NSF-funded Fairness, Ethics, Accountability, and Transparency (FEAT) Workshop Report

A report based on an NSF-funded workshop hosted by the Georgia Institute of Technology and Morehouse College

August 29 and 30, 2019

NSF-funded Fairness, Ethics, Accountability, and Transparency (FEAT) Workshop Report

Workshop Organizers
Ayanna Howard, Georgia Institute of Technology
Jason Borenstein, Georgia Institute of Technology
Kinnis Gosha, Morehouse College

October 2019

This report is based upon work supported by the National Science Foundation Division of Information & Intelligent Systems under Grant No. 1903909. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the workshop organizers and participants, and do not necessarily reflect the views of the National Science Foundation.

Executive Summary

Modern societies rely extensively on computing technologies. As such, there is a need to identify and develop strategies for addressing fairness, ethics, accountability, and transparency (FEAT) in computing-based research, practice, and educational efforts. To achieve this aim, a workshop, funded by the National Science Foundation, convened a working group of experts to document best practices and integrate disparate approaches to FEAT. The working group included different disciplines, demographics, and institutional types, including large research-intensive universities, Historically Black Colleges and Universities, Hispanic-Serving Institutions, teaching institutions, and liberal arts colleges. The workshop brought academics and members of industry together along with government representatives, which is vitally important given the role and impact that each sector can have on the future of computing. Relevant insights were gained by drawing on the experience of policy scholars, lawyers, statisticians, sociologists, and philosophers along with the more traditional sources of expertise in the computing realm (such as computer scientists and engineers). The working group examined best practices and sought to articulate strategies for addressing FEAT in computing-based research and education. This included identifying methodological approaches that researchers could employ to facilitate FEAT, instituting guidelines on what problem definition practices work best, and highlighting best practices for data access and data inclusion. The resulting report is the culmination of the working group activities in identifying systematic methods and effective approaches to incorporate FEAT considerations into the design and implementation of computing artifacts.

1. Introduction

Given how much modern societies have come to rely on computing technologies, including artificial intelligence (AI) and robotic applications, there is an urgent need to identify and develop guidelines and strategies for ensuring fairness, ethics, accountability, and transparency (FEAT) in computing-based research, practice, and educational efforts. As a starting point for addressing this need, a workshop funded by the National Science Foundation was hosted on the Georgia Tech campus on August 29 and August 30, 2019. The workshop was organized by PI Ayanna Howard, School of Interactive Computing and Co-PI Jason Borenstein, School of Public Policy and Office of Graduate Studies at the Georgia Institute of Technology, and Co-PI Kinnis Gosha, Division of Experiential Learning and Interdisciplinary Studies at Morehouse College.

The workshop convened a working group of experts across the four FEAT topical realms (fairness, ethics, accountability, and transparency). The charge of the group was to identify FEAT-related challenges, examine best practices, and seek to articulate strategies for ensuring FEAT in computing-based research and education. The group was diverse in various ways, including in terms of disciplinary makeup, the types of employers represented, and the demographics of the individual participants (see **Appendix A**). The aim was to bring together a group that possesses extensive expertise in areas such as artificial intelligence, broadening participation in computing, computing education, human-computer interaction, ethics in STEM, psychology, cognitive science, algorithm design, data science, statistics, technology design, human factors, philosophy, law, and public policy. The group contained a range of experts from a diverse collection of academic institutions, including from large research-intensive universities, Historically Black Colleges and Universities, Hispanic-Serving Institutions, teaching institutions, and liberal arts colleges.

The focus on the FEAT realm is crucial because of how computing is reshaping human life, and in some cases, contributing to significant harms or at least, not promoting human well-being as fully as it should.

2. The Process for Gathering Challenges and Best Practices from Workshop Experts

Two main strategies were implemented during the FEAT workshop to draw on the expertise of the working group participants (see **Appendix B**). The first strategy was to organize a series of four panels during the first day of the workshop; each panel focused on one of the FEAT topic areas. The panelists and panel moderators were supplied with a list of discussion questions in advance (see **Appendix C-F**) that was used as a jumping off point for each panel. The question list helped to facilitate discussion among the panelists and with other working group participants.

The second strategy was to host breakout group sessions associated with each of the FEAT topics. The goal of each breakout group was to discuss a list of guiding questions; these sets of questions were different from the ones provided to the panels. The breakout session questions (see **Appendix G**) served as a foundation for a 15-20 minute presentation from each breakout group during the second day of the workshop. The combination of information obtained from the panels and the breakout groups served as a foundation for a draft report. The draft report was then disseminated to the working group experts, and through iterative feedback, the final report was constructed for review and approval by the working group over a three-week time-period.

3. Defining Fairness, Ethics, Accountability, and Transparency

In order to identify challenges and best practices in relation to fairness, ethics, accountability, and transparency in computing-based research, practice, and education, a starting point was to seek definitions for the relevant terminologies. A summary of the working group's discussions about those definitions is provided below.

3.1 Defining Fairness

Fairness is a concept/principle that is difficult to define. Yet what may help shed light on the term is distinguishing between a fair result and a fair process. A **fair result** typically refers to an equitable distribution of goods that are required for human flourishing, where goods could include rights, resources, opportunities, and/or capabilities. The overarching aim is normally to seek an *equitable* distribution not just at a given point in time but also consistently over the long term. As such, what counts as equitable should be re-evaluated as a society's knowledge and values change. A key goal of those who seek to promote fair results is to redress historical inequities and to prevent such inequities from occurring in the future so that all parties have a meaningful opportunity to flourish. An example of this notion of fairness in the computing realm is whether each individual has a meaningful opportunity to attend a computing degree program and enter the computing workforce.

A **fair process** is one that has mechanisms or procedures which entail that similar situations are handled in similar ways. In other words, fairness can refer to a process that is consistently applied. A process is more likely to be fair if it is inclusive and representative. For example, if a politician is going to represent a district, then a fair process would entail having individuals from different genders, races, and religious backgrounds vote on the politician's candidacy with each person's vote counting the same. An example of this dimension of fairness in computing is whether each individual's employment application for a job is evaluated through a process that treats comparable candidates similarly. It is important to note that tension can emerge between the pursuit of the two main notions of fairness described in this section. For instance, consistently applying admission standards for a computing program (i.e., a fair process) might not necessarily remedy historical or current injustices (i.e., a fair result).

Concerns about the lack of fairness in terms of an outcome or process is often intertwined with bias. Not all types of bias are ethically problematic; in fact, many types of bias might bestow adaptive advantages. Yet bias can lead to arbitrary or unethical behavior. In the computing realm, there are many reasons to be concerned about unfair outcomes and processes, some of which are tied to algorithmic or other forms of bias; these issues will be discussed in more detail in section 4.1.

3.2 Defining Ethics

Ethics has many different definitions. The term is often used interchangeably with "morals"; yet many scholars distinguish between the two concepts. In academic circles, ethics often refers to the branch of philosophy dedicated to describing and analyzing the rightness or wrongness (or good and badness) of human behavior. Scholars in ethics often seek to articulate, recommend, and defend concepts of right and wrong behavior.

There are many different types of ethics in the realm of philosophy including:

- *Metaethics* - investigates where ethical concepts and principles come from and what they mean
- *Descriptive ethics* - seeks to describe how humans behave in situations that have ethical dimensions
- *Normative ethics* - seeks to provide standards that govern right and wrong conduct
- *Applied ethics* - involves studying the relationship of ethical concepts and principles to specific issues such as capital punishment, free speech, and the allocation of health care

Professional ethics is another related realm; it often seeks to identify the shared values, norms, or principles that guide the conduct of individuals and groups who are part of a profession.

