NATIONAL ACADEMIES Sciences Engineering Medicine

COMPUTER SCIENCE AND TELECOMMUNICATIONS BOARD

Fostering Responsible Computing Research:
Foundations and Practices

Report available from National Academies Press
www.nap.edu

Study Scope

- Explore ethics and governance issues associated with the personal and social consequences of computing research and its applications
- Recommend practical steps that National Science Foundation-supported researchers and others in the computing research community can take to address ethics in all phases of their research from proposal to publication

Not focused on:

- Government regulation of the tech sector
- Ethical issues associated with the conduct of research itself

Committee

- Barbara J. Grosz, NAE, Harvard University (Chair)
- Mark Ackerman, University of Michigan, Ann Arbor
- Steve M. Bellovin, NAE, Columbia University
- Mariano-Florentino Cuéllar, Carnegie Endowment for International Peace
- David Danks, University of California, San Diego
- Megan Finn, University of Washington
- Mary L. Gray, Microsoft Research
- John L. Hennessy, NAS/NAE, Stanford University and Alphabet Inc.
- Ayanna M. Howard, Ohio State University
- Jon M. Kleinberg, NAS/NAE, Cornell University
- Seth Lazar, Australian National University
- James Manyika, McKinsey Global Institute and Google, Inc.
- James Mickens, Harvard University
- Amanda Stent, Colby College

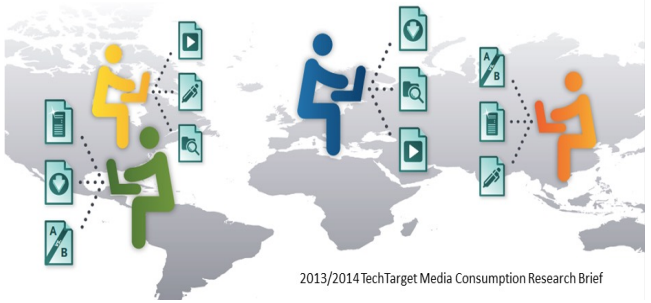
Why Now?

One person, one system, one place

→

Many people

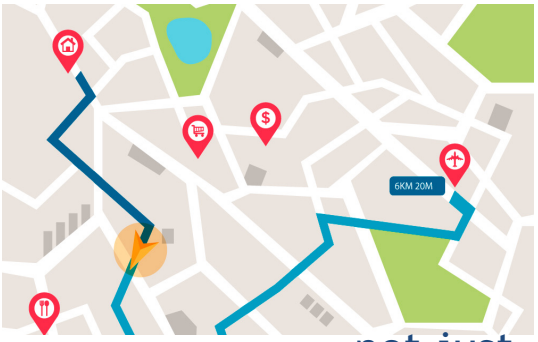
A multiplicity of systems



People & work distributed geographically

NATIONAL ACADEMIES Sciences Engineering Medicine

Use by “Every Citizen of the World”: Ethical & Societal Impact Concerns



not just AI/ML

not just faces

Two Ecosystems & Interactions Among Them

- Technology innovation
  - Academia, research performing institutions, industry
  - Multi-dimensional and multi-directional flow
- Computing research
  - Researchers, research community
  - Research sponsors
  - Research performing institutions

Corporate leaders, entrepreneurs, investors

Report Contents

Summary
1. Introduction
2. Theoretical Foundations from Ethical and Social Science Frameworks
3. Sources of Ethical Challenges and Societal Concerns for Computing Research
4. Conclusions and Recommendations

Appendixes

- Biographies of Committee Members
- Presentations to the Committee
- Federal Computing Research Programs Related to Ethical and Societal Impact Concerns

Recommendations in Brief

1. Reshape computing research to adequately include ethical and societal impact concerns
2. Foster & facilitate computing research aimed at furthering societal benefit and mitigating harm
3. Support the development of expertise in responsible computing in students and computing researchers in CS&E, social sciences and humanities
4. Provide researchers access to the knowledge and expertise needed to identify & assess the ethical & societal implications of their work
5. Integrate ethical and societal considerations into computing research sponsorship
6. Integrate ethical and societal considerations into publications and publicly released artifacts
7. Adhere to best practices for systems design, deployment, oversight & monitoring
8. Support engagement with the public and the public interest

