the career and the cared for. What's important is that caring is more about the concreteness of the relationship and the concreteness of the interests of those involved. These are things that you cannot simply describe as abstracts.

According to Carol Gilligan, the concept of care, rather than being deduced from, say, the nature of humanity or virtue, is an ethic grounded in voice and relationships and in the importance of everyone having a voice, being listened to carefully in their own right and on their own terms, and heard with respect, and additionally directs attention to the need for responsiveness and relationships.

So, care is based on the actual consideration in the actual voice of the people who are impacted by whatever decisions are made, and this would include the person being cared for, but it's not necessarily limited to the person being cared for. It's based not just on what we think. We're not doing mental exercises here. It's based on people's expressed needs. It's based on actually having conversations with the person being cared for, and possibly other people in the community, so that they can express their needs, because they are the ones that are in a unique position to say what those needs are.

According to Gilligan, an ethics of care starts from the premise that as humans, we are inherently relational. The human condition is one of connectedness or interdependence. So rather than being based on a rational argumentative calculation or sort of logic, morality is grounded, she says and a psychological logic reflecting the ways in which we experience ourselves in relation to others. Just so, care is represented as a feminist theory at least in part because proponents such as Gilligan believe women are more likely to make more decisions based on issues of care, inclusion, and personal connection, rather than on a more abstract and distant and distant notion of justice, based on their own direct experience of childbirth and connection with a child.

Care becomes a duty only in that the fact of the relation creates the urgency on the part of the other person to respond. It's not about being a better person, or a sort of reasoning about the moral status of the other person. So it's more than just a semantical thing. It is also an emotional or motivational thing. The ethics is you responding to the motivational or urgent sort of factor of that relationship. It doesn't even make sense to talk about a social contract between the carer and the cared for. It doesn't make sense to talk about a utilitarian calculation.

Pedagogy of care

Applying an ethics of care to education not only allows us to develop a unique «pedagogy of care», it also reshapes the role of analytics in learning. While traditionally, learning analytics has been deployed to assist instructors in shaping learning experiences, in a pedagogy of care the idea is to «ensure they (students) have enough knowledge and tools to make a well-informed choice, and know how much leeway they have to design their own path» (Bali 2020). «"Care" pedagogically expresses itself as recognition of the complex creative energies, desires and experiences of students as a place of knowing-possibility» (Motta and Bennett 2018, 637).

Moral sentiment

Feminist ethical theory deals a blow to the exclusively rational systems of thought that may have as their grounding and inherent disregard for the inherently personal, and sometimes, gender-based nature of knowledge construction (Craig Dunn and Brian Burton writing on Encyclopedia Britannica). What this means is that it moves ethical knowledge from the realm of explicit knowledge to what Polanyi would describe it as the realm of tacit knowledge.

Our ethical actions are not deductions. They're not inferences. So, what are they? We might say they're like what Jack Marshall calls «ethics alarms», the feelings in your gut. «Emotions and their embodiments thus become central to the construction of knowledge and knowing-subjects, and in particular knowledges about education and pedagogies of inclusion/exclusion, justice/injustice» (Motta and Bennett 2018, 634). It's more like a sensation than a type of cognition. We can call this a moral sense, or as David Hume would describe it, a moral sentiment.

To expand on this as a story about moral sense, we can draw from Elizabeth Radcliffe, who suggests that moral distinctions depend on our experience, sentiments or feelings. This is not a theory of innateness or natural morality,

nor are we saying that we have an inborn awareness of what morality is. It's not a sort of Cartesian certainty like «I think, therefore I am, therefore, I am moral». The idea that we can learn ethics, but we learn ethics in such a way that we feel or experience a moral sense rather than fully formed general principles.

It's important here to be clear that this is different from moral intuition. Speaking of the ethics of care, people may equate what they're talking about with intuition, as in «women's intuition», for example. But that's not what's intended. It's more like a sentiment or a feeling. It's more equivalent to your sense of balance; you get this feeling when you're off balance, and you wouldn't describe that as an intuition. It is often experienced at a subsymbolic (or «ineffable») level – ethics is not (*contra* Kant) not a matter of rationality but rather one of sympathy. How we react in a particular case depends on our ethical background and is the result of multiple simultaneous factors, not large-print key statements

How can you learn a sense? Think about training your taste buds. A sommelier, for example, a taster of wine, will over time, learn how to distinguish different types of wines. Similarly, someone who is a coffee aficionado will learn to distinguish different types of coffee. Moral sensations are like that, a sort of affective feeling that we might have, not an emotion, the sense of anger, or fear, or hope or desire, but actually a much more gentle and subtle kind of feeling. Similarly, it is arguable that such feelings «have shaped the cultural evolution of norms. For example, groups share autonomy norms in part because these norms resonated with moral feelings of respect and were therefore favoured in cultural transmission».

8. Ethical practices in learning analytics

As this is the ultimate section of this paper, it is important to take stock of what has been learned this far:

- we found that the application of Ai in learning (aka «learning analytics») has the potential to produce numerous benefits;
- in addition, we found that a wide range of ethical issues has also been raised;
- the development, application and testing of Ai depends on numerous decisions, all of which will have an ethical impact;
- through an examination of ethical codes, we determined that there is no consensus or *common* system of ethics describing what we would accept;

- indeed, a survey of ethical theory in general suggests that people have widely diverging ethical beliefs;
- in the duty of care we found an alternative approach where our ethical judgements in concrete circumstances are based on our sense of ethics.

In a nutshell, then: there is no particular point in Ai where ethics can be «applied», there is no individual person or set of people who can be held «responsible» for ethical outcomes, and there is no particular description of what constitutes «ethical Ai». The idea that we could somehow intervene in the process in order to produce «ethical» or even «explainable» Aiu appears to be misguided. But in specific instances, we *can* talk about the ethics of Ai, as we appeal to our own moral sense to decide what's right and what's wrong in any particular application of it.

In practice, we approach the ethics of *everything* differently depending on how we regard the potential consequences:

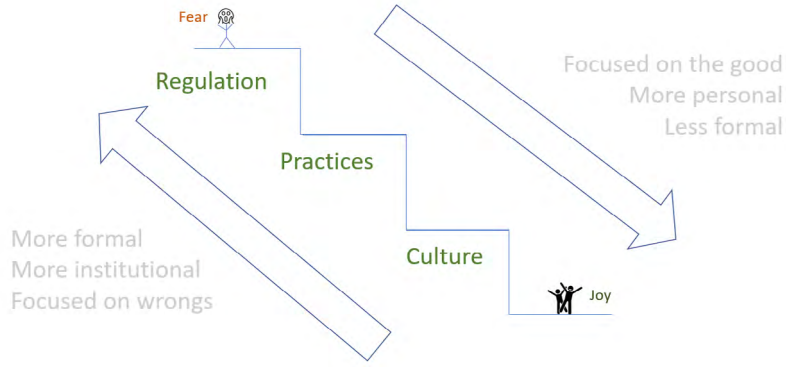

Fig. 4

The increasingly complex nature of Artificial intelligence lends to it a sense of mystery and opacity, writes Quito Tsui: «Echoing the genre of myth, emerging technologies and discussions around them are today infused with a sense of incomprehensibility, or a fundamental inability to understand or audit the "decision-making" of predictive tools, and an inviolable sense that these technologies defy our mortal ethical frameworks». To «tame» this mystification, Tsui argues that «it is vital to reorient the direction of Ai» and «we should be explicit about the direction we want Ai to face, the direction in which it should serve. By focusing the gaze of Ai on responding to the needs of the global majority, and mobilising Ai for those purposes in a directed manner, we can reign in its mystical status».