3.3 Defining Accountability

Accountability is a mechanism through which principles of fairness, ethics, and transparency are enforced. The concept is tied to being able to hold an individual or organization answerable for an action or the failure to act. Without it, risks increase and social benefits, such as upholding justice, might not be realized. Consequently, instillation and reinforcement of accountability has to be a sustained process. As with the other terms, accountability can be defined in many ways. The term often has both a moral and a legal sense; these two domains might overlap but are not necessarily the same thing. For example, someone might be accountable for lying to a friend, but that practice is not necessarily illegal. Accountability can also occur at different levels of granularity; for instance, an individual could be accountable for causing harm, or the responsible entity in question could be a group of individuals or an organization.

To shed further light on the concept, the following list of accountability dimensions may be helpful:

- Who is accountable?
    - Individuals (e.g., educators, researchers, and designers)
    - Organizations
    - Governments

- To whom are they accountable?
    - Colleagues
    - Clients or customers
    - Public (*everyone* is accountable to the public)

- What are they accountable for?
    - Upholding professional standards
    - Impacts on human rights and human well-being
    - Education and public awareness
    - Other social and ethical consequences (e.g., environmental impacts or economic inequality)

The above is not intended to cover exhaustively the various facets of accountability. Yet it highlights the point that to discuss accountability, we need to know who the actors are (individuals or groups), who has been or might be impacted by the actions, and what the (moral and/or legal) expectations are for the actors.

3.4 Defining Transparency

In daily life, transparency usually means being able to see through something such as a pane of glass. In academic and professional contexts, transparency often refers to whether something has been explained in such a way that one can understand how it operates or functions. Building on the latter definition, transparency in the computing realm typically entails whether information is clear and sufficient enough so that those who interact with the computing technology can, at least in principle, understand how the technology works and how decisions are made. At least some dimensions of transparency overlap with concepts such as explainability, interpretability, and intelligibility.

Several facets of transparency relevant to computing include:

- Are users and others aware of the existence of a computing system?
- Are they aware of what the computing system is doing?
- Are they aware of how the computing system affects them or others?

Transparency from computing companies, for example, does not just involve providing information concerning what type of data will be collected but it also involves providing information on how that data will be shared and used.

Transparency is also interconnected with the concept of traceability; in other words, is it possible to trace where an automated decision came from and which factors impacted that decision?

3.5 Discussion and Summary Remarks

Each of the relevant FEAT terms is notoriously difficult to define but even with that being the case, the working group sought to highlight facets of these terms that are important to address. The next section discusses the working group's insights regarding how FEAT is more specifically connected to the computing realm.

4. Application Areas of FEAT and Associated Challenges

4.1 Fairness in Computing and Associated Challenges

Fairness concerns are becoming pervasive in computing-based research, practice, and educational efforts. Among the issues to address include algorithmic (un)fairness; fairness in data collection, data storage, data access, and data analysis; fairness in terms of how to translate models, and where the biases in models and datasets originate. One recent example involves algorithms that are purportedly recommending disproportionately harsher prison sentences for minorities and cutting off people from lower-economic communities from health care benefits (Eubanks 2018). Part of the problem derives from the quality of the datasets used to inform the algorithms and whether the data accurately resemble reality (O'Neil 2016). Unless suitable remedies are implemented, computing may amplify and reinforce existing patterns of inequality and bias (Howard and Borenstein 2018). One source of the problem is connected to the research design, and the data collection and analysis methods being used. A benchmark strategy is typically connected to the accuracy of the model on a specific dataset, but it is not necessarily representative of the accuracy of the model on the relevant population as a whole. Thus, algorithms may perpetuate and potentially amplify biases in the training data set.

At the intersection of fairness and research, the working group highlighted the following considerations:

- Fairness for those doing the research (e.g., funding)
- Fairness for those impacted by research (e.g., using participatory action research methods)
- Fairness in/of the research
- Fairness in the choice of research questions
- Whether research methods are likely to promote fairness
- Whether the data are likely to lead to fair outcomes

The working group noted that if fairness concerns in research or other practices are not addressed properly, the following risks to the public may occur:

- Wealth differences or other disparities may be amplified
- Discrimination in workplace hiring, healthcare, housing, education, and other sectors may continue
- Automation might eradicate at least some job sectors
- Loss of global competitiveness may occur
- Historically disadvantaged groups will continue to experience significant discrimination

On the other hand, some might argue that computing algorithms might be "fairer" than humans in some contexts because of the potential to remove at least some of the subjectivity that pervades human decision-making. A person's shift in mood can certainly have an influence on decisions. For example, studies indicate that judges change their sentencing determinations depending on when they had their last meal (Danziger et al. 2011) or whether their favorite college football team won or lost (Deruy 2016). Thus, computing technology could help to uphold fairness in some contexts, but the working group noted that the ethical and legal ramifications of handing over decision-making authority to automated systems certainly needs to be addressed. Just because a computing device could possibly make a "fairer" decision, it does not necessarily follow that it should do so instead of a human being.

Moreover, the individuals and organizations that create an algorithm often lack diversity and may not adequately represent the perspectives of the population that will use or be affected by the algorithm. In such cases, the resulting algorithm may have negative biases embedded within it. In the realm of AI, this challenge has been called *AI's White Boy Problem* (Crawford 2016). The typical AI practitioner is a 40-year-old white male making approximately $92K per year (DataUSA 2018). In 2016, the National Science and Technology Council Committee on Technology published a report on the status of AI in the U.S., and in the report, it emphasized the need for racial and ethnic diversity in AI. In 2016, the AI Now Institute at New York University, an interdisciplinary research center dedicated to understanding the social implications of AI, made a similar recommendation (Crawford et al. 2016). The persistence of these diversity issues brings into question fairness in and access to computing and information science education. Overcoming current obstacles will require substantial improvements in those spheres.

At the intersection of fairness and computing education, the working group reiterated that significant problems must be overcome in terms of the lack of diversity in computing programs at academic institutions. For fairness in education, we need to reflect on:

- Who has access to the education?
- What does the education or curriculum consist of?
- How is the information delivered, for example – in-person or on-line or using a hybrid model?

Not only do significant barriers to entry for students from historically underrepresented groups remain (National Academies 2019; NSF 2019) but a relative lack of faculty from such groups persists as well (Clauset et al. 2105; Moody 2004). Failure to emphasize diversity efforts in computing education is certainly an ongoing problem. The group strongly asserted that diversity is a driver of innovation. Intertwined with the above concerns is the role of industry and its associated practices. Whether computing companies are doing enough to promote fairness is certainly open for debate. Arguably, the computing industry's current efforts to hire minority employees, including African Americans (Vara 2016), are insufficient.

4.2 Ethics in Computing and Associated Challenges

Many ethical issues are emerging in the computing realm. The (deliberate) spread of misinformation is a key ethical concern that warrants the attention of the computing community (Del Vicario et al. 2016). An issue that draws significant attention from the public is privacy. A pervasive sense of unease continues about whether we can maintain any semblance of privacy while online, especially when using social media websites (Shankland 2018). Something that seems relatively innocuous now (e.g., sharing pictures on a social media app) can become problematic later (e.g., losing control over the information). The recent bug with the Facetime app, which allowed users to hear a person that they were calling before the person answered the call, illustrates how it can be difficult to anticipate privacy violations (May 2019).

Moreover, the working group noted that diminished privacy can also erode the autonomy of those who interact with computing systems. The working group mentioned the need for fuller discussions on “opt-in” versus “opt-out” design of computing technology. For instance, hidden within legalistic terms of use for an app are a company’s data collection and sharing practices. The default is usually to collect the data unless the user opts out, but the user might not have been fully aware of the option or appreciate what the company’s data use practices mean.

The group also pointed out the tension between privacy and accuracy; efforts to improve data quality (e.g., by having a more robust and representative training data set) can be in conflict with privacy. For instance, if underrepresented groups seek to preserve their privacy, their data might not be included in a data set, which as a byproduct could intensify the bias in the data. In such instances, principles of fairness in algorithmic design might be in tension with the principle of preserving privacy. The computing community needs strategies that uphold privacy, especially for vulnerable groups, in conjunction with approaches that improve data accuracy.