Arriving at the Recommendations: Two sources of Expertise

Study committee expertise

- Computer science & engineering
  - Theory, systems, AI, HCI, cybersecurity, robotics
- Information science
- Computing tech development
- Social sciences
- Philosophy
- Law

Public meetings exploring ethical & societal impact issues

- Criminal and civil justice
- Public governance
- Work & labor
- Healthcare
- *Computing industry research*
- *Federal research sponsors*

Chapter 2: Engaging Expertise in Ethics and Sociotechnical Systems

- **Of fundamental importance:** The report and its recommendations carry no expectation that computer scientists & engineers will become experts in these areas of scholarship

- Effective engagement requires that computing researchers and scholars in the humanities and social & behavioral sciences
  - Acquire familiarity with each others' methods and approaches
  - Appreciate the value of different approaches in shared efforts toward identifying and addressing challenges of responsible computing

Ethics: What Matters (and What Doesn't)?

- Principles for developing and deploying computing technologies have proliferated
  - Important as goals; do not suffice for guiding responsible computing research
  - Divorced from practice, lack sufficient explanations of underlying assumptions and of ethical reasoning roots
- Theories (deontology, consequentialism): for philosophy, not practice
- This report roots its treatment of ethics in **ethical values**
  - Fundamental building blocks of moral theories
  - Conflicts among values are source of many ethical challenges
  - Analysis and decision making about trade-offs among them is a critical skill
- No checklist or other shortcut to considering values in context of use

Sociotechnical Perspective and Social Science Methods

- ***Perspective***: Computing research creates material, technical, and social worlds with which people interact and which affect them Design decisions in "purely technical" work, too
- ***Methods***: social theories, social scientific methodologies, & empirical observations and methods
- Contributing ***disciplines***: anthropology, information science, education, ethnic studies, history, qualitative sociology, political science, public health, urban studies, & women and gender studies

Chapter 3: Identifying (Some of) the Roots of Ethical and Societal Impact Challenges

- Chapter 2's theoretical concepts and methods provide foundations
- To identify practical steps, need to understand causal roots
- Of note: Many examples cited are of the impact of subsequent technology innovation: choices by downstream product designers, deployers and acquirers
  - Even though regulation is corporate and governmental responsibility,
  - still computing researchers have obligations to reduce misuse & misinterpretation by
    - Delineating limitations
    - Specifying intended scope of applicability of methods and artifacts

Causal Roots and Inherent Limitations

- Interactions of societal settings with computing technologies
- Limitations of human capabilities and their interactions with features of technologies in societal contexts
- Societal context influences on design and deployment choices
- Insufficient attention to best practices for system robustness and best practices in design and implementation
- Limitations of purely computing-technical approaches and the need for policy and regulation to work in tandem with computing technology design and development

Interactions of Societal Settings with Computing Technologies

- Conflicting values & goals of stakeholders
- Human dignity
- *Challenges of responsibly predicting & shaping individual behavior*
- *Technological solutionism*
- ***Alignment with existing norms, structures & practices***
- Environmental externalities
- Risks of extreme events
- Organizational & social structures
  - Diversity, equity and inclusion
  - Insufficient training in addressing ethical & societal issues

Social Factor: Aligning with Existing Norms, Structures & Practices

- Work and labor settings: varying values and power
  - Efficiency vs. employee burdens (work and home), surveillance
  - EHRs: data for the future vs. clinical practice
- Health care
  - Shift to telemedicine during pandemic was especially challenging for vulnerable children (foster care, unstable home situation)
    - Patients and their caregivers often lacked Internet-connected devices, the main delivery vehicle!
    - EHR and other systems presumed ready access to parents/guardians as point of contact for care and to provide consent