Regulation and governance

Regulation is at once narrower in scope and more formal than other approaches to ethical practice. A regulatory approach tends to focus on the most pertinent ethical issues raised by Ai such as fairness, transparency and privacy. Regulators are concerned with explainability and interpretability: for example, the idea of a «right to explanation» of algorithmic decisions. It often depends on a system of ethical auditing consisting of mechanisms that examine the inputs and outputs of algorithms for bias and harms (Cath 2018).

Above, in the section on models and interpretations, we discussed the issues inherent in explaining Ai. From the perspective of governance, the scope for «explainability» is also narrower in scope and based on more formal criteria. The Ico and The Alan Turing Institute, for example, identify six main varieties of explainability (Schildkraut 2021, 9):

- Rationale: reasons behind a decision;
- Responsibility: who made the Ai system and how to obtain a human review;
- Data: what data went into the model and how was data used;
- Fairness: how we know the Ai is unbiased and individuals are treated equitably;
- Safety and performance: how accuracy, reliability, security and robustness ensured;
- Impact explanation: how are effects and decisions monitored.

These provide explanations not in a scientific sense, but in a legal and political sense, seeking specific evidence of actions undertaken.

Often, a risk-based approach is taken. The European Ai regulation is one such example. Some Ai practices are banned (Title II), including technology that is subliminal, exploits vulnerabilities, social score, real-time biometrics (subject to conditions). Other jurisdictions, again citing the risk of abuse, have imposed facial recognition bans. The European Gdpr protects data rights, intellectual property law is concerned with defining «authorship» and «ownership» of Ai-generated content (which, currently, is none, as several ruling show Ai-generated content cannot be copyright). Regulations also address civil wrongs (torts) associated with Ai, such as manufacturing and design defects or failure to warn of risks (Hodgett *et al.* 2023).

Of particular interest is governance based in human rights. These would «not just set forth standards for how to "do no harm" or "be ethical", but it would help hold companies accountable for those standards» (Biddle and Zhang 2020) For example, Halbertal (2015) defines three categories of human rights violations: «being put in a state of helplessness, insignificance; losing

autonomy over your own representation; treating an individual as exchangeable and merely a means to an end; and making an individual superfluous, unacknowledging one's contribution, aspiration, and potential». Designing for human rights in Ai (Aizenberg and van den Hoven 2020). However, while a human rights framework is popular in some areas, the definition of these rights may be too Western and too individualistic, in addition to being too narrow in scope and too abstract to form the basis of sound Ai governance.

So far as an *ethics* of Ai is concerned, the regulation approach is inherently limited. What is legal is often not ethical. Regulation becomes subject to Goodhart's Law, the idea that any metric ceases to be a valid metric the moment it becomes a target for optimization (and gets «metric-hacked»). An oft-cited example of this law in action is the creation of a bounty on cobras, intended to reduce the cobra population. The bounty worked well at first, however, people began raising cobras in order to collect the bounty, resulting in a much larger population of cobras (Treviranus 2018, 32). Similarly, we see the beginnings of an industry developing around negotiating gaps and shortfalls and bypassing Ai regulations, without regard to Ai ethics (Rodrigues 2020).

Ethical practices frameworks

«Should we embed machines with ethics that we know to be good (ethics-by-design), or should we repose our faith in an ethics-making method that leads to ethics emerging by agreement in a society of machines?» (Nallur and Collier 2019, 534). As the authors suggest, the former sounds like the best option, but in a world with a multiplicity not only of ethical perspectives, but of interacting Ai systems, it's impossible to achieve. So, better to consider how Ais expressing the various perspectives can come to agreement on ethical questions. This suggests the use of ethical practices or frameworks, rather than specific codes or principles.

It is a truism that systems of rules and principles are compromised the moment they are applied in practice. A *common* expression, «the fog of war», captures the uncertainty that results when complex real-life situations are encountered. Thus many approaches supporting professional conduct focus on *common* practices rather than principles. The idea is that while the actual outcome and best decision cannot be predicted, following a standard will lead to an optimal outcome in the given situation (Courtney *et al.* 2013).

Some frameworks for ethical practices, for example, might include a management framework for ethics, a data governance framework, an It governance framework, or a human rights framework. Each will address specific aspects of Ai systems and practices.

A management framework for ethics, for example, might suggest a series of steps to be taken. The Markkula center for applied ethics at Santa Clara university offers a typical example (Kwan *et al.* 2021) whereby one would recognize an ethical issue, get the facts and evaluate alternative actions, then make a decision and reflect on the outcome. Similarly, the Sheila framework, an initiative using the Rapid outcome mapping approach (Roma), which was, according to proponents, specifically designed for policy-making derived from scientific evidence (Young *et al.* 2014), recommends identifying the problem, develop a strategy, and developing a monitoring and learning plan.

While these frameworks are reactive, responding to actual or potential problems, the Canadian privacy commissioner (2020) proposed that an appropriate law for Ai that outlines a prior regulatory framework that would allow personal information to be used for new purposes within a rights based framework while creating provisions specific to automated decision-making that would require businesses to demonstrate accountability

Alternatively, something like the Digital catapault acts as more of a checklist. Developed to help Ai companies design and deploy ethical Ai products, it consists of seven concepts: (deBruijn *et al.* 2020):

- be clear about the benefits of your product or service;
- know and manage your risks;
- use data responsibly;
- be worthy of trust;
- promote diversity, equality and inclusion;
- be open and understandable in communications;
- consider your business model.

Ethical frameworks fall short of being regulatory, though they are often intended to support the creation of regulations. They define what might be thought of as professional or institutional practice. They employ varying mechanisms for governing practice, for example, decision trees, checklists, frameworks and processes. None of these defines ethical practice, but all of them are useful as aids to either promote what is good or prevent what is wrong.

Ethical communities

From the perspective of an ethics of Ai, governance and ethical frameworks are not sufficient. They are designed for organisations, not wider society. They depend on agreement and shared presumptions. And most of all, they are not actually based in ethics. Violating these sanctions will be considered to be crossing a legal or institutional regulation, but not inherently a breach of ethics (except where «ethics» is strictly defined as a legal concept).

Compliance and enforcement therefore become constant issues of concern, and are addressed, unsuccessfully, through such mechanisms as business risk management, training and development programs, and high-level statements of principles and values (Blackman 2020).

Sometimes the motivation for these individual acts are defined in terms of citizenship, though conceptions of citizenship vary widely, and often appears to be based in a particular approach to ethics, ranging from Lockean liberalism («pursue the good life and be free from unreasonable government interference») to republicanism (widespread participation of citizenry as a duty toward the community. Mossberger *et al.* 2007, 6-7) to the so-called «digital citizen» («embrace rationalism, revere civil liberties and free-market economics». (Katz 1997). But the concept of digital citizenship often does not touch people where they live day-to-day. «Consider how infrequently many adults consider how the work they do, the things they buy, or the food they eat affects national or global citizenship. This is all big picture thinking that is, somehow, easy to miss» (TeachThought 2019).

