

# Getting from Commitment to Content in AI and Data Ethics: Justice and Explainability

**John Basl**  
**Ronald Sandler**  
**Steven Tiell**





# Atlantic Council

## GEOTECH CENTER

**Cover:** Lance Anderson/Unsplash



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

This report is written and published in accordance with the Atlantic Council Policy on Intellectual Independence. The author is solely responsible for its analysis and recommendations. The Atlantic Council and its donors do not determine, nor do they necessarily endorse or advocate for, any of this report's conclusions.

Atlantic Council  
1030 15th Street NW, 12th Floor  
Washington, DC 20005

For more information, please visit  
[www.AtlanticCouncil.org](http://www.AtlanticCouncil.org).

**ISBN:** 978-1-61977-189-5

August 2021

This report was designed by Donald Partyka and Liam Brophy

# **Getting from Commitment to Content in AI and Data Ethics: Justice and Explainability**

**John Basl**

**Ronald Sandler**

**Steven Tiell**

# TABLE OF CONTENTS

<b>3</b>	<b>I. Introduction</b>
<b>6</b>	<b>II. Example: Informed Consent in Bioethics</b>
<b>8</b>	<b>III. What Is Justice in Artificial Intelligence and Machine Learning?</b>
<b>9</b>	1. The Values and Forms of Justice
<b>10</b>	2. Justice in Artificial Intelligence and Data Systems
<b>13</b>	<b>IV. What Is Transparency in Artificial Intelligence and Machine Learning?</b>
<b>13</b>	1. The Values and Concepts Underlying Transparency
<b>14</b>	2. Specifying Transparency Commitments
<b>17</b>	<b>V. Conclusion: From Normative Content to Specific Commitments</b>
<b>19</b>	<b>VI. Contributors</b>

## EXECUTIVE SUMMARY:

There is widespread awareness among researchers, companies, policy makers, and the public that the use of artificial intelligence (AI) and big data raises challenges involving justice, privacy, autonomy, transparency, and accountability. Organizations are increasingly expected to address these and other ethical issues. In response, many companies, non-governmental organizations, and governmental entities have adopted AI or data ethics frameworks and principles meant to demonstrate a commitment to addressing the challenges posed by AI and, crucially, guide organizational efforts to develop and implement AI in socially and ethically responsible ways.

However, articulating values, ethical concepts, and general principles is only the first step—and in many ways the easiest one—in addressing AI and data ethics challenges. The harder work is moving from values, concepts, and principles to substantive, practical commitments that are action-guiding and measurable. Without this, adoption of broad commitments and principles amounts to little more than platitudes and “ethics washing.” The ethically problematic development and use

of AI and big data will continue, and industry will be seen by policy makers, employees, consumers, clients, and the public as failing to make good on its own stated commitments.

The next step in moving from general principles to impacts is to clearly and concretely articulate what justice, privacy, autonomy, transparency, and explainability actually involve and require in particular contexts. The primary objectives of this report are to:

- demonstrate the importance and complexity of moving from general ethical concepts and principles to action-guiding substantive content;
- provide detailed discussion of two centrally important and interconnected ethical concepts, justice and transparency; and
- indicate strategies for moving from general ethical concepts and principles to more specific substantive content and ultimately to operationalizing those concepts.

# I. INTRODUCTION

There is widespread awareness among researchers, companies, policy makers, and the public that the use of artificial intelligence (AI) and big data analytics often raises ethical challenges involving such things as justice, fairness, privacy, autonomy, transparency, and accountability. Organizations are increasingly expected to address these issues. However, they are asked to do so in the absence of robust regulatory guidance. Furthermore, social and ethical expectations exceed legal standards, and they will continue to do

so because the rate of technological innovation and adoption outpaces that of regulatory and policy processes.

In response, many organizations—private companies, non-governmental organizations, and governmental entities—have adopted AI or data ethics frameworks and principles. They are meant to demonstrate a commitment to addressing the ethical challenges posed by AI and, crucially, guide organizational efforts to develop and implement AI in socially and ethically responsible ways.

**Table 1:** Examples of Principles, Codes, and Value Statements

ORGANIZATION NAME	ORGANIZATION TYPE	PRINCIPLES NAME	FOCUS AREA
Royal Academy of Engineering	Professional organization	<i>Statement of Ethical Principles</i>	General AI (for engineering professionals)
FAT/ML	Professional organization	<i>Principles for Accountable Algorithms</i>	General AI
Google	Public company	<i>Artificial Intelligence at Google: Our Principles</i>	General AI
Université de Montréal	Academic/public partnership	<i>Montreal Declaration for Responsible AI</i>	General AI
Microsoft	Public company	<i>Microsoft AI Principles</i>	General AI
European Commission	Intergovernmental organization	<i>Ethics Guidelines for Trustworthy AI</i>	General AI
Accenture	Public company	<i>Universal Data Ethics Principles</i>	Data Ethics
Institute for the Future, Omidyar Network	Public/private partnership	<i>Ethical OS Toolkit</i>	Any information/computing technology
Partnership on AI	Nonprofit/industry/academic coalition	<i>Partnership Tenets of AI</i>	General AI
Organisation for Economic Co-operation and Development	Intergovernmental organization	<i>Principles on AI</i>	General AI
US Department of Defense	Governmental organization	<i>Principles of AI Ethics</i>	Defense/weaponry AI

As the ethical issues associated with AI and big data have become manifest, organizations have developed principles, codes, and value statements that (1) signal their commitment to socially responsible AI; (2) reflect their understanding of the ethical challenges posed by AI; (3) articulate their understanding of what socially responsible AI involves; and (4) provide guidance for organizational efforts to accomplish that vision.

Identifying a set of ethical concepts, and formulating general principles using those concepts, is a crucial component of articulating AI and data ethics. Moreover, there is considerable convergence among the many frameworks that have been developed. They coalesce around core concepts, some of which are individual oriented, some of which are society

oriented, and some of which are system oriented (see Table 2). That there is a general overlapping consensus on the primary ethical concepts indicates that progress has been made in understanding what responsible development and use of AI and big data involves.

**Table 2:** Normative Concepts

Individual Oriented	Society Oriented	System Oriented
Privacy	Justice	Transparency
Autonomy	Democracy	Explainability
Inclusivity/Accessibility	Social Good	Accountability

AI and data ethics statements and codes have coalesced around several ethical concepts. Some of these concepts are focused on the interests and rights of impacted individuals, e.g., protecting privacy, promoting autonomy, and ensuring accessibility. Others are focused on societal-level considerations, e.g., promoting justice, maintaining democratic institutions, and fostering the social good. Still others are focused on features of the technical systems, e.g., that how they work is sufficiently transparent or interpretable so that decisions can be explained and there is accountability for them.

However, articulating values, ethical concepts, and general principles is only the first step in addressing AI and data ethics challenges—and it is in many ways the easiest. It is relatively low cost and low commitment to develop and adopt these statements as aspirational documents.