Another thread that emerged within the working group pertains to social justice. Vast segments of society might be falling behind in terms of their digital literacy. What if parents do not own a computer to check on their children’s progress in school? Parents, along with teachers, at relatively poor schools can struggle to keep up with computing technology (Herold 2017). Lower income individuals or individuals who are isolated could significantly benefit from social media and similar technology if they can be connected. Usability of computing technology can be another problem for many segments of the population--for example, older adults who have difficulty navigating smart devices due to either a lack of digital proficiency or accessibility of interfaces.

With regard to research, it is important to develop ethical (accessible, equitable, and sustainable) research practices, especially when studying and engaging with marginalized communities. The role of ethics review boards, including IRBs, needs to be revisited, including whether such boards have the relevant expertise to review computing research protocols involving human subjects. Arguably, the Facebook emotional contagion study, for example, should have been more thoroughly reviewed (Kramer et al. 2014). Moreover, these boards need to address the tension between protecting people and obtaining an adequate and representative sample from the population.

From the perspective of computing education, the working group indicated that the methods for introducing ethics into the curriculum and the content being covered need to be revisited. For example, should ethics be embedded throughout the curriculum or should it just be included in a single required course? In terms of an educational approach, one strategy embraced by many universities is the hackathon. But the hackathon culture itself does not always have lessons about the ethics embedded within it. Content needs to be taught in a way so that students understand why ethics matters in general and why it specifically matters to the computing field. Furthermore, ethics needs to be embedded in the design process not only in academic circles but in industry as well.

Research ethics/RCR is also an important part of this puzzle. The lack of reproducibility, for example, is a crucial topic given the temptation to manipulate the process and outcomes of research to achieve a desired result. Another issue is who should teach ethics for computing students. It is a challenge to find people in different disciplines who are willing to work together (people in ethics collaborating with computer scientists for example) because incentive systems do not necessarily reward such collaborations.

Industry has been enmeshed in many ethical lapses such as the FaceTime privacy bug, and in some cases, more egregious and deliberate wrongdoing such as the Volkswagen's "Dieselgate" where automobile software was deliberately designed to fool emission tests (Patel 2015). Moreover, the working group argued that industry must re-examine its ethical responsibilities for the harms caused by its products, such as the deaths related to the use of electric scooters (Bussewitz 2019) and autonomous vehicles (Schmelzer 2019). More leadership from industry is needed on the topic of ethics. On a related note, the working group indicated that government and regulators are a key piece of the puzzle. External review of computing artifacts might be necessary. This is in part due to the inherent conflicts of interest that companies and other entities have; for example, the drive to be the first to market can conflict with the obligation to protect the public. Companies might indicate that they are putting forward initiatives to support ethics, but it must not be "ethics washing" or an attempt at merely being seen as ethical (Johnson 2019).

The "problem of many hands", a topic commonly discussed in engineering ethics (Harris et al. 2009), rears its head in relation to computing in the sense that many individuals can be involved in the creation of a complex computing artifact; this can obscure who is responsible when harm occurs, and thus can exacerbate accountability concerns. For example, when social networking websites are being used in ways that are not fully anticipated by users, it is not always clear how (or to whom) to assign blame. A related issue is that even though a machine might be able to make a "better" decision in some contexts, should it be allowed to do so when it involves human well-being (e.g., whether someone will remain in jail or be set free)?

4.3 Accountability in Computing and Associated Challenges

The pervasiveness of computing systems in society means that when accountability is absent, it impacts everyone. Yet accountability is only meaningful for those who are in a position to do something, and many of those who interact with computing technology may not have the power to hold someone accountable. A lack of accountability for bad outcomes, whether intended or not, can severely undermine public trust in computing technology and computing communities.

Socially and economically disadvantaged groups are among the most vulnerable segments of society. Those in rural areas, racial and ethnic minorities, workers in routine manual or cognitive jobs, and regions or countries with automatable labor are particularly at risk. Lack of familiarity with digital technologies also exacerbates vulnerability. Such individuals may, for example, have difficulty in detecting deception, be particularly susceptible to misinformation, and be at a loss in understanding privacy and data governance implications when using a computing product or service.

Another important accountability thread is that professionals and others are starting to delegate decision-making authority to computing technology. The FDA (2018) has allowed the use of an AI device to detect a diabetic condition related to blindness. In some cases, technology can make more accurate decisions than a professional (e.g., in certain applications of radiology), but ethical and legal considerations loom. Machines are making complex decisions, including some that have legal ramifications. Yet it is not clear whether technology can or should be held responsible in a moral or legal sense when harm occurs. The working group cautioned that it is also not clear how the legal system will handle the complexities here. The law was designed with the underlying assumption that humans would be the actors in legally relevant circumstances. How do we as a society want to address the fact that we have legally significant decisions that could be made by something other than a human being?

The working group also noted that computing devices may enable us to hold people more accountable for their actions, but it may occur in unsettling or disruptive ways (e.g., technologies in a workplace bathroom that can detect whether employees have washed their hands). The negative consequences (e.g., privacy concerns from feeling constantly monitored) would need to be weighed against an increase in accountability.

From the perspective of computing education, due to the way in which the curriculum is designed and other factors, computing students are not consistently considering the impact of what they are working on and/or may think it is someone else's problem to address. The computing community, in part by how it structures its pedagogy, needs to ensure that accountability is seen as an essential part of its mindset. Computing students are going into the world focused on technical aspects of their work, but they are not always thinking critically about their work's broader consequences. This seems to be a problem noted by Cech (2014) within the undergraduate engineering curriculum.

The working group suggested that a company should be more accountable than an individual person, but there are different kinds of accountability; for example:

- There is accountability for creating the technology
- There is accountability for putting controls in place pertaining to technology (or failing to do so)
- There is accountability for the operator or the customer
- There is accountability for how accessible/available the technology is

The lack of consistent standards, promulgated by industry or otherwise, means that we have a diminished ability to know the cause (much less respond) when things go wrong with computing artifacts. An added complexity is that human-machine teams can make it difficult to trace a specific behavior to a specific decision, and who has the responsibility and who should take the control in a particular context.

4.4 Transparency in Computing and Associated Challenges

We have reached a time when it is a common occurrence that developers, users, and others do not fully understand how a technology functions. This "black box" problem is a frequent concern in computing in the sense that it is often opaque how a sophisticated computing artifact makes a decision (Bleicher 2017). Users, and even designers, cannot necessarily discern how and why a computing device is offering a particular recommendation. Microsoft's Tay chatbot, for example, started to develop sexist, racist vocabulary a few hours after being deployed and had to be taken down (Price 2016). It is not clear how Tay worked (and how exactly it learned). On too many occasions, industry does not have mechanisms for evaluating the impact of the outcomes of its computing artifacts nor does industry make the functions of these artifacts or inner workings transparent to users.

Without understanding the inner workings of a computing artifact, it then becomes difficult to evaluate whether the data fed into the artifact is even appropriate for the specific task. For example, is it appropriate to use facial features to determine criminality? If so, under which circumstances? One of the differences seen in the machine learning community is training and validating their algorithms on conveniently available data and training algorithms to "force" causality (for example, assuming that facial features can be used to determine criminality and then training a model to do so). This approach contrasts with the traditional processes used in the social sciences, where data are carefully collected to determine whether causality exists.

In terms of research, the need for new methods in explainability and interpretability is growing in order to enhance the transparency of computing systems. From the perspective of education, designing transparency into computing systems is a crucial goal as it can allow students to understand more fully the societal impacts of the systems they are developing.

4.5 Discussion and Summary Remarks

The issues regarding FEAT-related challenges in the computing realm are varied, complex, and not easy to solve (for additional readings recommended by the working group, see **Appendix H**). Many barriers must be overcome, including the lack of sufficient incentives and resources, especially to pursue computing research that directly involves and benefits the public. To address these challenges, the following section addresses the working group's insights regarding best practices that can be employed. A variety of approaches are needed, including interdisciplinary collaboration that involves working across disciplines and employment sectors.