As has come to be recognized even in business circles, matters of ethics are ultimately matters of culture rather than governance or frameworks («culture eats strategy»), or in science, of the «legitimacy» of a certain practice or approach (Schintler *et al.* 2023, 9). The question of *ethics* in Ai therefore becomes a question of the culture of people involved in the development, delivery and use of Ai. We must ask here, how can the ethical culture around Ai be addressed?

What the concept of digital citizenship strives to achieve is some concept of an ethical community. This idea of «citizenship» is too narrow a notion, derived as it is from top-down principles of global ethics. Jones and Mitchel (2016, 2063) argue for a «narrower focus on (1) respectful behaviour online and (2) online civic engagement». They write, «both online respect and civic engagement were negatively related to online harassment perpetration and positively related to helpful bystander behaviours, after controlling for other variables».

Again, though, it is helpful to look at what will constitute ethical practices from the perspective of the Ai, describing the concept of ethical Ai as something like describing Ai as an ethical member of the community. And while we have a tendency to describe this in terms of virtues and duties, ultimately, it comes down to how well Ai and humans can interact with each other. And while we may be inclined to think of Ai as interacting with us at the output end, it is important to recognize that we also interact at the input end, the data used to train an Ai. Just like a human, an Ai learns from, and emulates, the culture with which it finds itself.

An alternative model of ethical community may be grounded in a concept of participation. Henry Jenkins describes this in relation to digital literacy. Citing «The breakdown of traditional forms of professional training and socialisation that might prepare young people for their increasingly public roles as media makers and community participants», he recommends training in a range of new skills, including distributed cognition, collective intelligence, networking and negotiation (Jenkins 2006, 3). Such an approach could include interactions with Ai as a *part* of that community and would both inform Ai design guidelines (Amershi *et al.* 2019, 3) as well as the skills needed to interact productively with an Ai, these ranging from Ai literacy (Long and Magerko 2020, 1) to prompt engineering (Wang *et al.* 2023, 1). There is a large literature on individual agency, community participation, and ethics, which is out of scope here, but which should be noted.

Ethics and culture

As mentioned above, in many ways, fostering an ethical Ai depends on an ethical culture. But here we ask, what is culture, much less, an ethical culture? As always, varying perspectives exist. We can think of culture as «the characteristics and knowledge of a particular group of people, encompassing language, religion, cuisine, social habits, music and arts» (Pappas and McKelvie 2022). Or as «shared patterns of behaviours and interactions, cognitive constructs and understanding that are learned by socialisation» (Damen 1987, 367). Or even as «the collective programming of the mind which distinguishes the members of one category of people from another» (Hosstede 1984, 51).

Culture – whatever it is – is grown or constructed in a relational space of communications, interactions, behaviours and traditions. It may be defined statically, as above, or dynamically. For example, Thomas and Seely Brown (2011, 37) draw the analogy of the culture that is found in a petri dish where «culture» is understood as the growth produced in that environment. Either way, culture is something that is grown and developed through interaction and participation in a community.

Culture – like the ethics of care – is grounded not in principles or generalisations, but in individual and collective acts. These, argue people like Saidiya Hartman, are based in forms of activism that focus on what we might call «quiet acts of caring» found in a «close narration» (Haffey 2023) rather than amplification of a message or platform, a movement «driven not by uplift or the struggle for recognition or citizenship, but by the vision of a world that would guarantee to every human being free access to earth and full enjoyment of the necessities of life, according to individual desires, tastes, and inclinations».

This can be characterised as a form of «connective action», as described by Bennet and Segerberg. Key ideas include «the idea of «platform inclination» as an alternative to «standing up to» or «against» something, the idea of producing hope rather than «looking for hope in the sky,» the distinction between «a performance of care» as opposed to «doing the work of care,» and connective action and the practice of «communicative labour» «at the point of organising rather than more visible forms of resistance» (Singh 2020). Just as «participatory culture shifts the focus of literacy from individual expression to community involvement» (Jenkins *et al.* 2009, 6) it in the same way reshapes the focus of ethics. Such a culture will encourage artistic expression and civic engagement, informal learning, and social connection.

An ethics of harmony

In 1973 Ivan Illich published *Tools for conviviality*, and as described in Wikipedia (2023), «Illich generalised the themes that he had previously applied to the field of education: the institutionalisation of specialised knowledge, the dominant role of technocratic elites in industrial society, and the need to develop new instruments for the reconquest of practical knowledge by the average citizen». This approach can be thought of as representative of a fostering of an ethical culture on the basis of practical action in the community.

Conviviality isn't an ethical theory at all. Oxford (2023) defines it as «the quality of being cheerful and friendly in atmosphere or character». It is the finding of joy in life, and extending that joy to others in the community. In a practical sense, we find it expressed in society as ambiguity and small things. It is the opposite of a rigid code of discipline and formal principles, the opposite of wanting to know «where the dividing line is». It is the margin we give each other on the road, neither of us pushing right up to the edge. It is attentiveness and reasonableness. It is the small politeness we offer each other in recognition that each of us has feelings and likes to be appreciated. It's the care we take when we do small things (Sheather 2020; Brody 2019).

In many ways, traditional ethics is directed at oneself, as a way, perhaps, of determining on some basis or another what ought to be done. But an ethics of harmony - so we will call it here - is an ethics directed away from oneself, and instead, like an ethics of care, directed at the other. It is, therefore, also essentially an ethics of openness, of seeing and hearing and taking in what is wanted, needed, and offered by the other. It is openness, not as a requirement (as so often it is depicted in the realm of open source software or open content) but as an *opportunity* to contribute to the public good through sharing. «Openness is receptive to others, inclusive rather than exclusive. It is welcoming of diver-

sity. It values (not just «tolerates») others, and seeks to discover the gifts and talents of those others» (Aerisman 1999).

Attention to connectedness and diversity can be seen as a natural consequence of being open to others and being attentive and responsive to their needs. But an ethics of harmony isn't marked by Dei (Diversity, equity and inclusion) initiatives – these would be unnecessary in a culture of conviviality. Our attention in such a community isn't focused on these overarching principles so much as it is on the specific attributes of a particular relationship.

9. Concluding unethical postscript

Ethics based on virtue, duty, or beneficial outcomes are not satisfactory in the case of fields like Ai and learning analytics. We don't agree on what «the good» is. We can't predict what the consequences will be. We can't repair bad consequences after the fact. Ethics – especially in the professions – are typically defined in terms of social contracts, rights or duties – and as such, as statements of rules or principles. But these don't take into account context and particular situations. They also don't take into account the larger interconnected environment in which all this takes place. And they don't take into account how analytics and Ai themselves work.

Instead, as the feminist philosophies of care show us, ethics – including the ethics of Ai – is about relationships, how we interact and care for each other. And as a key point of these interactions, our analytics are always going to reflect *us* (think Michael Wesch: the Machine is us/ing us; think of the case of Tay, the racist Ai based on Tweets). The ethics of Ai is based – in a concrete practical sense – on what we do and what we say *to each other*. This is the ethics we apply when we ask «what makes so-and-so think it would be appropriate to post such-and-such?». If there is a breakdown in the ethics of Ai, it is merely reflective of a breakdown in the social order *generally* (Belshaw 2011).