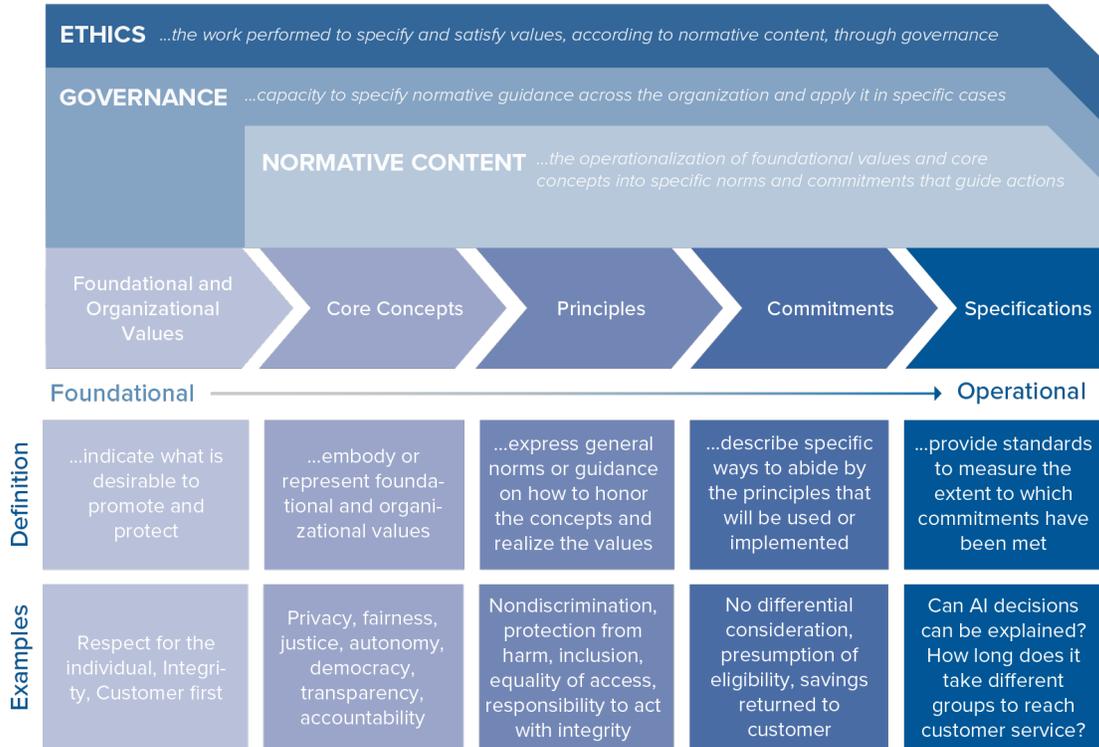
The **much harder** work is the following: (1) substantively specifying the **content** of the concepts, principles, and commitments—that is, clarifying what justice, privacy, autonomy, transparency, and explainability actually involve and require in particular contexts;<sup>1</sup> and (2) building professional, social, and organizational **capacity** to operationalize and realize

these in practice<sup>2</sup> (see Figure 1 for what these tasks encompass and how they are interconnected). Accomplishing these involves longitudinal effort and resources, as well as collaborative multidisciplinary expertise. New priorities and initiatives might be required, and existing organizational processes and structures might need to be revised. But these are the crucial steps in realizing the espoused values and commitments. Without them, any commitments or principles become mere “ethics washing.” Ethically problematic development and use of AI and big data will continue and industry will be seen by policy makers, employees, consumers, and the public as failing to make good on its own stated commitments.

1 Kate Crawford et al., AI Now 2019 Report, AI Now Institute, December 2019, [https://ainowinstitute.org/AI\\_Now\\_2019\\_Report.html](https://ainowinstitute.org/AI_Now_2019_Report.html), 20.

2 Ronald Sandler and John Basl, Building Data and AI Ethics Committees, Accenture, August 2019, [https://www.accenture.com/\\_acnmedia/PDF-107/Accenture-AI-And-Data-Ethics-Committee-Report-11.pdf](https://www.accenture.com/_acnmedia/PDF-107/Accenture-AI-And-Data-Ethics-Committee-Report-11.pdf).

**Figure 1:** Moving from Values to Actionable Commitments and Standards of Evaluation



The objectives of this report are to:

- demonstrate the importance and complexity of moving from general ethical concepts and principles to action-guiding substantive content, hereafter called **normative content**;
- provide detailed analysis of two widely discussed and interconnected ethical concepts, **justice** and **transparency**; and
- indicate strategies for moving from general ethical concepts and principles to more specific normative content and ultimately to **operationalizing** that content.

## Defining “Normative Content”

There are two uses of the term “norms.” One is descriptive, for which “norms” describe what is ordinary or typical to do. The other is ethical, for which “norms” prescribe what ought or should be done. Often, what ought or should be done (prescriptive) is different from what is currently being done (descriptive). The call for ethics for, and responsible development and use more generally of, AI and big data is a call for guidance on how it ought to be done because current practices are problematic in many respects. For example, justice-oriented considerations ought to be incorporated in ways that avoid perpetuating discriminatory practices and unjustified inequalities.

Therefore, as the term is used here, normative content refers to guidance on what “ought” or “should” be done. It is about developing well-justified standards, principles, and practices for what individuals, groups, and organizations should do, rather than merely describing what they currently do.

Normative content (in the ethical, prescriptive sense) is ultimately grounded on values, both social and organizational. This report is about the process of translating those general values into concrete and actionable guidance and commitments. This process is complex because specific principles and commitments need to be context sensitive, and because different values can sometimes come into tension with each other.

## II. EXAMPLE: INFORMED CONSENT IN BIOETHICS<sup>3</sup>

To illustrate the challenge of moving from a general ethical concept to action-guiding **normative content** it is helpful to reflect on an established case, such as informed consent in bioethics. Informed consent is widely recognized as a crucial component to ethical clinical practice and research in medicine. But where does the principle come from? What, exactly, does it mean? And what does accomplishing it involve?

Informed consent is taken to be a requirement of ethical practice in medicine and human subjects research because it protects individual autonomy.<sup>4</sup> Respecting people's autonomy means not manipulating them, deceiving them, or being overly paternalistic toward them. People have rights over themselves, and this includes choosing whether to participate in a research trial or undergo medical treatment. A requirement of informed consent is the dominant way of operationalizing respect for autonomy.

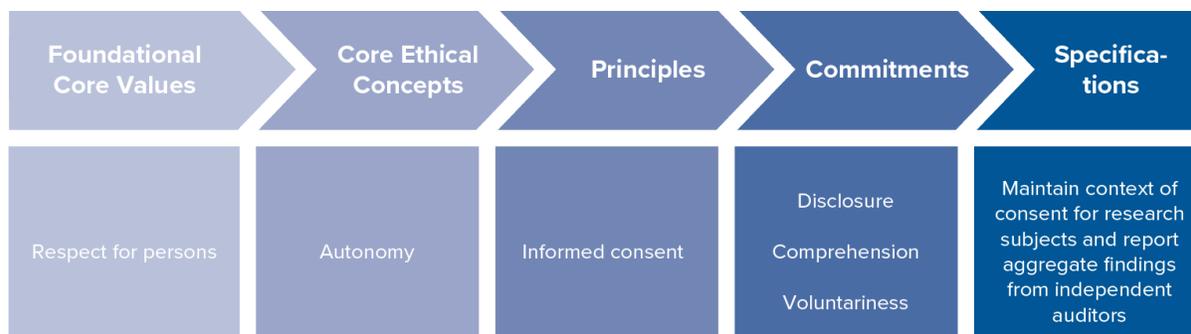
But what is required to satisfy the norm of informed consent **in practice**? When informed consent is explicated, it is taken to have three main conditions: disclosure, comprehension, and voluntariness. The disclosure condition is that patients, research subjects, or their proxies are provided clear, accurate, and relevant information about the situation and the decision they are being asked to make. The comprehension condition is that the information is presented to them in a way or

form that they can understand. The voluntariness condition is that they make the decision without undue influence or coercion.

But these three conditions must themselves be operationalized, which is the work of institutional review boards, hospital ethics committees, professional organizations, and bioethicists. They develop best practices, standards, and procedures for meeting the informed consent conditions—e.g., what information needs to be provided to potential research subjects, what level of understanding constitutes comprehension for patients, and how physicians can provide professional opinions without undue influence on decisions. Moreover, standards and best practices can and must be contextually sensitive. They cannot be the same in emergency medicine as they are in clinical practice, for example. It has been decades since the principle of informed consent was adopted in medicine and research, yet it remains under continuous refinement in response to new issues, concerns, and contexts to ensure that it protects individual autonomy.

So while informed consent is meant to protect the value of autonomy and express respect for persons, a general commitment to the principle of informed consent is just the beginning. The principle must be explicated and operationalized before it is meaningful and useful in practice. The same is true for principles of AI and data ethics.<sup>5</sup>

**Figure 2.** Example: Respect for persons in medical research (via IRB)



In clinical medicine and medical research, a foundational value is respect for the people involved—the patients and research subjects. But to know what that value requires in practice it must be clarified and operationalized. In medicine, respect for persons is understood in terms of autonomy of choice, which requires informed consent on the part of the patient or subject. Bioethicists, clinicians, and researchers have operationalized autonomous informed consent in terms of certain choice conditions: disclosure, comprehension, and voluntariness. These are then realized through a host of specific practices, such as disclosing information in a patient's primary language, maintaining context of consent for research subjects, and not obscuring information about risks. Thus, respect for persons is realized in medical contexts when these conditions are met through the specified practices for the context. It is therefore crucial to the ethics of medicine and medical research that these be implemented in practice in context-specific ways.