5. Promoting Best Practices

The challenges raised by FEAT in computing are certainly difficult to overcome but in at least some cases, there are best practices that could be followed. In other cases, the development of best practices is needed.

5.1 Fairness in Computing Best Practices

A key step towards addressing fairness concerns in computing is through promoting inclusion, in the various senses of the term. This involves in part diversifying the computing student population and the associated pipeline of developers. Inclusive curricular design and pedagogical methods, as well as educating computing students to understand the relevance of fairness and equity, can also help to promote fairness in computing. The working group asserted that recognizing how technology often disproportionately harms historically disadvantaged groups is an important step, and thus directly involving various target populations in technology and human subject research studies is key.

Initiatives that can provide a template for inclusive research include:

- Instituting external audits of computing artifacts and algorithmic decision making
- Creating and implementing standards that improve quality in data collection and analysis
- Developing and satisfying checklists that increase the likelihood of reproducible research
- Employing practices found in Participatory Action Research (PAR)
- Expanding FEAT in NSF Broadening Participation in Computing programs
- Providing incentives for interdisciplinary research that focuses directly on public involvement and outreach

Initiatives that can provide a template for fairness in education include:

- Making computing-related curriculum equally accessible (e.g., not every high school student gets access to advanced mathematics classes) (Toldson 2016)
- Designing a more inclusive curriculum
- Improving teacher quality in terms of their preparation and ability to teach
- Developing consistent metrics for evaluation
- Implementing evidence-based pedagogy that can reach a diverse range of students, including those with disabilities
- Implementing sustainable programs that can educate cohorts of students over time

Other practices that can promote fairness in research and education include:

- Mission statements that make a sincere commitment to fairness and other ethics-related goals
- The formation of diversity action pledges and councils
- Providing assistance and support to institutions who are seeking to develop an inclusive curriculum
- Developing culturally-relevant and personally-relevant training
- Implementing principles of universal design
- Developing codes of ethics and mechanisms of enforcement
- Engaging in advocacy, including efforts to assist disadvantaged and vulnerable communities

The working group also stated that industry could promote fairness by supporting inclusive educational and research efforts at academic institutions along with closing the funding gap at HBCUs (Toldson 2016).

5.2 Ethics in Computing Best Practices

Much of what the working group recommended pertains to improving ethics education in computing. In terms of academic institutions, it is clear that ethics should be embedded across the curriculum (not just seen as an afterthought in a single course). This can reinforce the mindset that ethics is an integral part of being a true computing professional (and is not an “externality” or add on). The message and content of promoting social and professional responsibility needs to be reiterated over time.

Ethics-related content that could bolster the computing curriculum and promote inclusiveness includes:

- Feminist Ethics, Ethics of Care
- Critical Race Theory
- Informational Ethics
- Disability Studies
- Readings attending to issues of power and democracy, stakeholder analysis, and differences (e.g., Value Sensitive Design)

The working group mentioned specific examples of effective pedagogical approaches, including James Madison University's Ethical Reasoning in Action Model and the University of Oxford's Ethical Hackathon.

Part of what needs to be incorporated in the computing curriculum are discussions about the different values that go into the design of technology; the myth of "value-neutrality" of technology can obscure what factors into the developer's decision-making process. Developers make choices and those choices are informed and shaped by values. For example, whether facial recognition technology should err on the side of a false positive or a false negative is a value judgment (pertaining to which type of error is seen as being more or less problematic). Teaching students about tradeoffs is a key lesson; for instance, less privacy might lead to less expensive access to technology but are users voluntarily agreeing to this arrangement?

Promoting ethical practices in research includes:

- Increasing the emphasis on RCR/research ethics in computing disciplines
- Requiring benchmarks for the reproducibility of results, which is a crucial issue to address in computing-based research efforts
- Increasing consistency of data documentation techniques. Just as in medical-based research, computing-based research needs to ensure the documentation of the data collection method, both positive and negative outcomes, and other parameters are published and shared.
- Employing inclusive research methodologies, including Community Based Participatory Research (e.g., Faridi et al. 2007) and Participatory Design (e.g., Spinuzzi 2005)

5.3 Accountability in Computing Best Practices

A key step toward accountability is documenting who did what and when; documentation is crucial in terms of holding individuals or groups responsible for their actions. Relevant types of information include:

- Documenting a procedure (e.g., the methods used and the steps involved)
- Documenting a decision (e.g., what the decision was, the bases for the decision, and the identity of the decision-makers)
- Establishing rules or guidelines that determine whether the procedure/decision is legal, ethical, and/or socially acceptable

Also, the creation of standards and principles can establish the conditions under which someone can be held accountable. When enforced correctly, accountability can:

- Prevent harmful events from happening
- Address harmful events that happen
- Help the people who have been negatively affected

In short, efforts should be put forward to create metrics for what accountability is and what all it entails in specific use cases (e.g., if an autonomous car collides with a pedestrian). The working group posited that keeping a human in the loop may help with some of the accountability complexities by allowing for contemporaneous documentation and, in some cases, real-time intervention.

Principles and habits of accountability should be instilled at the outset of training in computing--and, indeed, at the outset of being a digital citizen--and reinforced gradually throughout schooling. Introducing students and others to scholarly work, such as Langdon Winner's "Do Artifacts Have Politics?" (1980), might reveal how the development and use of technology can reshape society, and even in some cases, erode democracy.

From the research perspective, standards of accountability should be enforced and reinforced in computing artifacts as they are created. IEEE, for example, is in the process of developing technical standards, including P7010 which is at the intersection of human well-being and intelligent systems. Standards of accountability should be instilled at the outset of a research project to ensure proper conduct and methods beyond compliance and IRB monitoring.

Other efforts that could promote accountability at the individual level include:

- Considering the development of a computing "Hippocratic oath" (using IEEE or ACM codes of conduct as roadmaps)
- Developing a certification or licensing for computing professionals (but without creating additional barriers for underrepresented groups)

At the organization level, measures such as the following could be considered:

- Life cycle assessment/evaluation of computing products
- Anticipating potentially harmful secondary uses of products
- Ensuring proper training of workers
- Instilling principles of accountability across an organization
- Encouraging employers to take responsibility for supporting workers who could be affected by changes due to automation (e.g., efforts like the Emma Coalition)
- External auditing and governance
- Protections for whistleblowers
- Auditing requirements akin to those required of public companies
- Insurance underwriting requirements

5.4 Transparency in Computing Best Practices

As a starting point for promoting transparency, the working group recommended that if someone is going to deploy an algorithm or other form of computing technology, a corresponding statement of impact and reflection of what could go wrong should be developed alongside of it. There needs to be a renewed push for developers and researchers to report the limitations of their technology.

Members of the computing community need to foster a culture of explaining how computing technology works in a way in which users and other stakeholders are likely to understand. For example, a family who is applying for a home loan does not necessarily need to be taught Machine Learning techniques, but they should be told what the reasons are for the loan being accepted or rejected. This could be patterned after the informed consent process enforced by IRBs; consent forms in the realm of human subjects research must be written at the reading level of the target audience. The transparency stipulations within the European GDPR might be a helpful model as well. Also, enabling users of computing technologies to have clearer opportunities to opt in and opt out is crucial.

Establishing standards which require that test results and model evaluation be designed based on data sets other than what is originally being tested on is a key step. For example, for some DARPA programs, datasets are kept secret from the researchers until "demo" day – at which point, their algorithms are validated against a hold-out set. This then becomes a true test of how well the algorithms perform in a more realistic scenario as it tries to limit bias of the researchers which may occur if they design their algorithms around their own specific data.

Computing technology needs to be available for auditing. That is, if a product is to be made publicly available, it should be available for people to test it. Facial recognition systems, for instance, became better after their flaws were pointed out by others outside of the development team (Buolamwini and Gebru 2018); this process generates improvements. The working group consistently emphasized the practice of third-party inspection and evaluation of a product. This type of practice is typically performed when evaluating the outcomes of educational programs; it makes sense to expand this practice to computing-based research.