Limitations of Human Capabilities

- Cognitive complexity of oversight
  - "Human in the loop" has too large cognitive burden
  - Sullenberger: "requires much more training and experience, not less, to fly highly automated planes" [*Wired*, 2021]
- Pro-automation bias & automation aversion
  - Pro-bias: system designs and responsibility incentives lead people to defer to computer despite counter evidence
  - Aversion: system mistakes, people's own expertise
- Opaque systems
  - Data-intensive AI systems
  - Any system with human-computer interaction design flaws, inadequate training or training materials

Limitations of Human Capabilities (cont'd)
Designing for Open Worlds

- As computing moves into the physical and social world, computing researchers have limited knowledge about the situations in which a system will operate
- Autonomous vehicles: usage in unanticipated situations
  - Failures in unanticipated situations leading to pedestrian deaths
  - Driver over-reliance
- Bluetooth use for medical devices:
  - "SweynTooth" involved unexpected interoperability issues and inadequate testing
  - Wide array of medical devices at risk

Societal contexts affecting design & deployment

- Ideation and Design
  - **Specifying intended uses & functions**
  - Training & benchmark data
  - Defining objective functions
  - **Engaging relevant stakeholders**
  - Integrating computing & domain expertise
- Deployment
  - Institutional pressures
  - System characteristics
    - Continuous integration & continuous deployment
    - Validation
  - Appropriate system use
    - Mission, function & scale creep
    - Strategic behavior
  - Societal responsibilities
    - Disparate access
    - Governance principles for new technologies

Specifying Intended Functions and Uses of Research and Systems

- Many examples of faulty descriptions causing misperceptions of capabilities:
  - **Large language model**s described as “learning a language” rather than “learning large-scale statistics of word co-occurrence”
  - **Dynamic employee scheduling** software described as “empowering employees,” but actual design instead empowers employers
  - Description of **facial recognition systems** as “computer vision” implying relative domain- and data-independence, but actual systems are highly dependent on particular training data sets.

Engaging Relevant Stakeholders

- Contrast design of SepsisWatch, an early warning system for sepsis, with EHR design
  - Interdisciplinary team of researchers studied existing systems and their stakeholders, gleaned expert knowledge of the ways they handled sepsis cases and their workflows
  - Effectively engaged nurses and other “on the ground” stakeholders

System Robustness
Limits of Purely Computing Technical Solutions

Underlying causes: Failure to follow best practices for design & implementation

- Trustworthy, secure & safe systems
- Software engineering
- Data cleaning & provenance tracking
- Designing for responsibility

Limits of a Purely Computing-Technical Approach

- Privacy
- Content moderation

Recommendations in Brief

1. Reshape computing research to adequately include ethical and societal impact concerns
2. Foster & facilitate computing research aimed at furthering societal benefit and mitigating harm
3. Support the development of expertise in responsible computing in students and computing researchers in CS&E, social sciences and humanities
4. Provide researchers access to the knowledge and expertise needed to identify & assess the ethical & societal implications of their work
5. Integrate ethical and societal considerations into computing research sponsorship
6. Integrate ethical and societal considerations into publications and publicly released artifacts
7. Adhere to best practices for systems design, deployment, oversight & monitoring
8. Support engagement with the public and the public interest

1. Reshape Computing Research

What: ensure that ethical and societal consequences are considered and addressed appropriately from the start

- Include participants with expertise in
  - Social & behavioral sciences and ethics
  - In any domains of application involved
- Involve relevant stakeholders throughout the research
- Report possible limitations and downstream risks of artifacts, algorithms and other computing methods

Who: Computing research community, researchers

2. Foster and Facilitate New Types of Research

What: Foster and encourage responsible computing research

- Research and support for new kinds projects that
  - Could lead to societal benefits and ethical societal impacts
  - Could help avoid or mitigate negative outcomes and harms
  - Might involve new types of partnerships with companies and philanthropies
- Provide sufficient resources for the participation of scholars from fields other than CS&E and of stakeholders
- Performance review processes and tenure & promotion committees that value scholarship on the ethical & societal impacts of computing research

Who: Research community, research sponsors and research institutions

3. Develop student & researcher expertise in responsible computing

What: Better equip students, computing researchers, and faculty to address the ethical and societal impacts of computing