<sup>3</sup> This section is adapted from Ronald Sandler and John Basl, Building Data and AI Ethics Committees, Accenture, August 2019, [https://www.accenture.com/\\_acnmedia/PDF-107/Accenture-AI-And-Data-Ethics-Committee-Report-11.pdf](https://www.accenture.com/_acnmedia/PDF-107/Accenture-AI-And-Data-Ethics-Committee-Report-11.pdf), 13.

<sup>4</sup> National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, April 18, 1979, [https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c\\_FINAL.pdf](https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf).

The case of informed consent in bioethics has several lessons for understanding the challenge of moving from general ethical concepts to practical guidance in AI and data ethics:

- To get from a general ethical concept (e.g., justice, explanation, privacy) to practical guidance, it is first necessary to specify the normative content (i.e., specify the general principle and provide context-specific operationalization of it).
- Specifying the normative content often involves clarifying the foundational values it is meant to protect or promote.
- What a general principle requires in practice can differ significantly by context.
- It will often take collaborative expertise—technical, ethical, and context specific—to operationalize a general ethical concept or principle.
- Because novel, unexpected, contextual, and confounding considerations often arise, there need to be ongoing efforts, supported by organizational structures and practices, to monitor, assess, and improve operationalization of the principles and implementation in practice. It is not possible to operationalize them once and then move on.

**Figure 3.** The Process of Specifying Normative Content



AI and data ethics encompass the following: clarifying foundational values and core ethical concepts, specifying normative content (general principles and context-specific operationalization of them), and implementation in practice.

In what follows, the focus is on the complexities involved in moving from core concepts and principles to operationalization—the normative content—for two prominently discussed and interconnected AI and data ethics concepts: justice and transparency.

# III. WHAT IS JUSTICE IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING?

## Problems of Fairness and Justice in AI and Data Ethics

Fairness and justice are widely recognized as key components of AI and data ethics. One reason for this is that there already have been many high-profile cases where machine learning and automated decision systems have led to biased, discriminatory, or otherwise problematically unequal results. Here are just a few examples:

- AI recruiting and job advertising systems have been found to replicate existing racial and gender biases and stereotypes in the workforce<sup>1</sup>
- AI systems used to assign risk scores in medical contexts assigned lower risk scores to African-American patients as compared to white patients, resulting in their receiving less care or less urgent care<sup>2</sup>
- Facial recognition systems have been found to be racially biased<sup>3</sup>
- AI systems used in social service eligibility determinations are making it more difficult for people to access benefits for which they are qualified<sup>4</sup>
- Hate speech detection systems have been found to wrongly classify speech patterns associated with African-Americans as hate speech<sup>5</sup>
- A recidivism prediction system used in parole and bail setting overpredicted recidivism by African-Americans and underpredicted recidivism by whites<sup>6</sup>
- AI advertising systems have been found to racially differentiate with apartment listings<sup>7</sup>

- 1 Miranda Bogen, "All the Ways Hiring Algorithms Can Introduce Bias," *Harvard Business Review*, May 6, 2019, <https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias>.
- 2 Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (October 25, 2019): 447–53, <https://doi.org/10.1126/science.aax2342>.
- 3 Sophie Bushwick, "How NIST Tested Facial Recognition Algorithms for Racial Bias," *Scientific American*, December 27, 2019, <https://www.scientificamerican.com/article/how-nist-tested-facial-recognition-algorithms-for-racial-bias/>.
- 4 Ed Pilkington, "Digital Dystopia: How Algorithms Punish the Poor," *The Guardian*, October 14, 2019, <http://www.theguardian.com/technology/2019/oct/14/automating-poverty-algorithms-punish-poor>.
- 5 Charlotte Jee, "Google's Algorithm for Detecting Hate Speech Is Racially Biased," *MIT Technology Review*, August 13, 2019, <https://www.technologyreview.com/2019/08/13/133757/googles-algorithm-for-detecting-hate-speech-looks-racially-biased/>.
- 6 Julia Angwin Mattu, Jeff Larson, and Surya Lauren Kirchner, "Machine Bias," *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=f3FajB9MLCTEMdgLRTeO6gSsoYw-yhCA>.
- 7 Louise Matsakis, "Facebook's Ad System Might Be Hard-Coded for Discrimination," *Wired*, April 6, 2019, <https://www.wired.com/story/facebook-ad-system-discrimination/>.

There is consensus that justice is a crucial ethical consideration in AI and data ethics.<sup>5,6</sup> Machine learning (ML), data analytics, and AI more generally should, at a minimum, not exacerbate existing injustices or introduce new ones. Moreover, there is potential for AI/ML to reduce injustice by uncovering and correcting for (typically unintended) biases and unfairness in current decision-making systems.

## The Meanings of "Fairness" and "Justice"

"Fairness" and "justice" are often used interchangeably in the discourse around AI and data ethics, and that is how they are used here. The reason for this is to encompass the full range of considerations that fall under them (as indicated below). However, the terms also sometimes have more specific meanings. For example, among computer scientists who work on these issues, "fairness" is often used to refer to certain forms of parity among groups. So, for example, an algorithmic system would be "fair" if it had the same level of accuracy or false positive rates across protected groups. A common question raised is which, if any, of those forms of parity is the appropriate operationalization of fairness for that context (since it is often impossible to realize multiple forms of parity together). In institutional contexts, "fairness" is often used to refer to treatment in a particular context, whereas "justice" refers to the institutional features that structure the context—a process (e.g., lending or criminal justice) can be unfair because the system is unjust (e.g., organized in a way that advantages some groups and disadvantages others).

However, the concept of "justice" is complex, and can refer to different things in different contexts. To determine what justice in AI and data use requires in a particular context—for example, in deciding on loan applications, social service access, or healthcare prioritization—it is necessary to clarify the normative content and underlying values. Only then is it possible to specify what is required in specific cases, and in turn how (or to what extent) justice can be operationalized in technical and techno-social systems.

- 5 Cyrus Rafar, "Bias in AI: A Problem Recognized but Still Unresolved," *TechCrunch*, July 25, 2019, <https://techcrunch.com/2019/07/25/bias-in-ai-a-problem-recognized-but-still-unresolved/>
- 6 Reuben Binns et al., "'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, Association for Computing Machinery, 2018, <https://doi.org/10.1145/3173574.3173951>, 1–14.

### III. 1: THE VALUES AND FORMS OF JUSTICE

The foundational value that underlies justice is the equal worth and political standing of people. The domain of justice concerns how to organize social, political, economic, and other systems and structures in accordance with these values. A law, institution, process, or algorithm is unjust when it fails to embody these values. So, the most general principle of justice is that *all people should be equally respected and valued in*

*social, economic, and political systems and processes.*

However, there are many ways these core values and this very general principle of justice intersect with social structures and systems. As a result, there is a diverse set of more specific justice-oriented principles, examples of which are given in Table 3.