Other measures that may assist with promoting transparency include:

- Interdisciplinary collaboration and academic-industry collaboration
- Changing the current model of optimizing for accuracy by implementing transparency metrics for performance on all groups
- Bias detection and explanation at a high level
- Understanding feature sets for machine learning
- Validating transparency claims (HCI) with actual subjects

Overlapping with the issue of transparency is whether to trust in self-regulation versus external regulation when it comes to the development of computing artifacts. On one hand, some in the computing community would argue that external regulation stifles innovation. Yet on the other hand, it is clear that the computing community has failed to develop sufficient internal mechanisms to protect the public. The working group discussed whether a regulatory or other entity should have jurisdiction over reviewing computing artifacts, especially AI algorithms. The working group briefly considered whether it should take the form of an "FDA for AI" or whether it would be more appropriate to create sector-specific agencies (for example, one for financial AI and a different one for AI in the criminal justice realm). Even if it is not settled whether a single or multiple entities should be involved in the AI space, developing a process similar to the FDA review phases might be important. The working group noted that the European Regulation of Chemical Industry (REACH) might be a model for the regulation of AI, as it relies on the combination of transparency, openness, and crowdsourced activity to detect risks. Furthermore, the FAA has a vetting process for the use of algorithms in airplanes, which perhaps is a model that can be more widely embraced.

5.5 Discussion and Summary Remarks

A key overarching comment in this section is addressing the question of who is in the room when computing-related education and research decisions are being made. The FEAT best practices identified above will fail to work if the various stakeholders that are impacted by computing artifacts are not represented in the relevant processes. In addition, systematic efforts are needed to determine when and how humans should remain in the decision-making loop when computing technology is being used.

6. The Evaluation of FEAT

The collection of the best practices mentioned above may help to gauge the effectiveness of measures to promote FEAT-related goals. In addition to those practices, the working group described other measures that support the evaluation of FEAT in computing. One strategy is to develop metrics for assessing whether changes to research practice (such as the external auditing of algorithms) contributes to the creation of improved computing artifacts (e.g., fairer algorithms). The working group specifically encouraged the development and assessment of new methods to detect, quantify, and mitigate bias.

In the realms of both education and research, formal assessments can detect whether efforts at promoting inclusion, including those supporting cognitive diversity in a genuine, authentic sense, are working. Assessments could also capture the degree to which gaps resulting from historical inequities are closing, including gaps in funding and representation. Additionally, systematic analyses of the benefits and harms of computing artifacts could be undertaken, including the degree to which such artifacts promote or erode ethical ideals such as respect for persons, justice, and fairness. Along these lines, data could be collected on whether computing education and research support human rights and are genuinely fostering human well-being locally, nationally, and internationally.

The working group also delineated other measures for evaluating the success of FEAT-related efforts including metrics for identifying:

- How widespread the adoption of internal and external regulation (governance boards, standards, etc.) is
- Whether consensus on standards and best practices (from entities such as IEEE, NIST, and the European Union) emerges
- Gauging the internalization of social and ethical considerations beyond mere compliance in different organization types
- Measuring attitude changes in educational and research settings; this could for example involve conducting surveys of first-year students, graduates, and professionals now and in 10 years through the use of tools such as the Engineering Professional Responsibility Assessment (Canney and Bielefeldt 2016) or the Generalized Professional Responsibility Assessment

- The degree to which the creation of “ethical” computing products occurs, including products that uphold universal design principles and are genuinely usable by those with disabilities or impairments
- The degree to which public interest technology outputs and outcomes (camps, programs, funding, etc.) emerge

An overarching notion expressed by the working group is that researchers and others could formally examine the degree to which promising practices are being adopted; the group specifically mentioned that important developments to follow include the use of datasheets for datasets (Gebru et at. 2019), energy usage monitoring (such as how much energy an AI system requires), and life cycle and well-being impact assessment.

7. Responsibilities of Educators and Researchers

Given the extent to which computing is reshaping the lives of people around the globe, ranging from the Cambridge Analytica case to different forms of algorithmic bias, computing educators, researchers, and practitioners clearly have a responsibility to protect the public. Arguably this responsibility extends beyond mere harm avoidance to an aspirational goal to do good. Members of the computing community need to embrace anticipatory ethics and more fully realize their responsibility to reflect on how the work they are undertaking is reshaping society.

Is the computing profession taking its role seriously enough? Often when a profession recognizes its importance to society and the level of trust placed in it, formal mechanisms are put in place (such as licensing). Perhaps a form of certification should be revisited for different sectors of the computing profession. It must be balanced against the concern that it might be a barrier for entry for those who have been historically blocked from entering the profession.

The computing community, broadly defined, needs to develop a stronger sense of responsibilities for the information and technology that it is helping to generate. In important senses, experiments on a local, national, and global scale are being run on humans due to the introduction of computing applications into our lives.

The working group also mentioned the idea of nurturing a “pro bono” ethos in computing; building a mindset of service could be an important step towards lessening the disconnect between the computing community and the public.

8. Other Points of Discussion

Many other ideas were mentioned during the workshop that intersect with FEAT concerns; one of those ideas is whether people should be paid for the personal information that researchers and companies are monetizing. Many of the “free” online services are making exorbitant revenues from our personal data, which computing researchers are also freely scrubbing and downloading for their own research uses. Users are often unaware of such practices or do not fully realize what the risks are.

Another issue that the working group brought to light is the reuse of algorithms in contexts for which they were not originally designed. This practice is part of the reason why algorithms are generating fairness problems or other harms. Also, the group warned that when you think that every problem can be solved by algorithms, you start applying algorithms to situations where they probably should not be used, such as to determine whether someone will remain in prison or be set free, or whether to allow someone to immigrant into a country (Molnar and Gill 2018).

The working group also suggested that the energy costs of computing (to run AI algorithms for example) needs a fuller examination. The environmental, sustainability dimensions of computing should be an important focus area for both researchers and educators.

An overarching thread through several of the workshop discussions is the role of industry as it pertains to FEAT. If, for example, industry shows leadership in research and development, the prevention of harmful impacts (such as inaccurate facial recognition) might follow. Industry has a key role to play with regard to ethics. Companies are in the process of trying to figure out how to guide the development of AI (Simonite 2018). Their efforts have had mixed success, but CEO directives, social impact statements, and promoting inclusive practices can alter the landscape of computing. Meanwhile, inclusive approaches, such as developing accessible technologies like SMS for hearing impaired users, can improve usability and benefit a wider range of users. To uphold accountability of industry and other entities, the creation of a computer and information science equivalent of the National Transportation Safety Board (NTSB) could be an option as well.

9. Conclusion

The fairness, ethics, accountability, and transparency (FEAT) workshop convened a diverse collection of experts into a working group to identify challenges in computing-based research and education. The working group noted that such challenges are not easy to solve in part because the FEAT terminology can be difficult to define precisely. Yet through inclusive educational and research initiatives, meaningful progress can be made toward improving computing practice and promoting the public's well-being.

Acknowledgments

The workshop organizers would like to thank the expert participants who contributed to the workshop and/or the workshop report: John Banja, Justin Biddle, Aylin Caliskan, Erin Dalton, Tawanna Dillahunt, Edward Dillon, Carl DiSalvo, Nitcelle Emanuels, Sandeep Gopisetty, Christina Harrington, Leshell Hatley, Monique Head, Kaye Husbands Fealing, Jacquelyn Krones, Joseph Lyons, Keith Miller, Errika Moore, Jaye Nias, Nathan Nobis, Adriane Randolph, Monique Ross, Matt Scherer, Daniel Schiff, Kristen Shinohara, Karen Silverman, Jamila Smith-Loud, Felesia Stukes, Suresh Venkatasubramanian, Phil Ventimiglia, Gloria Washington, and Ellen Zegura.