This breakdown is what motivates us to study the Duty of care, a feminist philosophical perspective that uses a relational and context-bound approach toward morality and decision making, and more importantly, looks at moral and ethical relationships that actually *work*. These are based on different objectives – not «rights» or «fairness» but rather things like a sense of compassion, not on a rigid set of principles but rather an attitude or approach of caring and kindness, not on constraining or managing our temptation to do wrong, but in finding ways to do good.

In the end, ethics are derived from our own lived experiences, and thus reflect the nature of a community as an entire system, rather than one indivi-

dual making a decision. We need to keep in mind how we're all connected. What's important here is how we learn to be ethical in the first place (as opposed to the specific statement of a set of rules defining what it is to be ethical). How should this be approached in practice, in learning, in a workplace, and in society? By creating an ethical culture (rather than emphasis on following the rules), by encouraging a diversity of perspective to create a wider sense of community, and by encouraging openness and interaction (art, drama, etc.) to develop empathy and capacity to see from the perspective of others. None of these are *ethical* principles, but they are the ways we arrive at an ethical society.

References

Ackerman, E. (2019), *My Fight With a Sidewalk Robot*, CityLab, 10 November, https://www.citylab.com/perspective/2019/11/autonomous-technology-ai-robot-delivery-disability-rights/602209/.

Aerisman (1999), *Benchmark Ethics: Openness*, Ethix, https://ethix.org/1999/08/01/openness.

Aizenberg, E. and van den Hoven, J. (2020), *Designing for Human Rights in AI*, in «Big Data & Society», 7(2), doi:10.1177/2053951720949566.

Akselrod, O. (2021), *How Artificial Intelligence Can Deepen Racial and Economic Inequities*, ACLU News & Commentary, July 13, https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R. and Horvitz, E. (2019), *Guidelines for Human-AiInteraction*, Microsoft CHI 2019, May, https://www.microsoft.com/en-us/research/publication/guidelines-for-human-ai-interaction/.

Anderson, C. (2008), *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, Wired, 23 June, https://www.wired.com/2008/06/pb-theory/.

Andrejevic, M. and Selwyn, N. (2019), *Facial Recognition Technology in Schools: Critical Questions And Concerns*, in «Learning, Media and Technology», 45(2), pp. 115–128, doi:10.1080/17439884.2020.1686014.

Aristotle (BCE 350-2003), *The Nicomachean Ethics of Aristotle*, introduction by J. A. Smith, *Gutenberg Project*, https://www.gutenberg.org/files/8438/8438-h/8438-h.htm.

Armstrong, K. and Sheckler, C. (2019), *Why Are Cops Around the World Using This Outlandish Mind-Reading Tool?*, Pro Publica, 7 December, https://www.propublica.org/article/why-are-cops-around-the-world-using-this-outlandish-mindreading-tool.

Bali, M. (2020), *Pedagogy of Care: Covid-19 Edition*, Reflecting Allowed (weblog), 28 May, https://blog.mahabali.me/educational-technology-2/pedagogy-of-care-covid-19-edition/.

Banoula, M. (2023), *What is Perceptron: A Beginners Guide for Perceptron. Online course: Simplilearn*, 10 May, https://www.simplilearn.com/tutorials/deep-learning-tutorial/perceptron.

Belshaw, D. (2011), *Anarchy in The UK? The Reasons Behind The Breakdown of Social Order. Open Thinkering*, 22 August, https://dougbelshaw.com/blog/2011/08/22/anarchy-in-the-uk-the-reasons-behind-the-breakdown-of-social-order/.

Biddle, E. R. and Zhang, J. (2020), *Moving Fast and Breaking Us All: Big Tech's Unaccountable Algorithms, Ranking Digital Rights*, https://rankingdigitalrights.org/index2020/spotlights/unaccountable-algorithms.

Blackman, R. (2020), *A Practical Guide to Building Ethical AI*, Harvard Business Review, 15 October, https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai.

Bodie, R. (2013), *The Ethics of Online Anonymity or Zuckerberg vs. «Moot»*, in «Computers and Society», 43(1), pp. 22-35, https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/Communications/BodleRobert.pdf.

Bostrom, N. and Yudkowsky, E. (2014), *The Ethics of Artificial Intelligence*, in K. Frankish and W. M. Ramsey (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge, Cambridge University Press, pp. 316-334.

Boyd, D. and Crawford, K. (2012), *Critical Questions for Big Data*, in «Information, Communication & Society», 15(5), pp. 662-679, https://www.dhi.ac.uk/san/waysofbeing/data/communication-zangana-boyd-2012.pdf.

Boyer, A. and Bonnin, G. (2019), *Higher Education and the Revolution of Learning Analytics*, International Council For Open And Distance Education, https://static1.squarespace.com/static/5b99664675f9eea7a3ecee82/t/5beb449703ce644d00213dc1/1542145198920/anne_la_report+cc+licence.pdf.

Brandon, S., Arthur, J., Ray, D., Meissner, C., Kleinman, S., Russano, M. and Wells, S. (2019), *The High-Value Detainee Interrogation Group (HIG): Inception, Evolution, and Impact*, in S. C. Harvey and M. A. Staal (eds.), *Operational Psychology: A New Field to Support National Security and Public Safety*, London, Bloomsbury Publishing, pp. 263-285.

Brodsky, A., Shao, G., Krishnamoorthy, M., Narayanan, A., Menascé, D. and Ak, R. (2015), *Analysis and Optimization in Smart Manufacturing based on a Reusable Knowledge Base for Process Performance Models*, paper presented at the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, Usa, doi: 10.1109/BigData.2015.7363902.

Brody, S.H. (2019), *The Law of Small Things: Creating a Habit of Integrity in a Culture of Mistrust*, Oakland, Berrett-Koehler Publishers.

Broomhead, D.S. and Lowe, D. (1988), *Multivariable Functional Interpolation and Adaptive Networks*, in «Complex Systems», 2, pp. 321-355, https://content.wolfram.com/sites/13/2018/02/02-3-5.pdf.

Buckingham Shum, S. J. and Crick, R. D. (2012), *Learning Dispositions And Transferable Competencies: Pedagogy, Modelling and Learning Analytics, paper presented at the 2nd International Conference on Learning Analytics & Knowledge*, Vancouver, Canada, 29 April-2 May, http://oro.open.ac.uk/32823/1/SBS-RDC-LAK12-ORO.pdf.

Calvano, E., Calzolari, G., Denicolò, V. and Pastorello, S. (2020), *Artificial Intelligence, Algorithmic Pricing, and Collusion*, in «American Economic Review», 110 (10), pp. 3267-97, https://www.aeaweb.org/articles?id=10.1257/aer.20190623.

Carpenter, Todd A. (2020), *If My Ai Wrote this Post, Could I Own the Copyright?*, in «The Scholarly Kitchen», 12 February, https://scholarlykitchen.sspnet.org/2020/02/12/if-my-ai-wrote-this-post-could-i-own-the-copyright/.

Cath, C. (2018), *Governing Artificial Intelligence: Ethical, Legal And Technical Opportunities and Challenges*, in «Philosophical Transactions of the Royal Society A», http://doi.org/10.1098/rsta.2018.0080.

Cavoukian, A. (2013), *Information and Privacy Commissioner*, Ontario, Canada, https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-surveillance.pdf.

Chesney, R. and Citron, D. K. (2018), *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, in «California Law Review», 107, pp. 1753-1819, doi:10.15779/Z38RVOD15J.