**Table 3.** Justice-Oriented Principles

PROCEDURAL	DISTRIBUTIVE	RECOGNITION
<p><b><i>Non-discrimination</i></b></p> <p>A system or decision-making process should not be biased against certain groups.</p>	<p><b><i>Equality of Access</i></b></p> <p>Everyone should be provided the same/similar access to benefits and services.</p>	<p><b><i>Same Treatment</i></b></p> <p>Everyone should be treated the same regardless of the groups to which they belong.</p>
<p><b><i>Equality of Opportunity</i></b></p> <p>Everyone should have equal/similar chances of success (e.g. educational, social, or economic).</p>	<p><b><i>Benefit Sharing</i></b></p> <p>Everyone who contributes to a collective endeavor should share in its benefits.</p>	<p><b><i>Representational Accuracy</i></b></p> <p>People or groups of people should not be mischaracterized.</p>
<p><b><i>Equality of Participation</i></b></p> <p>People should be similarly empowered in social and political decision-making contexts and processes.</p>	<p><b><i>Decent Outcome</i></b></p> <p>Everyone should have good enough (or minimally decent) social and economic outcomes.</p>	<p><b><i>Inclusion</i></b></p> <p>People or groups of people (and their interests) should not be marginalized or excluded.</p>
<p><b><i>Just Deserts</i></b></p> <p>People’s social and economic outcomes should reflect their efforts and contributions.</p>	<p><b><i>Prioritizing Worst-Off</i></b></p> <p>Practices and policies should prioritize those who are most vulnerable, dependent, or in need.</p>	<p><b><i>Reparative Justice</i></b></p> <p>Past wrongful harms should be made up for and addressed so as not to create further harms or future disadvantages.</p>

Each of these principles concerns social, political, or economic systems and structures. Some of the principles address the processes by which systems work (procedural justice). Others address the outcomes of the system (distributive justice). Still others address how people are represented within systems (recognition justice). There are many ways for things to be unjust, and many ways to promote justice.

Each of these justice-oriented principles is important in specific situations. Equality of opportunity is often crucial in

discussions about K-12 education. Equality of participation is important in political contexts. Prioritizing the worst off is important in some social services contexts. Just deserts is important in innovation contexts. And so on. The question to ask about these various justice-oriented principles when thinking about AI or big data is not which of them is the correct or most important aspect of justice. The questions are the following: *Which principles are appropriate in which contexts or situations? And what do they call for or require in those contexts or situations?*

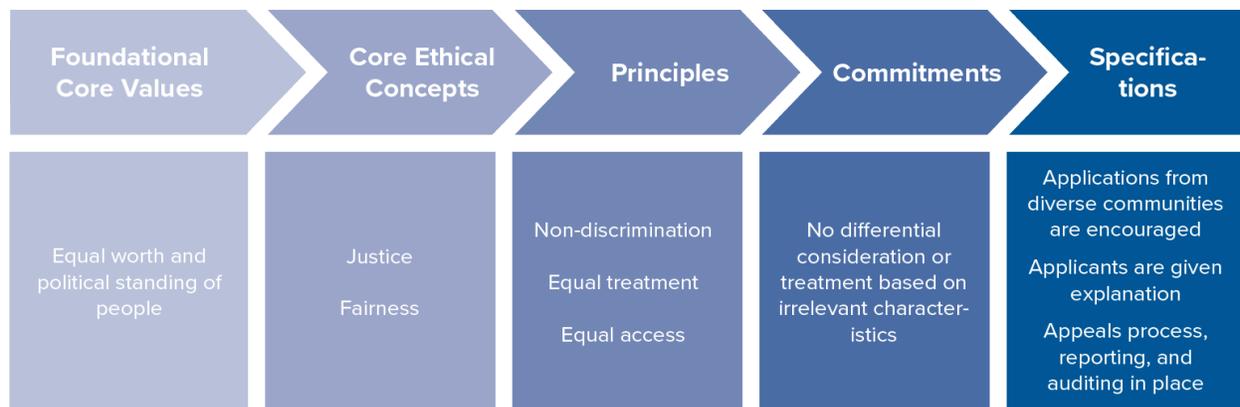
### III. 2: JUSTICE IN ARTIFICIAL INTELLIGENCE AND DATA SYSTEMS

What is an organization committing to when it commits to justice in AI and data practices? Context is critically important in determining which justice-oriented principles are operative or take precedence. Therefore, the first step in specifying the meaning of a commitment—i.e., specifying the normative content—is to identify the justice-oriented principles that are crucial to the work that the AI or data system does or the contexts in which they operate. Only then can a commitment to justice be effectively operationalized and put into practice.

Legal compliance will of course be relevant to identifying salient justice-oriented considerations—for example, statutes on nondiscrimination. But, as discussed above, legal compliance alone is insufficient.<sup>7</sup> Articulating the relevant justice-oriented principles will also require considering organizational missions, the types of products and services involved, how those products and services could impact communities and individuals, and organizational understanding of the causes of injustice within the space

in which they operate. In identifying these, it will often be helpful to reflect on similar cases and carefully consider the sorts of concerns that people have raised about AI, big data, and automated decision-making. Consider again some of the problem cases discussed earlier (see *Problems of Fairness and Justice in AI and Data Ethics*). Several of them, such as the sentencing and employment cases, raise issues related to equality of opportunity and equal treatment. Others, such as those concerning facial and speech recognition, involve representation and access. Still others, such as the healthcare cases, involve access and treatment as well as prioritization, restitutive, and benefit considerations. Here are two hypothetical and simplified cases—a private financial services company and a public social service organization—to illustrate this. They share the same foundational values (equal worth and political standing of people) and core ethical concepts (fairness and justice). However, they diverge in how these are specified, operationalized, and implemented due to what the algorithmic system does, its institutional context, and relevant social factors (see Figure 4 and Figure 5).

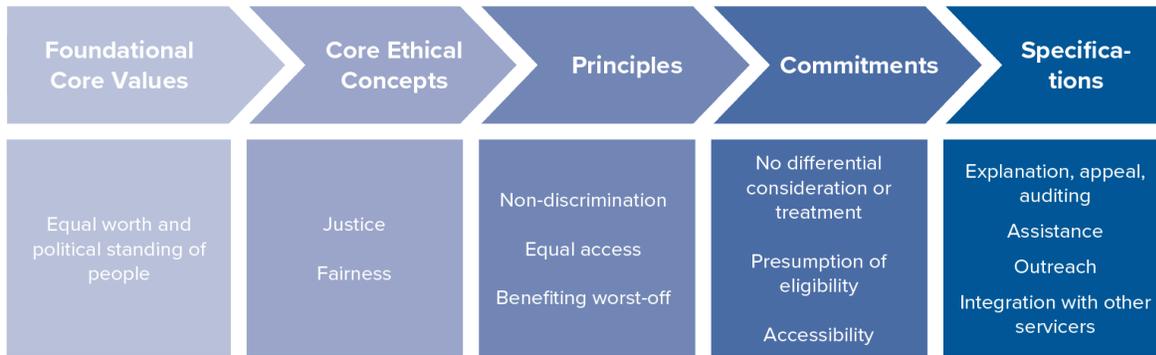
**Figure 4.** Example: Private Financial Services Company



Financial services firms trying to achieve justice and fairness must, at a minimum, be committed to nondiscrimination, equal treatment, and equal access. These cannot be accomplished if there is differential consideration, treatment, or access based on characteristics irrelevant to an individual's suitability for the service. To prevent this, a financial services firm might encourage applications from diverse communities, require explanations to applicants who are denied services, and participate in regular auditing and reporting processes, for example. Implementing these and other measures would promote justice and fairness in practice. To be clear, these may not be exhaustive of what justice requires for financial services firms. For instance, if there has been a prior history of unfair or discriminatory practices toward particular groups, then reparative justice might be a relevant justice-oriented principle as well. Or if a firm has a social mission to promote equality or social mobility, then benefiting the worst off might also be a relevant justice-oriented principle.

<sup>7</sup> Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review* 104 (2016): 671, <https://heinonline.org/HOL/LandingPage?handle=hein.journals/calr104&div=25&id=&page=>

**Figure 5.** Example: Public Social Service Organization



Social service organizations are publicly accountable and typically have well-defined social missions and responsibilities. These are relevant to determining which principles of justice are most salient for their work. It is often part of their mission that they provide benefits to certain groups. Because of this, justice and fairness for a social service organization may require not only nondiscrimination, but also ensuring access and benefitting the worst off. This is embodied in principles such as no differential consideration or treatment based on irrelevant characteristics, presumption of eligibility (i.e., access is a right rather than a privilege), and maximizing accessibility of services. Operationalizing these could involve requiring explanations for service denial decisions and having an available appeals/recourse process, creating provisions for routine auditing, providing assistance to help potential clients access benefits, conducting outreach to those who may be eligible for services, and integrating with other services that might benefit clients.