The workshop organizers would also like to thank Toya Jordan at Georgia Tech who assisted with event logistics and the Georgia Tech student notetakers at the workshop Adriana Alvarado, De'Aira Bryant, Katelyn Fry, and Shubhangi Gupta.

References

Bleicher, Ariel. 2017. Demystifying the Black Box That Is AI. Scientific American, August 9, https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/.

Buolamwini, Joy and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81: 1-15.

Bussewitz, Cathy. 2019. The E-Scooter Boom Has Caused At Least 11 Deaths Since the Beginning of 2018. Business Insider, June 7, https://www.businessinsider.com/boom-in-electric-scooters-leads-to-more-injuries-fatalities-2019-6.

Canney, N. E. and A. R. Bielefeldt. 2016. Validity and Reliability Evidence of the Engineering Professional Responsibility Assessment Tool. Journal of Engineering Education, 105: 452-477. doi: 10.1002/jee.20124.

Cech, Erin. 2014. Culture of Disengagement in Engineering Education? Science, Technology, & Human Values, 39(1): 42-72.

Clauset, Aaron, Samuel Arbesman, and Daniel B. Larremore. 2015. Systematic Inequality and Hierarchy in Faculty Hiring Networks. Science Advances, Vol. 1, no. 1, e1400005.

Crawford, K. 2016. Artificial Intelligence's White Guy Problem,' The New York Times, http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html.

Crawford, K., M. Whittaker, M.C. Elish, S. Barocas, A. Plasek, and K. Ferryman. 2016. AI Now 2016 Report, https://ainowinstitute.org/AI_Now_2016_Report.html.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. 2011. Extraneous Factors in Judicial Decisions. Proceedings of the National Academy of Sciences, 108 (17) 6889-6892. doi: 10.1073/pnas.1018033108.

Data USA. 2018. DATAUSA: Artificial Intelligence, https://datausa.io/profile/cip/110102/.

Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. 2016. Echo Chambers in the Age of Misinformation. Proceedings of the National Academy of Sciences, 113 (3) 554-559. doi: 10.1073/pnas.1517441113.

Deruy, Emily. 2016. Judge's Football Team Loses, Juvenile Sentences Go Up: No, Seriously. The Atlantic, September 7, https://www.theatlantic.com/education/archive/2016/09/judges-issue-longer-sentences-when-their-college-football-team-loses/498980/.

Eubanks, Virginia. 2018. Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor. St. Martin’s Press.

Faridi Z., J.A. Grunbaum, B.S. Gray, A. Franks, and E. Simoes. 2007. Community-Based Participatory Research: Necessary Next Steps. Preventing Chronic Disease, http://www.cdc.gov/pcd/issues/2007.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, and Kate Crawford. 2019. Datasheets for Datasets. arXiv.org. arXiv:1803.09010.

Harris, C.E., M.S. Pritchard, and M.J. Rabins. 2009. Engineering Ethics: Concepts and Cases (4th ed). Belmont, CA: Wadsworth Cengage Learning.

Herold, Benjamin. 2017. Poor Students Face Digital Divide in How Teachers Learn to Use Tech. Education Week, June 15, https://www.edweek.org/ew/articles/2017/06/14/poor-students-face-digital-divide-in-teacher-technology-training.html.

Howard, A. and J. Borenstein. 2018. The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. Science and Engineering Ethics 24: 1521, https://doi.org/10.1007/s11948-017-9975-2.

Johnson, Khari. 2019. How AI Companies Can Avoid Ethics Washing. Venture Beat, July 17, https://venturebeat.com/2019/07/17/how-ai-companies-can-avoid-ethics-washing/.

Kramer, Adam D. I., Jamie E. Guillory, Jeffrey T. Hancock. 2014. Emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111 (24) 8788-8790. doi: 10.1073/pnas.1320040111.

Mayo, Benjamin. 2019. Major iPhone FaceTime Bug Lets You Hear the Audio of the Person You Are Calling … Before They Pick Up. 9to5Mac, January 28, https://9to5mac.com/2019/01/28/facetime-bug-hear-audio/.

Molnar, Petra and Lex Gill. 2018. Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada’s Immigration and Refugee System. International Human Rights Program and the Citizen Lab, University of Toronto.

Moody, J. 2004. Faculty Diversity: Problems and Solutions. Routledge.

National Academies of Sciences, Engineering, and Medicine. 2019. Minority Serving Institutions: America's Underutilized Resource for Strengthening the STEM Workforce. Washington, DC: The National Academies Press, https://doi.org/10.17226/25257.

National Science and Technology Council, Committee on Technology, Executive Office of the President. 2016. Preparing for the Future of Artificial Intelligence, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

National Science Foundation (NSF), National Center for Science and Engineering Statistics. 2019. *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019.* Special Report NSF 19-304. Alexandria, VA, https://www.nsf.gov/statistics/wmpd.

O’Neil, Cathy. 2016. Weapons of Math Destruction. Crown Random House.

Patel, Prachi. 2015. Engineers, Ethics, and the VW Scandal. IEEE Spectrum, September 25, http://spectrum.ieee.org/cars-that-think/at-work/education/vw-scandal-shocking-but-not-surprising-ethicists-say.

Price, Rob. 2016. Microsoft Took Its New A.I. Chatbot Offline After It Started Spewing Racist Tweets. Slate, March 24, https://slate.com/business/2016/03/microsoft-s-new-ai-chatbot-tay-removed-from-twitter-due-to-racist-tweets.html.

Schmelzer, Ron. 2019. What Happens When Self-Driving Cars Kill People? Forbes, September 26, https://www.forbes.com/sites/cognitiveworld/2019/09/26/what-happens-with-self-driving-cars-kill-people/.

Shankland, Stephen. 2018. Facebook, Cambridge Analytica Face Lawsuit Over Privacy Loss. Cnet, March 22, https://www.cnet.com/news/facebook-cambridge-analytica-face-lawsuit-over-privacy-loss/.

Simonite, Tom. 2018. Tech Firms Move to Put Ethical Guard Rails Around AI. Wired, May 16, https://www.wired.com/story/tech-firms-move-to-put-ethical-guard-rails-around-ai/.

Spinuzzi, Clay. 2005. The Methodology of Participatory Design. Technical Communication, 52(2): 163-174.

Toldson, Ivory A. 2016. The Funding Gap between Historically Black Colleges and Universities and Traditionally White Institutions Needs to be Addressed* (Editor’s Commentary). The Journal of Negro Education Vol. 85, No. 2, The 36th Annual Charles H. Thompson Lecture: Why Black Lives (and Minds) Matter: Race, Freedom Schools & the Quest for Educational Equity (Spring 2016), pp. 97-100.

U.S. Food and Drug Administration (FDA). 2018. FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems, April 11, https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye.

Vara, Vauhini. 2016. Why Doesn’t Silicon Valley Hire Black Coders? Howard University Fights to Join the Tech Boom. Bloomberg, January 21, https://www.bloomberg.com/features/2016-howard-university-coders/.

Winner, Langdon. 1980. Do Artifacts Have Politics? Daedalus, 109(1), 121-136.

Appendices

A. Workshop Participants
B. Workshop Agenda
C. Fairness Panel Questions
D. Ethics Panel Questions
E. Accountability Panel Questions
F. Transparency Panel Questions
G. Breakout Group Questions
H. Other Recommended FEAT Readings

Appendix A: List of Workshop Participants

Workshop Organizers

Ayanna Howard
Georgia Institute of Technology

Jason Borenstein
Georgia Institute of Technology

Kinnis Gosha
Morehouse College

NSF Program Officers

Fay Cobb Payton

Tonya Smith-Jackson

Workshop Participants

John Banja
Emory University

Justin Biddle
Georgia Institute of Technology

Aylin Caliskan
George Washington University

Erin Dalton
Depart. of Human Services, Allegheny County

Tawanna Dillahunt
University of Michigan

Edward Dillon
Morgan State University

Carl DiSalvo
Georgia Institute of Technology

Nitcelle Emanuels
Dell

Sandeep Gopisetty
IBM

Errika Moore
TAG-Ed Education Collaborative

Jaye Nias
Spelman College

Nathan Nobis
Morehouse College

Adriane Randolph
Kennesaw State University

Monique Ross
Florida International University

Matt Scherer
Littler Mendelson P.C.