Clay, J. (2020). *Data Does Matter, and so Does Ethics*, in Jisc, 24 January, https://analytics.jiscinvolve.org/wp/2020/01/24/data-does-matter-and-so-does-ethics/.

Cohen, I.G., Amarasingham, R., Shah, A., Xie, B. and Lo, B. (2014), *The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics*, in «Health Affairs», 33(7), doi:10.1377/hlthaff.2014.0048.

Courtney, H., Lovallo, D. and Clarke, C. (2013), *Deciding How to Decide*, Harvard Business Review, 13 November, https://hbr.org/2013/11/deciding-how-to-decide.

Cunningham, R. (2008), *How Standards Proliferate*. XKCD (web comic) number 927, https://xkcd.com/927/.

Damen, L. (1987), *Culture Learning: The Fifth Dimension on the Language Classroom*, Reading, MA, Addison-Wesley.

Danzig, L. (2020). *The Road to Artificial Intelligence: An Ethical Minefield*, InfoQ, 6 January, https://www.infoq.com/articles/algorithmic-integrity-ethics/.

De Bruijn, B., Désillets, A., Fraser, K., Kiritchenko, S., Mohammad, S., Vinson, N., Bloomfield, P., Brace, H., Brzoska, K., Elhalal, A., Ho, K., Kinsey, L., McWhirter, R., Nazare, M. and Ofori-Kuragu, E. (2019), *Applied Ai ethics: Report 2019*, National Research Council Canada, https://nrc-publications.canada.ca/eng/view/fulltext/?id=a9064070-feb7-4c97-ba87-1347e41ec06a.

Demiaux, V. and Abdallah, Y. S. (2017), *How Can Humans Keep the Upper hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence*, French Data

Protection Authority, Commission Nationale Informatique & Libertés (CNIL), https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf.

Devlin, H. (2017), *Ai Programs Exhibit Racial and Gender Biases*, Research Reveals, The Guardian, 13 April, https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals.

Dhuri, S. (2020), *Explainable AI: Making Sense of the Black Box*, The Startup, 6 October, https://medium.com/swlh/explainable-ai-making-sense-of-the-black-box-32ebf2d16c61.

Donges, N. (2019), *A Guide to Recurrent Neural Networks: Understanding RNN and LSTM Networks*, BuiltIn, 28 February, https://builtin.com/data-science/recurrent-neural-networks-and-lstm.

Dressel, J. and Farid, H. (2018), *The Accuracy, Fairness, and Limits Of Predicting Recidivism*, in «Science», 4(1), doi:10.1126/sciadv.aao5580.

Drew, C. (2018), *Design For Data Ethics: Using Service Design Approaches To Operationalize Ethical Principles On Four Projects*, in «Philosophical Transactions of the Royal Society A», doi:10.1098/rsta.2017.0353.

Dringus, L. P. (2012), *Learning Analytics Considered Harmful*, in «Journal of Asynchronous Learning Networks», 16(2), 87-100.

Eckersley, P., Gillula, J. and Williams, J. (2017), *Electronic Frontier Foundation – Written evidence (AIC0199) (House of Lords Select Committee on Artificial Intelligence)*, http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69720.html.

Edureka (2023), *Artificial Intelligence Algorithms: All you Need to Know*, Edureka, 2 August, https://www.edureka.co/blog/artificial-intelligence-algorithms/.

Ellemers, N., van der Toorn, J., Paunov, Y. and van Leeuwen, T. (2019), *The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017*, in «Personality and Social Psychology Review», 23(4), 332-366, doi:10.1177/1088868318811759.

Emory University Libraries (2019), *Policy on the Collection, Use, and Disclosure of Personal Information*, http://web.library.emory.edu/privacy-policy/personal-information.html.

European Commission's High-Level Expert Group on Artificial Intelligence (2019), *Ethics Guidelines for Trustworthy AI*, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M. (2020), *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI, Berkman Klein Center for Internet & Society,* http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2018), *AI4People—An Ethical Framework for a Good Ai*

Society: Opportunities, Risks, Principles, and Recommendations, in «Minds and Machines», 28(4), 689-707, doi:10.1007/s11023-018-9482-5.

Floridi, L. and Cowls, J. (2019), *A United Framework of Five Principles For Ai in Society*, in «Harvard Data Science Review», 1 (1), doi:10.15779/Z38RVOD15J.

Foot, P. ([1967]1978), *The Problem of Abortion and the Doctrine of the Double Effect in Virtues and Vices*, Oxford, Basil Blackwell, (Originally appeared in the Oxford Review, 5, 1967, http://www2.econ.iastate.edu/classes/econ362/hallam/Readings/FootDoubleEffect.pdf).

Haffey, K. (2023), *Posthuman Scale and the Care to Come / Wayward Voices: Intimate Narration in Saidiya Hartman*, ASAP Journal, 3 May, https://asapjournal.com/posthuman-scale-and-the-care-to-come-wayward-voices-intimate-narration-in-saidiya-hartman-kate-haffey.

Hamel, S. (2016), *The Elasticity of Analytics Ethics*, Medium, 28 June, https://stephane-hamel.medium.com/the-elasticity-of-analytics-ethics-7d8ac253a3b9.

Harris, E.L. (2017), *The Philosophy of the Mozi: The First Consequentialists, Notre Dame Philosophical Reviews*, 7 May, https://ndpr.nd.edu/reviews/the-philosophy-of-the-mozi-the-first-consequentialists/.

Hebb D.O. (1949), *The Organization of Behavior*. *A Neuropsychological Theory*, New York, Wiley.

Hobbes, T. (1994), *Leviathan with Selected Variants From the Latin Edition of 1668,* ed. by E. Curley, Indianapolis, Hackett.

Gašević, D., Dawson, S. and Siemens, G. (2015), *Let's Not Forget: Learning Analytics are About Learning*, in «TechTrends», 59, pp. 64-71.

Gilbert, S. and Lynch, N. (2002), *Brewer's Conjecture and The Feasibility Of Consistent*, Available, Partition-Tolerant Web Services, «ACM SIGACT News», 33(2), pp. 51-59, doi:10.1145/564585.564601.

Greller, W. and Drachsler, H. (2012), *Translating Learning Into Numbers: A Generic Framework for Learning Analytics*, in «Educational Technology & Society», 15(3), pp. 42-57.

Griggs, M. B. (2019), *Google Reveals «Project Nightingale» After Being Accused of Secretly Gathering Personal Health Records*, The Verge, 14 November, https://www.theverge.com/2019/11/11/20959771/google-health-records-project-nightingale-privacy-ascension.

Griffiths, D., Drachsler, H., Kickmeier-Rust, M., Steiner, C., Hoel, T. and Greller, W. (2016), *Is Privacy A Show-Stopper For Learning Analytics? A Review of Current Issues and Solutions*, in «Learning Analytics Review», 6(15), http://www.laceproject.eu/learning-analytics-review/files/2016/04/LACE-review-6_privacy-show-stopper.pdf.

Guillaud, H. (2020), *Des limites du recrutement automatisé*, in InternetActu.net, 28 February, http://www.internetactu.net/a-lire-ailleurs/des-limites-du-recrutement-automatise/.

Hassoun, M. (1995), *Fundamentals of Artificial Neural Networks*, Cambridge, MIT Press, https://neuron.eng.wayne.edu/tarek/MITbook/t_contents.html.