When there are multiple justice-oriented considerations that are relevant, it can frequently be necessary to balance them. For example, prioritizing the worst off or those who have been historically disadvantaged can sometimes involve moving away from same treatment. Or accomplishing equal opportunity and access can sometimes require compromises on just deserts. What is crucial is that all the relevant justice-oriented principles for the situation are identified; that the process of identification is inclusive of the communities most impacted; that all operative principles are considered as far as possible; and that any compromises are justified by the contextual importance of other justice-oriented considerations.

### Recidivism Predication and the Importance of Clarifying the Appropriate Conception of Justice

In the much-discussed case of the COMPAS recidivism prediction algorithm, there was a trade-off between equal accuracy among groups on the one hand and equal false positive and negative rates among groups on the other hand. The prediction system was roughly equally accurate with respect to recidivism predictions for whites and Blacks. However, it differed with respect to the kinds of mistakes that it made, when it made mistakes. It overpredicted recidivism by Blacks and underpredicted it by whites. Moreover, a plausible explanation for these differences across groups is the use of training data that embodies a history of overpolicing minority communities. So while there is a sense in which the algorithm accomplishes equal treatment on one parameter (i.e., equal accuracy in prediction), it fails to address the more justice-relevant treatment consideration in this context (i.e., overprediction of recidivism) and it does not take into account other relevant justice-oriented considerations, such as restitutive and representational considerations.<sup>1</sup>

<sup>1</sup> Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner. "Machine Bias," ProPublica, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

The diversity of justice-relevant principles and the need to make context-specific determinations about which are relevant and which to prioritize reveal the limits of a strictly algorithmic approach to incorporating justice into ML, data analytics, and autonomous decision-making systems.

First, *there is no singular, general justice-oriented constraint, optimization, or utility function; nor is there a set of hierarchically ordered ones.* Instead, context-specific expertise and ethical sensitivity are crucial to determining what justice requires.<sup>8</sup> A system that is designed and implemented in a justice-sensitive way for one context may not be fully justice sensitive in another context, since the aspects of justice that are most salient can be different.

Second, *there will often not be a strictly algorithmic way to fully incorporate justice into decision-making, even once the relevant justice considerations have been identified.*<sup>9</sup> For example, there can be data constraints, such as the necessary data might not be available (and it might not be the sort of data that could be feasibly or ethically collected). There can be technical constraints, such as the relevant types of justice considerations not being mathematically (or statistically) representable. There can be procedural constraints, such as justice-oriented considerations that require people to be responsible for decisions.

Therefore, it is crucial to ask the following question: *How and to what extent can the salient aspects of justice be achieved algorithmically?* In many cases, accomplishing justice in AI will require developing justice-informed techno-social (or human-algorithm) systems. AI systems can support social workers in service determinations, admissions officers in college admissions determinations, or healthcare professionals in diagnostic determinations—the systems might even be able to help reduce biases in those processes—while not being able to replicate or replace them. Moreover, it might require not using an AI or big data approach at all, even if it is more accurate or efficient. This could be for procedural, recognition, or relational justice reasons, or because it would be problematic to try to use or collect the necessary data. For example, justice might require a jury-by-peer criminal justice system even if it were possible to build a more accurate (in terms of guilt/innocent determinations) algorithmic system. A commitment to justice in AI involves remaining open to the possibility that sometimes an AI-oriented approach might not be a just one.

For these reasons, organizations that are committed to justice in AI and data analytics will require significant organizational capacity and processes to operationalize and implement the commitment, in addition to technical capacity and expertise. *Organizations cannot rely strictly on technical solutions or on standards developed in other contexts.*

## Justice in AI and Big Data: Key Takeaways

- Justice is grounded in the foundational value of equal worth and political standing of all people.
- AI and data analytics should not create or perpetuate injustice, and they should make every effort to promote justice.
- Specifying the normative content of justice in AI—i.e., what that commitment means—involves clearly articulating the justice-oriented considerations or principles that must inform automated decision systems.
- What justice requires is multifaceted and context specific.
- Which justice-oriented principles and considerations are most salient often depends upon an organization's mission, products/services, and the contexts in which it operates.
- Determining which justice-oriented principles and considerations are most salient therefore involves context-specific expertise—and crucially the perspectives of those who experience the injustice—AI/data ethics expertise, organizational governance, and technical and techno-social system expertise.
- There is usually not a strictly technical or algorithmic way to accomplish justice in AI and data analytics. It requires value sensitive design, monitoring, and auditing of both technical and techno-social (or human-machine) systems.
- Accomplishing justice in AI and data analytics is a substantial, collaborative, and ongoing process that requires developing organizational capacity.

8 Jon Kleinberg, "Inherent Trade-Offs in Algorithmic Fairness," In Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS '18, Association for Computing Machinery, June 2018, <https://doi.org/10.1145/3219617.3219634>.

9 Reuben Binns, "Fairness in Machine Learning: Lessons from Political Philosophy," in Conference on Fairness, Accountability, and Transparency, PMLR, 2018, <http://proceedings.mlr.press/v81/binns18a.html>, 149–59.

## IV. WHAT IS TRANSPARENCY IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING?

Assessing whether a decision-making system is just or fair and holding decision-makers accountable often requires being able to understand what the system is doing and how it is doing it. So, from the concepts of justice and accountability follow principles requiring transparency, for example, ensuring that algorithmic systems can be sufficiently understood by decision-subjects.

One feature of algorithmic decision-making that has been met with concern is that many of these systems are “black boxes”—how they make decisions is sometimes entirely opaque, even to those who build them. If an algorithmic decision-making system is a black box, then parties subject to its decisions will be in a poor position to assess whether they have been treated fairly and users of the system will be in a poor position to know whether it satisfies their commitment to

justice. Parties to hiring and firing decisions and parole decisions and those that use products, such as pharmaceuticals, that are discovered or designed using algorithms often have a significant interest in understanding how decisions are made or about how the products they use are designed.

Since black box algorithms can serve as an obstacle to realizing justice and other important values, concerns about opacity in algorithmic decision-making have been met with calls for *transparency* in algorithmic decision-making.<sup>10,11</sup> However, like justice, transparency has many meanings, and translating calls for transparency into guidance for the design and use of algorithmic decision-making systems requires clarifying why transparency is important in that context, to whom there is an obligation to be transparent, and what forms transparency might take to meet those obligations.<sup>12</sup>

### IV. 1: THE VALUES AND CONCEPTS UNDERLYING TRANSPARENCY

Transparent decision-making is important for and sometimes essential to realizing a variety of values and concepts. One such example, already mentioned, is justice. Opacity can sometimes prevent decision-subjects—e.g., the person who is seeking a loan, who is up for parole, who is applying for social services, or whose social media content has been removed—from determining whether they have in fact been treated respectfully and fairly. In many contexts, decision-makers owe decision-subjects due consideration. It is often important to hear others out, try to understand their perspective, and take that perspective seriously when making decisions. It is often disrespectful or shows lack of due consideration to make unilateral decisions that have significant impacts on others without taking the specific details of their circumstances into account. When a parole board decides whether to grant parole, for example, decision-subjects have a right to know that decision-makers have made a good faith effort to understand relevant details about their lives and take those details into account—i.e., that they have treated them as an individual.

In addition to the role that transparency plays in helping to achieve justice, it can also play an important role in realizing other concepts and values.

#### Compliance

Compliance is a normative concept associated with the value

of respecting the law. Opaque decision-making systems can make it difficult to determine whether a system or a model is compliant with applicable laws, regulations, codes of conduct, or other ethical norms. If a recidivism prediction system, targeted advertising system, or ranking system for dating apps were completely opaque and there was no way to tell whether it was using, for example, race or ethnicity (or proxies thereof) in its decision-making, it would be impossible to tell whether the system was in compliance with anti-discrimination law or requirements of justice. Thus, transparency is often justified or important because it makes it possible to determine whether it satisfies legal and ethical obligations.