Daniel Schiff
Georgia Institute of Technology

Kristen Shinohara
Rochester Institute of Technology

Karen Silverman
Latham & Watkins LLP

Christina Harrington
Northwestern University

Leshell Hatley
Coppin State University

Monique Head
Paypal

Kaye Husbands Fealing
Georgia Institute of Technology

Jacquelyn Krones
Microsoft

Joseph Lyons
Air Force Research Laboratory

Keith Miller
University of Missouri - Saint Louis

Jamila Smith-Loud
Google

Felesia Stukes
Johnson C. Smith University

Suresh Venkatasubramanian
University of Utah

Phil Ventimiglia
Georgia State University

Gloria Washington
Howard University

Ellen Zegura
Georgia Institute of Technology

Workshop Assistants

Adriana Alvarado
Georgia Institute of Technology

De'Aira Bryant
Georgia Institute of Technology

Katelyn Fry
Georgia Institute of Technology

Shubhangi Gupta
Georgia Institute of Technology

Appendix B: Workshop Agenda

Meeting Time: Thursday, August 29th, 8:30am–5:00pm and Friday, August 30th, 8:30am–1:30pm
Location: Tech Square Research Building (TSRB), Rooms 132 and 133
85 Fifth Street NW, Atlanta, GA 30308

DAY ONE: Thursday August 29th

8:00-8:30AM – Breakfast

8:30-8:45AM – Goals of the FEAT Workshop

- Fay Cobb Payton, NSF
- Tonya Smith-Jackson, NSF

8:45-9:30AM – Attendee Introductions

9:30-11:00AM – Fairness Panel

- Edward Dillon, Morgan State University (moderator)
- Nitcelle Emanuels, Dell
- Suresh Venkatasubramanian, University of Utah
- Gloria Washington, Howard University

11:00AM-12:30PM – Ethics Panel

- Tawanna Dillahunt, University of Michigan (moderator)
- Keith Miller, University of Missouri - Saint Louis
- Jamila Smith-Loud, Google
- Felesia Stukes, Johnson C. Smith University

12:30-1:15PM – Lunch

1:15-2:45PM – Accountability Panel

- Joseph Lyons, Air Force Research Laboratory (moderator)
- Monique Head, Paypal
- Monique Ross, Florida International University
- Matt Scherer, Littler Mendelson P.C.
- Karen Silverman, Latham & Watkins LLP

2:45-3:00PM – Break

3:00-4:30PM – Transparency Panel

- Jaye Nias, Spelman College (moderator)
- Aylin Caliskan, George Washington University
- Erin Dalton, Depart. of Human Services, Allegheny County
- Sandeep Gopisetty, IBM
- Jacquelyn Krones, Microsoft

4:30-5:00PM – Topical Breakout Group Membership and Charge

6:00-8:00PM – Working Dinner (location Georgia Tech Hotel restaurant)

DAY TWO: Friday August 30th

8:00-8:30AM – Breakfast

8:30AM-11:30AM – Breakout Group Meeting Time

11:30AM-12:30PM – Breakout Group Presentations

12:30-1:30PM – Working Lunch

- Open discussion
- Next steps

Appendix C: Fairness Panel Questions

1. Could you briefly explain how your expertise/work intersects with the topic area?

2. What does fairness mean to you?

3. What are the most significant fairness-related challenges that you see in the realm of computing-based research and education?

4. What does bias mean and in which circumstances does it become ethically problematic?

5. What are the ethical obligations of computing professionals and communities to uphold fairness and prevent bias?

6. How could "fairer" computing technologies be designed?

7. Are there best practices related to the acquisition and use of training data that could help to mitigate bias?

Appendix D: Ethics Panel Questions

1. Could you briefly explain how your expertise/work intersects with the topic area?

2. What are the most significant ethical challenges/issues that you see in the realm of computing-based research and education?

3. As we near the era of AI, what does privacy mean, and in which circumstances do privacy violations become ethically problematic?

4. Are there effective methods for upholding privacy in the digital age?

5. Are there effective methods for ensuring that users place an appropriate amount of trust in computing technologies?

6. How are computing technologies affecting human-human relationships? Is there evidence to indicate that computing technologies may be causing these relationships to erode?

7. Is it a significant concern that users might become addicted to computing devices, and if so, what measures could be put in place to address the concern?

8. When an AI program learns and changes, what are some ethically significant opportunities or benefits and what are some ethically significant risks? Also, how should the benefits be balanced against the risks?

Appendix E: Accountability Panel Questions

1. Could you briefly explain how your expertise/work intersects with the topic area?

2. What does accountability mean to you?

3. What do you see as the similarities and differences between responsibility and accountability?

4. What are the most significant accountability-related challenges that you see in the realm of computing-based research and education?

5. Under which circumstances would you hand over decision-making authority to an autonomous system for a low risk task? How about for a high risk task?

6. How does the momentum towards increasingly autonomous technologies impact efforts to hold an individual or group accountable for computing technologies?

7. Are there strategies for holding individuals or groups accountable when a computing technology causes harm?

8. What do you think are the most effective methods for preventing potential harms caused by computing technology?

Appendix F: Transparency Panel Questions

1. Could you briefly explain how your expertise/work intersects with the topic area?

2. What does transparency mean to you?

3. What are the most significant transparency-related challenges that you see in the realm of computing-based research and education?

4. Is the “black box” problem a significant technical and ethical concern? What kinds of harms could result from it or more generally, a lack of transparency in the computing realm?

5. What are your views on “explainable AI”? For example, is it a feasible approach?

6. Which methods could be implemented to make computing technologies more transparent to users, including the professionals that rely on them and consumers?

7. Is the traditional model of informed consent robust enough in the AI/robotics age?

Appendix G: Breakout Group Questions

Questions guiding the breakout group discussion include:

1. Provide a working definition(s) of your topic area. Which communities are involved in defining this topic area? In which ways is there agreement and/or disagreement in these communities about the definition of the topic area?

2. Why is your topic area important to the realm of:
    - Research?
    - Education?
    - Industry?

3. Provide specific examples of how your topic area is impacting current computing-based research, educational, and industry efforts. Please provide information sources for these examples (for instance, as identified via news stories, blogs, papers, etc.).
    - Research:
    - Education:
    - Industry:

4. Which strategies have academics, practitioners, and others implemented to address these issues (as identified above)? For example, new research methods, legislations, presidential directives, industrial practices, new ethics boards, etc. Please provide information sources.

5. Which strategies are the most promising to help address specific aspects of your topic area? Are there some that have not been proposed yet by the different communities?

6. How might societal or other aspects related to your topic area change over the next 5-10 years?

7. Which computing-based research, educational, and/or industry efforts pose the most risk to the public if your topic area is not adequately addressed? Which stakeholders (e.g., vulnerable populations) are most at risk and in which ways?

8. Based on the working definition(s) of your topic area, how can progress and success be measured? What would success look like?

9. Are there any other important issues or topics related to FEAT in computing-based research, practice, and education that have not been addressed by the above questions? Please describe.

Appendix H: Other Recommended FEAT Readings

Adam, Alison and Jacqueline Ofori-Amanfo. 2000. Does Gender Matter in Computer Ethics? Ethics and Information Technology, 2(1): 37-47.

Abdul, Ashraf, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, doi: https://doi.org/10.1145/3173574.3174156.

Ananny, Mike and Kate Crawford. 2018. Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. New Media & Society, 20(3): 973-989.

Andrew, Alexa. 2006. The Ethics of Using Dolls and Soft Toys in Dementia Care. Nursing and Residential Care, 8(9): 419-421.