Hochreiter, S. and Schmidhuber, J. (1997), *Long Short-Term Memory*, in «Neural Computation», 9 (8), pp. 1735-1780, https://direct.mit.edu/neco/article-abstract/9/8/1735/6109/Long-Short-Term-Memory.

Hodgett, S., Liu, T. and Ip, S. (2023), *AI, Machine Learning & Big Data Laws and Regulations 2023*, Canada, Global Legal Insights, https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/canada.

Hofstede, G. (1984), *National Cultures and Corporate Cultures*, in L.A. Samovar and R.E. Porter (eds.), *Communication Between Cultures*, Belmont, Wadsworth.

Hopfield, J. J. (1982), *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, in «Proceedings of the National Academy of Sciences», 79(8), pp. 2554–2558, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC346238/.

Hotchkiss, K. (2019), *With Great Power Comes Great (Eco) Responsibility – How Blockchain is Bad for the Environment*, Georgetown Law, 6 April, https://www.law.georgetown.edu/environmental-law-review/blog/with-great-power-comes-great-eco-responsibility-how-blockchain-is-bad-for-the-environment/.

International Union of Psychological Science-IUPSYS (2008), *Universal Declaration of Ethical Principles for Psychologists, International Union of Psychological Science*, https://www.iupsys.net/about/governance/universal-declaration-of-ethical-principles-for-psychologists.html.

Jambekar, S. (2017), *GDPR: Data Subjects, Controllers and Processors*, Oh My!, Twilio Blog, 4 October, https://www.twilio.com/blog/gdpr-data-subjects-controllers-processors-html.

Jaschik, S. (2016), *Are At-Risk Students Bunnies to Be Drowned?*, in «Inside Higher Ed.», 20 January, https://www.insidehighered.com/news/2016/01/20/furor-mount-st-marys-over-presidents-alleged-plan-cull-students.

Jenkins, H. (2006), *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*, John D. and Catherine T. MacArthur Foundation, https://www.macfound.org/media/article_pdfs/jenkins_white_paper.pdf.

Jones, H. (2011), *Taking Responsibility for Complexity. How Implementation Can Achieve Results in The Face of Complex Problems*, ALNAP, https://www.alnap.org/help-library/taking-responsibility-for-complexity-how-implementation-can-achieve-results-in-the-face.

Jones, L. M. and Mitchell, K. J. (2016), *Defining and Measuring Youth Digital Citizenship*, in «New Media & Society», 18(9), pp. 2063-2079, doi:10.1177/1461444815577797.

Katz, J. (1997), *The Digital Citizen and Birth of a Digital Nation*, Wired, 1 December, https://www.wired.com/1997/12/netizen-29/.

Keppler, N. (2020), *Cost Cutting Algorithms Are Making Your Job Search a Living Hell*, Vice, 12 February, https://www.vice.com/en_us/article/pkekvb/cost-cutting-algorithms-are-making-your-job-search-a-living-hell.

Khalil, M. and Ebner, M. (2015), *Learning Analytics: Principles and Constraints.* 1789-1799, paper presented at EdMedia 2015, Montreal, Quebec, Canada, June 22-24, https://pure.tugraz.at/ws/portalfiles/portal/3217534/edmedia2015.pdf.

Kowalewski, M. (2020), *Data Cleaning In 5 Easy Steps + Examples, Iterators*, 23 September, https://www.iteratorshq.com/blog/data-cleaning-in-5-easy-steps/.

Kumar, S. (2021), *3 Types of Classification Problems in Machine Learning: Deep Dive Analysis of Binary Classification, Multi-Class Classification, and Multi-Label Classification*, in «The Startup», 21 January, https://medium.com/swlh/3-types-of-classification-problems-in-machine-learning-1cffd3765ca1.

Kumar, V. and Campbell, R. (2023), *Response to Louise Antony (Review of A Better Ape: The Evolution of the Moral Mind and How it Made us Human, NDPR, 4 August 2023)*, Medium, 21 September, https://medium.com/@victor.c.kumar/response-to-louise-antony-4e6b43ce94bf.

Kwan, J., Mclean, M. and Raicu, I. (2021), *Introduction to «A Framework for Ethical Decision-Making»*, Markkula Center for Applied Ethics, https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/introduction-to-a-framework-for-ethical-decision-making/.

Li, Y. and Lyu, S. (2019), *Exposing DeepFake Videos By Detecting Face Warping Artifacts*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops 2019, https://github.com/danmohaha/CVPRW2019_Face_Artifacts.

Liberman, M. (2020), *Perils of Topic Modeling*, Language Log, 5 May, https://languagelog.ldc.upenn.edu/nll/?p=46996.

Lioeberman, M. (2020), *Using Student Data to Identify Future Criminals: A Privacy Debacle*, EducationWeek, 24 November, https://www.edweek.org/technology/using-student-data-to-identify-future-criminals-a-privacy-debacle/2020/11.

Liu, H. Y., Maas, M., Danaher, J., Scarcella, L., Lexer, M. and Van Rompaey, L. (2020), *Artificial Intelligence and Legal Disruption: A New Model for Analysis*, in «Law, Innovation and Technology», 12(2), pp. 205-258, https://philpapers.org/archive/DANAIA-6.pdf.

Long D. and Magerko, B. (2020), *What is Ai Literacy? Competencies and Design*, CHI 2020, April 25-30, 2020, Honolulu, HI, Usa, doi:10.1145/3313831.3376727.

Loshin, D. (2002), *Knowledge Integrity: Data Ownership*, Datawarehouse, 8 June, http://www.datawarehouse.com/article/?articleid=3052.

Lu, X. (2019), *An Empirical Study on the Artificial Intelligence Writing Evaluation System in China CET*, in «Big Data», 7(2), pp. 121-129, doi:10.1089/big.2018.0151.

Mackie, J. L. (1983), *Ethics: Inventing Right and Wrong*, London, Penguin Books.

Meinecke, S. (2018), *Ai Could Help Us Protect the Environment — Or Destroy It*, DW, 16 July, https://p.dw.com/p/31X4Z.

Metz, R. (2020), *There's A New Obstacle to Landing A Job After College: Getting Approved By AI*, CNN Business, 15 January, https://www.cnn.com/2020/01/15/tech/ai-job-interview/index.html.

Miikkulainen, R. (2011), *Neuroevolution*, in Sammut, C., Webb, G.I. (eds.) *Encyclopedia of Machine Learning*, Springer, Boston, MA, https://doi.org/10.1007/978-0-387-30164-8_589.

Mill, J. S. ([1879]2004), *Utilitarianism*. Reprinted from Fraser's Magazine. Seventh Edition. Longmans, Green, and Co, 1879. Gutenberg Project, February 1, 2004 [eBook #11224], Most recently updated: 25 December, 2020, https://www.gutenberg.org/cache/epub/11224/pg11224-images.html.

Mitra, A. (2018), *We Can Train Ai To Identify Good and Evil, and Then Use It To Teach Us Morality*, Quartz, 5 April, https://qz.com/1244055/we-can-train-ai-to-identify-good-and-evil-and-then-use-it-to-teach-us-morality/.

Morris, D. Z. (2016), *Mercedes-Benz's Self-Driving Cars Would Choose Passenger Lives Over Bystanders, Fortune*, 15 October, https://fortune.com/2016/10/15/mercedes-self-driving-car-ethics/.