#### Autonomy

Autonomy is a normative concept associated with the value of respecting persons. Enabling autonomy involves allowing people to make informed decisions, especially when those decisions make significant impacts on their life prospects. Opaque decision-making systems can make it difficult for decision-subjects to use decisions to guide future behavior or choices. If a decision-subject is rejected for a loan, job, school admission, or social service but has no access to how or why the decision was made, or what they can do to achieve a more satisfactory outcome, they will have a difficult time determining what they can do to achieve a favorable outcome in the future should they reapply. Thus, transparency is often

10 Cynthia Rudin, “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead,” *Nature Machine Intelligence* 1, no. 5 (May 2019): 206–15, <https://doi.org/10.1038/s42256-019-0048-x>.

11 Jonathan Zittrain, “The Hidden Costs of Automated Thinking,” *The New Yorker*, July 23, 2019, <https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking>.

12 Kathleen A. Creel, “Transparency in Complex Computational Systems,” *Philosophy of Science* 87, no. 4 (April 30, 2020): 568–89, <https://doi.org/10.1086/709729>.

valuable because it informs and empowers decision-making subjects to change behaviors in ways that impact future decisions about the trajectory of their lives.

### Knowledge

Knowledge is a normative concept grounded in the value of understanding. Knowledge enables us to better understand the world and our place in it. In some instances, algorithmic systems are used to uncover hidden patterns in large data sets. These patterns can help develop diagnostics or treatments even while knowledge of the patterns and/or of the underlying mechanisms that cause or explain those patterns remains opaque.<sup>13</sup> Knowledge in these areas could help practitioners not only intervene to realize better outcomes but generate new understanding, cures, and insights about

a domain.<sup>14</sup> Thus, transparency is often valuable because it enables future discoveries.

### Trust

Trust is a normative concept related to a variety of foundational values, including developing good relationships, and trust grounds principles of transparency because transparency can build trust. To fully realize the benefits of algorithmic tools, from medicine to route mapping, stakeholders must trust these systems and the ways they exercise power over their lives. Stakeholders that do not understand the how and why of such decisions are less likely to accept the use of such technologies and its decisions. Thus, transparency is often valuable because it can foster trust in decision-making systems.

## IV. 2: SPECIFYING TRANSPARENCY COMMITMENTS

Just as there are many concepts, including justice, that ground principles of transparency, there are many ways in which a decision-system could be made transparent, and so there are many forms that commitments to transparency might take, such as a commitment to interpretability of algorithmic decision-making systems, a commitment to the explainability of algorithmic decisions, or a commitment to accountability for such decisions. For example, a system can be transparent in the sense that individual decisions or behaviors of a system can be understood by technical experts or decision-subjects. For instance, a computer scientist might be able to explain how sensitive an algorithmic decision-making system used to set insurance rates is to factors such as age or driving experience. A broader form of transparency involves being transparent about why certain decision-making systems are in use, even if those decision-making systems are opaque. For instance, a decision-maker might show decision-subjects that the use of a largely opaque recommendation engine results in vastly more accurate recommendations with very low risk.

Here are some ways of understanding some common commitments to transparency:

### Interpretability

An often-discussed way to commit to transparency in algorithmic decision-making is to require these systems to be *interpretable*.<sup>15</sup> Some algorithmic systems are so complex—for example, because they rely on models built on such large data sets or involve many layers—that data and computer scientists themselves cannot determine in any precise way how system inputs are transformed to outputs. The interpretability of a system is relevant to many of the values above. For example, an uninterpretable system is a serious obstacle to discovering underlying patterns of inference the system

relies on and could make it impossible to ensure compliance with certain norms. However, not all AI and ML systems are uninterpretable, and computer and data scientists are working to develop tools to recover interpretability from black box systems.

### Explainability

Another kind of commitment to transparency is to require that algorithmic decision-making systems be *explainable*.<sup>16</sup> A decision-making system is explainable when it is possible to offer stakeholders an explanation that can be understood as justifying a given decision, behavior, or outcome. An interpretable system does not necessarily generate outputs that are explainable in this sense; knowing the inferences that underwrite a systems' decisions will not always serve to justify those decisions. Explainability is essential to realizing the values discussed above. Decision-subjects often need to know the reasons or considerations that led to the decisions about them to see if they have been treated with respect or for them to trust decision-makers. It is also often crucial for ensuring compliance with legal and ethical norms.

### Justified Opacity

Being transparent in the sense of ensuring interpretability or explainability can sometimes come at a cost to the accuracy of a system, i.e., there are sometimes trade-offs between accuracy on the one hand and interpretability or explainability on the other. In such cases, the trade-off might not be justified—i.e., in some contexts maximizing accuracy might be more important than ensuring that an algorithmic decision-making system is interpretable or explainable. Even so, there is a broader commitment to transparency that can still help realize core concepts such as trust. For example, patients might be willing to accept an uninterpretable diagnostic tool

13 Zittrain, "The Hidden Costs of Automated Thinking."

14 Rich Caruana et al., "Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-Day Readmission," in Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15: AC

15 Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead."

16 Hani Hagras, "Toward Human-Understandable, Explainable AI," *Computer* 51, no. 9 (September 2018): 28–36, <https://doi.org/10.1109/MC.2018.3620965>

if that tool is significantly more accurate than alternatives. Transparency about the reasons for adopting opaque systems can serve to justify other forms of opacity.

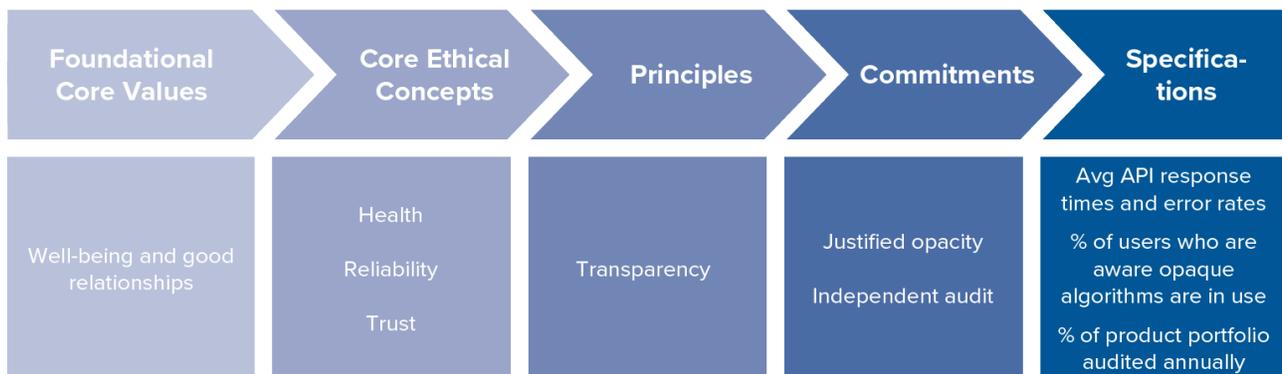
**Auditability**

A final kind of commitment to transparency is *auditability*.<sup>17</sup> It is not always feasible to provide transparency about all the individual decisions made by a system. Still, it is possible to evaluate the system by auditing some set of decisions periodically or based on certain observed patterns in decision-making to determine whether core values are realized. For example, while it might be infeasible to provide an explanation for every decision made by an autonomous vehicle system, it might be important for increasing trust that audits can reveal explanations for particular decisions. *One virtue*

*of auditing procedures is that they provide transparency not solely about algorithmic decision-making but about the broader systems in which they are embedded. A carefully constructed audit can provide assurance that decision-making systems broadly are trustworthy, reliable, and compliant.*