Arkin, Ronald C. 2010. The Case for Ethical Autonomy in Unmanned Systems. Journal of Military Ethics, 9(4): 332-341.

Barocas, Solon and Andrew D. Selbst. 2016. Big Data's Disparate Impact. Calif. L. Rev. 104: 671.

Benjamin, Ruha. 2019. Race After Technology. Polity.

Blackwelder, B., K. Coleman, S. Colunga-Santoyo, J.S. Harrison, and D. Wozniak. 2016. The Volkswagen Scandal, https://scholarship.richmond.edu/robins-case-network/17/.

Bolukbasi, Tolga, et al. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Advances in Neural Information Processing Systems.

Borenstein, Jason and Ron Arkin. 2016. Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being. Science and Engineering Ethics, 22(1): 31-46.

Brey, Philip. 2010. Values in Technology and Disclosive Computer Ethics. The Cambridge Handbook of Information and Computer Ethics, 41-58.

Bryson, Joanna J. 2010. Robots Should be Slaves. Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues, 63-74.

Bynum, Terrell Ward. 2008. Milestones in the History of Information and Computer Ethics. The Handbook of Information and Computer Ethics.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics Derived Automatically from Language Corpora Contain Human-Like Biases. Science, 356(6334): 183-186.

Camp, Tracy. 1997. The Incredible Shrinking Pipeline. Communications of the ACM, 40(10): 103-110.

Chokshi, Niraj. 2018. Is Alexa Listening? Amazon Echo Sent Out Recording of Couple’s Conversation. The New York Times.

Collins, W. Robert and Keith W. Miller. 1992. Paramedic Ethics for Computer Professionals. Journal of Systems and Software, 17(1): 23-38.

Costanza-Chock, S. 2019. Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice. DRS2018: Catalyst. doi: https://doi.org/10.21606/drs.2018.679.

Datta, Anupam, Shayak Sen, and Yair Zick. 2016. Algorithmic Transparency Via Quantitative Input Influence: Theory and Experiments with Learning Systems. 2016 IEEE Symposium on Security and Privacy (SP).

Davis, Michael. 2012. "Ain't No One Here But Us Social Forces": Constructing the Professional Responsibility of Engineers. Science and Engineering Ethics, 18(1): 13-34.

Dolmage, Jay Timothy. 2017. Academic Ableism: Disability and Higher Education. University of Michigan Press, Ann Arbor, https://doi.org/10.3998/mpub.9708722

Doshi-Velez, Finale, and Been Kim. 2017. Towards a Rigorous Science of Interpretable Machine Learning. arXiv preprint arXiv:1702.08608.

Dwork, Cynthia, et al. 2012. Fairness Through Awareness. Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. ACM.

Floridi, Luciano. 1999. Information Ethics: On the Philosophical Foundation of Computer Ethics. Ethics and Information Technology, 1(1): 33-52.

Friedman, Batya, Peter H. Kahn, and Alan Borning. 2008. Value Sensitive Design and Information Systems. The Handbook of Information and Computer Ethics, 69-101.

Gillborn, David, Paul Warmington, and Sean Demack. 2018. QuantCrit: Education, Policy, 'Big Data' and Principles for a Critical Race Theory of Statistics. Race Ethnicity and Education, 21(2): 158-179.

Goethe University. 2019. Diffracting AI and Robotics: Decolonial and Feminist Perspectives. Symposium and Workshop, Goethe University Frankfurt, 11-12 October.

Gotterbarn, Don, and Simon Rogerson. 2005. Responsible Risk Assessment with Software Development: Creating the Software Development Impact Statement. Communications of the Association for Information Systems, 15(1): 40.

Huff, Chuck and C. Dianne Martin. 1995 Computing Consequences: A Framework for Teaching Ethical Computing. Communications of the ACM, 38(12): 75-84.

Johnson, Deborah G. 1985. Computer Ethics. Englewood Cliffs.

Johnson, Deborah G. and Keith W. Miller. 2002. Is Diversity in Computing A Moral Matter? ACM SIGCSE Bulletin, 34(2): 9-10.

Kohlberg, Lawrence. 1971. Stages of Moral Development. Moral Education, 1(51): 23-92.

Kosinski, Michal, David Stillwell, and Thore Graepel. 2013. Private Traits and Attributes Are Predictable from Digital Records of Human Behavior. Proceedings of the National Academy of Sciences, 110(15): 5802-5805.

Kroll, Joshua A., et al. 2016. Accountable Algorithms. U. Pa. L. Rev. 165: 633.

Le Dantec, C.A. 2016. Design Through Collective Action / Collective Action Through Design. Interactions, 24(1): 24–30. doi: https://doi.org/10.1145/3018005.

Le Dantec, C.A. and Fox, S. 2015. Strangers at the Gate: Gaining Access, Building Rapport, and Co-Constructing Community-Based Research. CSCW ’15: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (Vancouver, BC, Canada), 1348–1358.

Lipton, Zachary C. 2016. The Mythos of Model Interpretability. arXiv preprint arXiv:1606.03490.

Lohr, Steve. 2018. Facial Recognition is Accurate, If You’re a White Guy. The New York Times.

Mankoff, Jennifer, Gillian R. Hayes, and Devva Kasnitz. 2010. Disability Studies as a Source of Critical Inquiry for the Field of Assistive Technology. Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility. ACM.

Mellström, Ulf. 2009. The Intersection of Gender, Race and Cultural Boundaries, or Why Is Computer Science in Malaysia Dominated by Women? Social Studies of Science, 39(6): 885-907.

Miller, Keith. 1988. Integrating Computer Ethics into the Computer Science Curriculum. Computer Science Education 1(1): 37-52.

Moor, James H. 1985. What is Computer Ethics? Metaphilosophy, 16(4): 266-275.

Mullen, Hilary, and David Horner. "Ethical problems for e-government: an evaluative framework." Electronic Journal of e-government 2.3 (2004): 179-188.

Nissenbaum, Helen. 2009. Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press.

Noble, Safiya Umoja. 2018. Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

Orbit.com. Ethical Hackathons, https://www.orbit-rri.org/services/ethical-hackathon/.

Partovi, Hadi. 2015. A Comprehensive Effort to Expand Access and Diversity in Computer Science. ACM Inroads, 6(3): 67-72.

Racadio, R. et al. 2014. Research at the Margin: Participatory Design and Community Based Participatory Research. PDC ’14: Proceedings of the 13th Participatory Design Conference: Short Papers, Industry Cases, Workshop Descriptions, Doctoral Consortium papers, and Keynote abstracts - Volume 2 (Windhoek), 49–52.

Rawls, John. 2009. A Theory of Justice. Harvard University Press.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. 2016. Why Should I Trust You? Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Richards, Neil M. and Jonathan H. King. 2014. Big Data Ethics. Wake Forest Law Review, 49: 393.

Silverman, Arielle M., Jason D. Gwinn, and Leaf Van Boven. 2014. Stumbling in Their Shoes. Social Psychological and Personality Science, 6(4): 464–471, https://doi.org/10.1177/1948550614559650.

Selvaraju, Ramprasaath R., et al. 2017. Grad-Cam: Visual Explanations from Deep Networks via Gradient-Based Localization. Proceedings of the IEEE International Conference on Computer Vision.

Sweeney, Latanya. 2013. Discrimination in Online Ad Delivery. arXiv preprint arXiv:1301.6822.

Torralba, Antonio and Alexei A. Efros. 2011. Unbiased Look at Dataset Bias. CVPR, 1(2).

Turilli, Matteo and Luciano Floridi. 2009. The Ethics of Information Transparency. Ethics and Information Technology, 11(2): 105-112.

Zook M, S. Barocas, D. Boyd, K. Crawford, E. Keller, S.P. Gangadharan, et al. 2017. Ten Simple Rules for Responsible Big Data Research. PLoS Comput Biol 13(3): e1005399.