Mossberger, K., Tolbert, C.J. and McNeal, R.S. (2007), *Digital Citizenship: The Internet, Society, and Participation*, Cambridge, MIT Press.

Motta, S. and Bennett, A. (2018), *Pedagogies of Care, Care-full Epistemological Practice and «Other» Caring Subjectivities in Enabling Education*, in «Teaching in Higher Education», 23(5), pp. 631-646, doi:10.1080/13562517.2018.1465911.

Mozi. (1929), *The Ethical and Political Works of Motse, Probsthain*, 1929, revised, 1978. Trans. Mei, Yi-Pao, Chinese Text project, https://ctext.org/mozi.

Narayan, A. (n.d.), *How to Recognize Ai Snake Oil. In Promises and Perils of AI*, Retrieved November 27, 2019, from Google Docs, https://docs.google.com/document/d/1s_AgoeL2y_4iuedGuQNH6Fl1744twhe8Kj2qSfTqyHg/edit?fbclid=IwAR0QSuS-QXJB8rxgni_zGm5KU0oQPa9AJPFv-NpKcBlOkKIJZ0J4uefhg0o#heading=h.ypt4v4y21eo5.

Newell, A., Shaw, J.C. and Simon, H.A. (1959), *Report on a General Problem-solving Program*, Proceedings of the International Conference on Information Processing (pp. 256-264), http://bitsavers.informatik.uni-stuttgart.de/pdf/rand/ipl/P-1584_Report_On_A_General_Problem-Solving_Program_Feb59.pdf.

Nietzsche, F. (1967), *The Birth of Tragedy and the Case of Wagner*, trans. W. Kaufmann, New York, Vintage.

Nietzsche, F. (1999), *Thus Spake Zarathustra*, trans. Thomas Common, eBook #1998, Gutenberg Library, https://www.gutenberg.org/files/1998/1998-h/1998-h.htm#link2H_4_0009.

Nozick, R. (1981), *Philosophical Explanations*, Cambridge, MA, Harvard University Press.

O'Leary, K. and Murphy, S. (2019), *Anonymous Apps Risk Fuelling Cyberbullying But They Also Fill A Vital Role*, The Conversation, 11 July, https://theconversation.com/anonymous-apps-risk-fuelling-cyberbullying-but-they-also-fill-a-vital-role-119836.

Oravec, J. (2022), *AI, Biometric Analysis, and Emerging Cheating Detection Systems: The Engineering of Academic Integrity?*, in «Education Policy Analysis Archives», 30(175), pp. 11-18, doi:10.14507/epaa.30.5765.

Oxford Learner Dictionaries (2023), *Conviviality*, https://www.oxfordlearnersdictionaries.com/definition/english/conviviality.

Pappas, S. and McKelvie, C. (2022), *What is Culture?*, LiveScience, 17 October, https://www.livescience.com/21478-what-is-culture-definition-of-culture.html.

Parkes, D. (2019), *A Responsibility to Judge Carefully In the Era of Prediction Decision Machines*, Harvard Business School Digital Initiative, 2 December, https://d3.harvard.edu/a-responsibility-to-judge-carefully-in-the-era-of-prediction-strikethrough-decision-machines/.

Pandya, A., Gilbar, T. and Kim, K. (2005), *Neural Network Training Using a GMDH Type Algorithm*, in «International Journal of Fuzzy Logic and Intelligent Systems», 5, pp. 52-58, https://koreascience.kr/article/JAKO200516610534768.pdf.

Passi, S. and Jackson, S. (2018), *Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects*, in «Proceedings of the ACM on Human-Computer Interaction», 2, pp. 1-28, doi:10.1145/3274405.

Pojman, L. P. (1990), *Ethics: Discovering Right and Wrong*, Wadsworth, Wadsworth Publishing.

Pariser, E. (2012), *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*, New York, Penguin.

Paul, C. and Posard, M. N. (2020), *Artificial Intelligence and the Manufacturing of Reality*, 20 January, The RAND Blog, January, https://www.rand.org/blog/2020/01/artificial-intelligence-and-the-manufacturing-of-reality.html.

Privacy Comissioner of Canada (2020), *A Regulatory Framework for AI: Recommendations for PIPEDA Reform*, Government of Canada, https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/reg-fw_202011/.

Radan, N. (2019), *Ethical Use of Artificial Intelligence for Actuaries. Society of Actuaries*, https://www.soa.org/globalassets/assets/files/resources/research-report/2019/ethics-ai.pdf.

Rawls, J. (1999), *A Theory of Justice*, Cambridge, MA, Harvard University Press.

Rawls, J. (2001), *Justice as Fairness: A Restatement*, Cambridge, MA, Harvard University Press.

Rienties, B. and Jones, A. (2019), *Evidence-Based Learning: Futures*, in R. Ferguson and A. Jones (eds.), *Educational Visions: Lessons from 40 Years of Innovation*, London, Ubiquity Press, pp. 109-125, doi:org/10.5334/bcg.g.

Rieke, A., Bogen, M. and Robinson, D.G. (2018), *Public scrutiny of automated decisions: early lessons and emerging methods*, Analysis & Policy Observatory, 12 December, https://apo.org.au/node/210086.

Rodrigues, R. (2020), *Legal And Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities*, in «Journal of Responsible Technology», 4, pp. 1-12, https://doi.org/10.1016/j.jrt.2020.100005.

Roscoe, R. D., Wilson, J. and Johnson, A. C. (2017), *Presentation, Expectations, and Experience: Sources of Student Perceptions of Automated Writing Evaluation*,

in «Computers in Human Behavior», 70, pp. 207-221, doi:org/10.1016/j.chb.2016.12.076.

Rosenblatt, F. (1958), *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, Cornell Aeronautical Laboratory, in «Psychological Review», 65(6), pp. 386-408, doi:10.1037/h0042519.

Rousseau, J-J. ([1762] 2004), *The Social Contract*, London, Penguin Books.

Rumelhart, D., Hinton, G. and Williams, R. (1986), *Learning Representations by Back-propagating Errors*, in «Nature» 323, pp. 533-536, doi:org/10.1038/323533a0.

Rushkoff, D. (1994), *Cyberia: Life in the Trenches of Cyberspace*, New York, HarperCollins.

Schildkraut, P. (2021), *Ai Regulation: What You Need to Know to Stay Ahead of the Curve*, Arnold & Porter Kaye Scholer LLP, https://www.arnoldporter.com/-/media/files/perspectives/publications/2021/06/ai-regulationstaying-ahead-of-curveschildkraut0621.pdf?la=en.

Schintler, L.A., McNeely, C.L. and Witte, J. (2023), *A Critical Examination of the Ethics of AI-Mediated Peer Review*, in «arXiv», https://doi.org/10.48550/arXiv.2309.12356.

Schneier, B. (2020), *We're Banning Facial Recognition. We're Missing the Point*, New York Times, 20 January, https://www.nytimes.com/2020/01/20/opinion/facial-recognition-ban-privacy.html.

Sclater, N. and Bailey, P. ([2015]2023), *Code of Practice for Learning Analytics*, Jisc, https://beta.jisc.ac.uk/guides/code-of-practice-for-learning-analytics.

Sclater, N., Peasgood, A. and Mullan, J. (2016), *Learning Analytics in Higher Education a Review of UK and International Practice*, Jisc, https://www.jisc.ac.uk/sites/default/files/learning-analytics-in-he-v3.pdf.