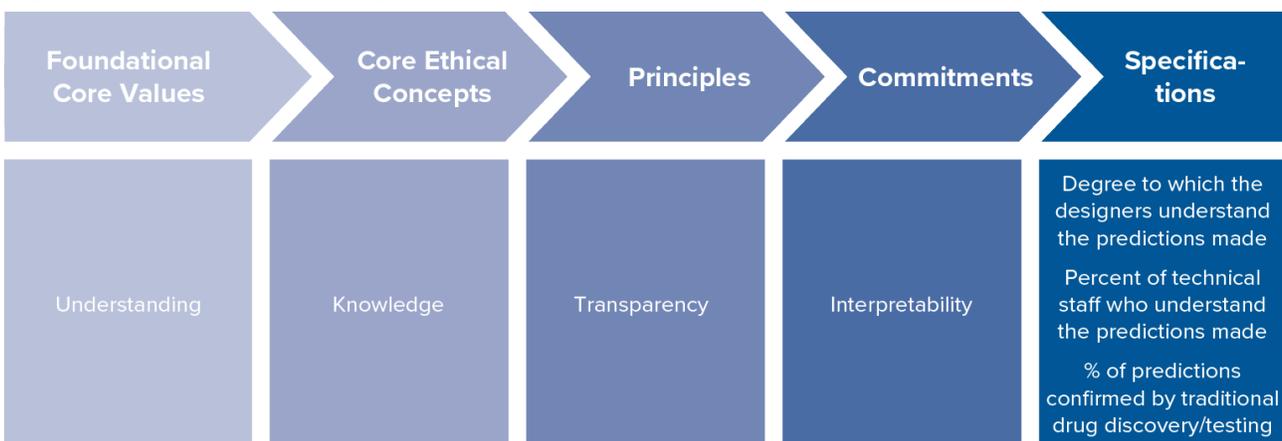
Moving from a principle of transparency to the appropriate commitment and ultimately to a measurable specification in a given context involves determining which core concepts are trying to be achieved in a decision-context with in that decision-context with a principle of transparency and which specific commitments to transparency best instantiate that principle. Increasing system interpretability might help minimize safety concerns but does little to increase trust or provide evidence of respect, for example.

**Figure 6.** Example: Low-Risk Health Diagnostic Tools



A low-risk health diagnostic tool is an algorithm that reliably provides a diagnosis for some conditions in patients but where failed diagnosis comes at low cost. In this case, the principle of transparency stems from concepts of health, reliability, and safety, which are grounded in the values of well-being and good relationships. Because transparency here is grounded in these core concepts, the transparency that is appropriate is a commitment to justifying the use of an opaque system in terms of its accuracy and a commitment to auditing the results to ensure that it remains accurate so as to demonstrate reliability and build trust.

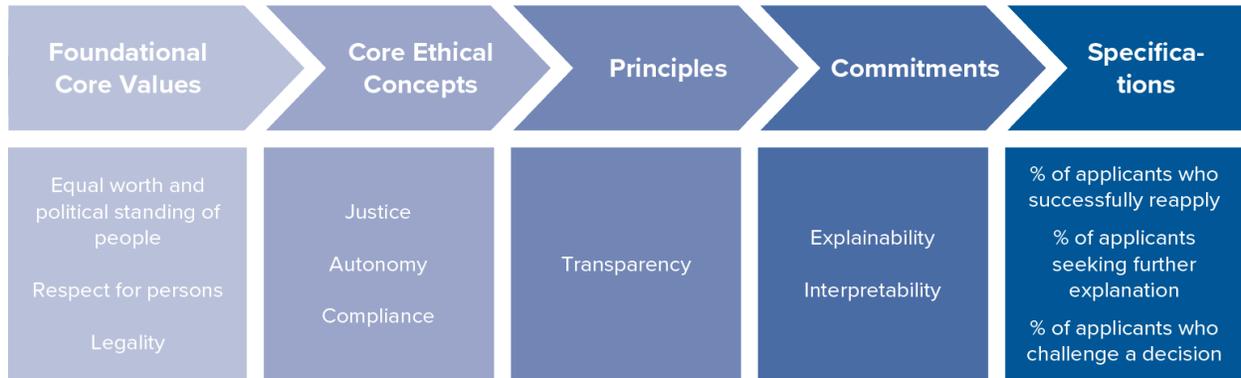
**Figure 7.** Example: Drug Discovery Algorithm



Drug discovery algorithms rely on the use of algorithmic tools to identify promising treatments for disease. In a case where the principle to be transparent is grounded in the core concept of knowledge, which stems from the value of understanding, the commitment that best embodies the principle might be interpretability since understanding the underlying structure by which decisions are made might yield knowledge.

<sup>17</sup> Bruno Lepri et al., "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes," *Philosophy & Technology* 31, no. 4 (December 1, 2018): 611–27, <https://doi.org/10.1007/s13347-017-0279-x>.

**Figure 8.** Example: Algorithmic Lending



In the case of a lending application, a principle of transparency is grounded, at least partly, in concepts of justice, autonomy, and compliance. These are important concepts because they help realize values of recognizing the equal worth of persons, respect for persons, and acting according to legitimate laws. Given the way the principle of transparency, in this case, relates to the core concepts and foundational values, it grounds commitments to both interpretability and explainability in some form. Explainability helps decision-subjects see that decisions about them were justified and that they have been respected, for example. Interpretability can ensure that algorithmic decision-making is compliant with the law. And jointly, explainability and interpretability can help ensure that various requirements of justice have been met.

### Transparency and Explainability: Key Takeaways

- Opacity is an obstacle to achieving a variety of values and core ethical concepts, including justice.
- Developing transparency standards that overcome these obstacles requires recognizing why transparency is valuable or important in a given context.
- Developing transparency standards also requires identifying which commitments to transparency—interpretability, explainability, justified opacity, or auditability—are necessary and sufficient for realizing relevant values and concepts in a given context.
- Satisfying transparency standards does not always require that everyone understands how every decision is made. In some contexts, such a high level of transparency is not needed and would compromise other important system values, such as efficiency and accuracy.

## V. CONCLUSION: FROM NORMATIVE CONTENT TO SPECIFIC COMMITMENTS

General commitments to ethics in AI and big data use are now commonplace. A first step in realizing them is adopting principles, value statements, frameworks, and codes that broadly articulate what responsible development and use encompass. However, if these concepts and principles are to have an impact—if they are to be more than aspirations and platitudes—then organizations must move from loose talk of values and commitments to clarifying how normative concepts embody foundational values, what principles embody those concepts, what specific commitments to those principles are for a particular use case, and ultimately to some specification of how they will evaluate whether they have lived up to their commitments. It is common to hear organizations claim to favor justice and transparency. The hard work is clarifying what a commitment to them means and then realizing them in practice.

This report has emphasized the complexities associated with moving from general commitments to substantive specifications in AI and data ethics. Much of this complexity arises from three key factors:

- Ethical concepts—such as justice and transparency—often have many senses and meanings.
- Which senses of ethical concepts are operative or appropriate is often contextual.
- Ethical concepts are multidimensional—e.g., in terms of what needs to be transparent, to whom, and in what form.

An implication of these factors is that specifying normative content in AI and data ethics is not itself a strictly technical problem. It often requires ethical expertise, knowledge of the relevant contextual (or domain-specific) factors, and expertise in human-machine interactions, in addition to knowledge of technical systems. Moreover, it is not possible to substantively specify a general standard of justice or transparency (or other ethical concept) that can then be applied in each instance of AI development and use.

A corollary to this is that for organizations to be successful in realizing their ethical commitments and accomplishing responsible AI, they must think broadly about how to build ethical capacity within and between their organizations. The field of AI and data ethics is still developing, but there already are initiatives at many organizations to do this, including the following:

### Organizational Structures and Governance

- Creating AI and data ethics committees that can help develop policies, consult on projects/products, and conduct reviews
- Establishing responsible innovation officer positions whose role includes translating ethical codes and principles into practice
- Developing tools, organizational practices, and structures/processes that encourage employees to identify and raise potential ethical concerns

### Broadening Perspectives and Collaborations

- Meaningfully engaging with impacted communities to better understand ethical concerns and the ways in which AI and data systems perpetuate inequality, as well as identifying opportunities for promoting equality, empowerment, and the social good
- Hiring in-house ethics and social science experts and/or consulting with external experts

### Training and Education

- Developing ethics trainings/curricula for employees, partners, and customers that raise awareness and understanding of the organization's commitments
- Creating programs that encourage integrating ethics education into computer science, data science, and leadership programs

## Integrating Ethics in Practice

- Incorporating value-sensitive design early into product development—often via responsible innovation practices—by including members who represent legal, ethical, and social perspectives on research and project teams
- Conducting “ethics audits” on data and AI systems with both technical tools and case reviews
- Conducting privacy- and ethics-oriented risk and liability assessments that are considered in decision-making processes
- Developing ways to signal to users and customers that products and services meet robust ethical standards
- Ensuring compliance of responsible innovation protocols and best practices throughout the organization

## Building an AI and Data Ethics Community

- Developing industry and cross-industry collaborative groups and fora to share experiences, ideas, knowledge, and innovations related to accomplishing responsible AI
- Creating teams within organizations and collaborations across organizations focused on promoting AI as a social good along well-defined specifications
- Supporting foundational AI and data ethics research on problems/issues of organizational concern

This collection of interconnected strategies is being innovated across sectors—e.g., business, medicine, education, financial services, government, nonprofits—and across scales—e.g., start-ups to multinationals and local to federal. Collectively, they provide a rich set of possibilities for generating ethical capacity within organizations. They demonstrate that the challenge of translating general commitments to substantive action is fundamentally a techno-social one that requires multi-level system solutions incorporating diverse groups and expertise. They also show that translating general AI ethics principles into substantive operationalized efforts is not easily done. It involves significant ongoing commitment.

## VI. CONTRIBUTORS

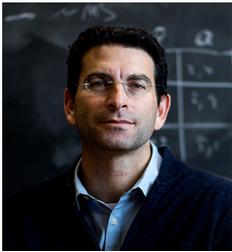
### AUTHORS



**John Basl, PhD**  
**Associate Professor of**  
**Philosophy**  
**Northeastern University**

John Basl is an author of this report. He is an associate professor of philosophy in the Department of Philosophy & Religion at

Northeastern University, a researcher at the Northeastern Ethics Institute, and a faculty associate of both the Edmond J. Safra Center for Ethics and the Berkman Klein Center for Internet and Society at Harvard University. He works primarily in moral philosophy, the ethics of AI, and the ethics of emerging technologies.



**Ronald Sandler, PhD**  
**Professor of Philosophy; Chair,**  
**Department of Philosophy**  
**and Religion; Director, Ethics**  
**Institute**  
**Northeastern University**

Ronald Sandler is an author of this report. He is a professor of philos-

ophy, Chair of the Department of Philosophy and Religion, and Director of the Ethics Institute at Northeastern University. He works in ethics of emerging technologies, environmental ethics, and moral philosophy. Sandler is author of *Nanotechnology: The Social and Ethical Issues* (Woodrow Wilson International Center for Scholars), *Environmental Ethics: Theory in Practice* (Oxford), *Food Ethics: The Basics* (Routledge), and *Character and Environment* (Columbia), as well as editor of *Ethics and Emerging Technologies* (Palgrave MacMillan), *Environmental Justice and Environmentalism* (MIT), and *Environmental Virtue Ethics* (Rowman and Littlefield).

### CONTRIBUTING EDITOR



**Steven Tiell**  
**Nonresident Senior Fellow,**  
**GeoTech Center**  
**Atlantic Council**

Steven is a contributing editor to this report. He is a Nonresident Senior Fellow with the Atlantic Council's GeoTech Center. He is an expert in

data ethics and responsible innovation working at Accenture, where he helps clients to integrate responsible product development practices and executives to manage risks brought on by digital transformation and widespread use of artificial intelligence. He founded the Data Salon Series, now a program at the GeoTech Center, in 2018. Since embarking on Data Ethics research in 2013, Steven has contributed to and published more than a dozen papers and has worked with dozens of organizations in high-tech, media, telecom, financial services, public safety, public policy, government, and defense sectors. He often speaks on topics such as governance, trust, data ethics, surveillance, deepfakes, and industry trends.



### **CHAIRMAN**

\*John F.W. Rogers

### **EXECUTIVE CHAIRMAN EMERITUS**

\*James L. Jones

### **PRESIDENT AND CEO**

\*Frederick Kempe

### **EXECUTIVE VICE CHAIRS**

\*Adrienne Arsht

\*Stephen J. Hadley

### **VICE CHAIRS**

\*Robert J. Abernethy

\*Richard W. Edelman

\*C. Boyden Gray

\*Alexander V. Mirtchev

\*John J. Studzinski

### **TREASURER**

\*George Lund

### **DIRECTORS**

Stéphane Abrial

Todd Achilles

\*Peter Ackerman

Timothy D. Adams

\*Michael Andersson

David D. Aufhauser

Barbara Barrett

Colleen Bell

Stephen Biegun

\*Rafic A. Bizri

\*Linden P. Blue

Adam Boehler

Philip M. Breedlove

Myron Brilliant

\*Esther Brimmer

R. Nicholas Burns

\*Richard R. Burt

Teresa Carlson

James E. Cartwright

John E. Chapoton

Ahmed Charai

Melanie Chen

Michael Chertoff

\*George Chopivsky

Wesley K. Clark

Beth Connaughty

\*Helima Croft

Ralph D. Crosby, Jr.

\*Ankit N. Desai

Dario Deste

\*Paula J. Dobriansky

Joseph F. Dunford, Jr.

Thomas J. Egan, Jr.

Stuart E. Eizenstat

Thomas R. Eldridge

Mark T. Esper

\*Alan H. Fleischmann

Jendayi E. Frazer

Courtney Geduldig

Meg Gentle

Thomas H. Glocer

John B. Goodman

\*Sherri W. Goodman

Murathan Günal

Amir A. Handjani

Frank Haun

Michael V. Hayden

Amos Hochstein

Tim Holt

\*Karl V. Hopkins

Andrew Hove

Mary L. Howell

Ian Ihnatowycz

Wolfgang F. Ischinger

Deborah Lee James

Joia M. Johnson

\*Maria Pica Karp

Andre Kelleners

Henry A. Kissinger

\*C. Jeffrey Knittel

Franklin D. Kramer

Laura Lane

Jan M. Lodal

Douglas Lute

Jane Holl Lute

William J. Lynn

Mark Machin

Mian M. Mansha

Marco Margheri

Michael Margolis

Chris Marlin

William Marron

Gerardo Mato

Timothy McBride

Erin McGrain

John M. McHugh

Eric D.K. Melby

\*Judith A. Miller

Dariusz Mioduski

\*Michael J. Morell

\*Richard Morningstar

Georgette Mosbacher

Dambisa F. Moyo

Virginia A. Mulberger

Mary Claire Murphy

Edward J. Newberry

Thomas R. Nides

Franco Nuschese

Joseph S. Nye

Ahmet M. Ören

Sally A. Painter

Ana I. Palacio

\*Kostas Pantazopoulos

Alan Pellegrini

David H. Petraeus

W. DeVier Pierson

Lisa Pollina

Daniel B. Poneman

\*Dina H. Powell

McCormick

Ashraf Qazi

Robert Rangel

Thomas J. Ridge

Gary Rieschel

Lawrence Di Rita

Michael J. Rogers

Charles O. Rossotti

Harry Sachinis

C. Michael Scaparrotti

Ivan A. Schlager

Rajiv Shah

Kris Singh

Walter Slocombe

Christopher Smith

Clifford M. Sobel

James G. Stavridis

Michael S. Steele

Richard J.A. Steele

Mary Streett

\*Frances M. Townsend

Clyde C. Tuggle

Melanne Verveer

Charles F. Wald

Michael F. Walsh

Ronald Weiser

Olin Wethington

Maciej Witucki

Neal S. Wolin

\*Jenny Wood

Guang Yang

Mary C. Yates

Dov S. Zakheim

### **HONORARY DIRECTORS**

James A. Baker, III

Ashton B. Carter

Robert M. Gates

James N. Mattis

Michael G. Mullen

Leon E. Panetta

William J. Perry

Colin L. Powell

Condoleezza Rice

Horst Teltschik

William H. Webster



This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



The Atlantic Council is a nonpartisan organization that promotes constructive US leadership and engagement in international affairs based on the central role of the Atlantic community in meeting today's global challenges.

© 2021 The Atlantic Council of the United States. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means without permission in writing from the Atlantic Council, except in the case of brief quotations in news articles, critical articles, or reviews. Please direct inquiries to:

Atlantic Council  
1030 15th Street, NW, 12th Floor,  
Washington, DC 20005  
(202) 463-7226, [www.AtlanticCouncil.org](http://www.AtlanticCouncil.org)