Seufert, S., Meier, C., Soellner, M. and Rietsche, R. (2019), *A Pedagogical Perspective on Big Data and Learning Analytics: A Conceptual Model for Digital Learning Support*, in «Technology, Knowledge and Learning», 24, pp. 599-619, doi:org/10.1007/s10758-019-09399-5.

Shaw, J. (2017), *The Watchers: Assaults on Privacy In America*, Harvard Magazine, January-February, https://www.harvardmagazine.com/2017/01/the-watchers.

Sheather, J. (2020), *The Ethics of Small Things*, The RSA, 24 June, https://medium.com/@thersa/the-ethics-of-small-things-e24f03a5b62a.

Shwayder, M. (2020), *Clearview AI's Facial-Recognition App is A Nightmare for Stalking Victims*, Digital Trends, 22 January, https://www.digitaltrends.com/news/clearview-ai-facial-recognition-domestic-violence-stalking/.

Siemens, G. (2012), *Learning Analytics: Envisioning a Research Discipline and a Domain of Practice*, Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, https://dl.acm.org/citation.cfm?id=2330605.

Singer, N. (2017), *How Google Took Over the Classroom*, New York Times, 13 May, https://www.nytimes.com/2017/05/13/technology/google-education-chromebooks-schools.html.

Singh, R. (2020), *Resistance in a Minor Key: Care, Survival and Convening on the Margins*, in First Monday, 25 (5-4), https://firstmonday.org/ojs/index.php/fm/article/download/10631/9418.

Suler, J. (2004), *The Online Disinhibition Effect*, in «CyberPsychology and Behavior», 7(3), pp. 321-326, doi:10.1089/1094931041291295.

Tanmay, K., Khandelwal, A., Agarwal, U. and Choudhury, M. (2023), *Exploring Large Language Models' Cognitive Moral Development through Defining Issues Test (Preprint under review)*, in «arXiv», https://arxiv.org/pdf/2309.13356.pdf

TeachThought Staff (2019), *Moving Students From Digital Citizenship To Digital Leadership*, TeachThought, 26 November, https://www.teachthought.com/the-future-of-learning/digital-leadership-2/.

Thomas, D. and Seely Brown, J. (2011), *A New Culture of Learning: Cultivating the Imagination for a World of Constant Change, CreateSpace Independent Publishing Platform*, 4 January, http://www.newcultureoflearning.com/newcultureoflearning.pdf.

Treviranus, J. (2018), *The Three Dimensions of Inclusive Design: A Design Framework for a Digitally Transformed and Complexly Connected Society*, PhD thesis, University College Dublin, http://openresearch.ocadu.ca/id/eprint/2745/.

Tsui, Q. (2023), *Dethroning the All-powerful AI: Developing Ethics for a Demystified AI, Bot Populi*, 26 September, https://botpopuli.net/dethroning-the-all-powerful-ai-developing-ethics-for-a-demystified-ai/.

Tufekci, Z. (2018), *YouTube, the Great Radicalizer*, New York Times, 10 March, https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

Tuomi, I. (2018), *The Impact of Artificial Intelligence on Learning, Teaching, and Education*, in M. Cabrera, R. Vuorikari and Y. Punie (eds.), *JRC Science for Policy Report European Union*, doi:10.2760/12297.

Turing, A. M. (1936), *On Computable Numbers, with an Application to the Entscheidungsproblem*, in «Proceedings of the London Mathematical Society», 2(42), pp. 230-265, doi:10.1112/plms/s2-42.1.230.

Upadhyay, Y. (2019), *Introduction to FeedForward Neural Networks*, Towards Data Science, 7 March, https://towardsdatascience.com/feed-forward-neural-networks-c503faa46620.

van Rees R. (2003), *Clarity in The Usage of The Terms Ontology*, Taxonomy and Classification, in «CIB REPORT», 284(432), pp. 1-8, http://itc.scix.net/paper/w78-2003-432.

Vance, J. (2021), *Vicious Perceptual Expertise*, The Brains Blog, 29 September, https://philosophyofbrains.com/2021/09/29/jonna-vance-vicious-perceptual-expertise.aspx.

Vesset, D. (2018), *Descriptive analytics 101: What happened?*, IBM Blog, 10 May, https://www.ibm.com/blogs/business-analytics/descriptive-analytics-101-what-happened/.

Vivek, N. and Collier, R. (2019), *Ethics by Agreement in Multi-agent Software Systems*, Proceedings of the 14th International Conference on Software Technologies 2019, Scitepress, https://www.scitepress.org/Papers/2019/79581/79581.pdf.

Wahing, R. J. C. (2021), *Confucius on the Five Constant Virtues*, in «Modern Asian Thoughts», pp. 1-17, https://philpapers.org/archive/WAHCOT-2.pdf.

Wang, J., Liu, Z., Zhao, L., Wu, Z., Ma, C., Yu, S., Dai, H., Yang, Q., Liu, Y., Zhang, S. and Shi, E. (2023), *Review of Large Vision Models and Visual Prompt Engineering*, in «arXiv», https://arxiv.org/abs/2307.00855.

Watters, A. (2019), *HEWN 317*, Hack Education, 17 August, https://hewn.substack.com/p/hewn-no-317.

Weinberger, D. (2021a), *Parler and the Failure of Moral Frameworks*, Joho the Blog, 11 January, https://www.hyperorg.com/blogger/2021/01/11/parler-and-the-failure-of-moral-frameworks/.

Weinberger, D. (2021b), *Learn from Machine Learning*, Aeon, 15 November, https://aeon.co/essays/our-world-is-a-black-box-predictable-but-not-understandable.

Wikipedia (2023), *Tools for Conviviality*, 28 August, https://en.wikipedia.org/w/index.php?title=Tools_for_Conviviality&oldid=1172690716.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M. and Tinkler, J. (2015), *The Metric Tide. Report of The Independent Review of the Role of Metrics in Research Assessment and Management*, https://blogs.lse.ac.uk/impactofsocialsciences/files/2015/07/2015_metrictide.pdf.

Yeung, D. (2023), *The AiConspiracy Theories Are Coming*, The Rand Blog, 22 June, https://www.rand.org/blog/2023/06/the-ai-conspiracy-theories-are-coming.html.

Young, J., Shaxson, L., Jones, H., Hearn, S., Datta, A. and Cassidy, C. (2014), *Rapid Outcome Mapping Approach (ROMA): A Guide to Policy Engagement and Influence*, https://i2s.anu.edu.au/wp-content/uploads/2018/01/9011.pdf.

Young, N.T. (2020), *I Know Some Algorithms Are Biased — because I Created One*, Scientific American, 31 January, https://blogs.scientificamerican.com/voices/i-know-some-algorithms-are-biased-because-i-created-one/.

Zawacki-Richter, O., Marín, V. I., Bond, M. and Gouverneur, F. (2019), *Systematic Review of Research On Artificial Intelligence Applications In Higher Education – Where Are The Educators?*, in «International Journal of Technology in Higher Education», 16(39), doi:10.1186/s41239-019-0171-0.

Zimmermann, A., Rosa, E. D. and Kim, H. (2020), *Technology Can't Fix Algorithmic Injustice*, Boston Review, 9 January, https://bostonreview.net/science-nature-politics/annette-zimmermann-elena-di-rosa-hochan-kim-technology-cant-fix-algorithmic.