- Enable students by
  - Reshaping computer science and engineering curricula
  - Reshaping curricula in social & behavioral sciences and the humanities
  - Supporting graduates' gaining abilities to assist public and private sector institutions in making better decisions
- Enable computing researchers & faculty to
  - Gain proficiency in carrying out and assessing responsible computing research
  - Share best practices for research, publication and proposal review
  - Evaluate effectiveness of different approaches

Who: Universities, scientific/professional societies, research sponsors

4. Access to expertise: ethical & societal implications

What: Provide researchers access to the knowledge and expertise needed to identify & assess the ethical & societal implications of their work

- Assist computing researchers in finding scholars with the ethical, societal impact & domain expertise their projects require
- Support such scholars in collaborating in the research
- Support the development & sharing of educational materials and descriptions of best practices

Who: Research institutions, scientific societies, research sponsors

5. Ethical & societal considerations in proposals & projects

What: Integrate ethical and societal considerations into computing research sponsorship

Not a separate section

- Research proposals should integrate ethical and societal considerations into project description
- Project review panels should have
  - Requisite expertise to evaluate ethical & societal considerations
  - Appropriate evaluation criteria for such considerations
- Require project reports to address ethical and societal issues

Who: Research sponsors

6. Ethical & societal considerations in publications

What: Ensure that ethical and societal considerations are integrated into publications and publicly released artifacts

- Establish criteria & metrics for assessing treatment of ethical &societal impacts
- Establish criteria for determining whether and how to release artifacts
- Provide guidelines for authors and reviewers to meet/assess criteria
- Encourage researchers to:
  - Report unanticipated ethical or societal consequences of research
  - Provide guidance to future researchers interested in using the results of their research

Who: Conferences and journals

7. Adhere to Best Practices

What: Adhere to best practices for systems design, deployment, oversight & monitoring

- Best practices for: accessibility, integrating with organizational practices, involving diverse expertise & stakeholders, ensuring security and privacy, identifying potential unanticipated uses, mitigating harm
- Transparency about the capabilities, maturity & limitations of any artifacts produced
- Documenting design assumptions

Who: Computing researchers

8. Engage with the Public

What: Support engagement with the public and the public interest

- Engagement activities include
  - Informing the public about emerging technologies
  - Assisting public and private sector acquirers of computing technologies
  - Bringing potential adverse consequences of emerging technologies to the attention of governments and other public organizations
- Enable engagement by preparing computing researchers to serve effectively in advisory capacities

Who: Universities, research sponsors, and scientific societies

Responsibilities for Recommended Actions (Recommendations 1-3)

Researchers & research community	Reshape computing research to integrate ethical, behavioral & social science expertise
Research sponsors & research institutions	Provide support that enables research community to broaden scope and define new kinds of projects & partnerships
Academic Institutions	Reshape curricula in computing, social & behavioral sciences, and humanities
Scientific & professional societies and research sponsors	Provide & support training in designing, carrying out and evaluating responsible computing research

Responsibilities for Recommended Actions (Recommendations 4-8)

Research institutions, research sponsors, scientific & professional societies	Provide computing researchers access to scholarly expertise in ethics, social & behavioral sciences
Research sponsors & scholarly publishers	Vet computing research and assess adequacy of consideration of ethical & societal impacts
Computing researchers	Follow best practices for developing systems and releasing computing artifacts
All actors	Support better public understanding of computing research & its outcomes

Computing Research Ecosystem: Responsibilities Beyond the Research

Innovative technologies draw on multiple research results

Design of new technologies involves many tradeoffs (values, incentives)

Science and engineering are necessary but not sufficient

Scientists & engineers: provide decision makers information they need

Societies and communities	• Determine norms
Governments	• Instate mechanisms to realize or enforce norms
Entrepreneurs, investors, and corporate leaders	• Shape incentives & establish governance mechanisms

NATIONAL ACADEMIES Sciences Engineering Medicine

Final observations

- Computing researchers cannot eliminate every ethical or societal problem but can be proactive in identifying risks and avoiding potential harms by broadening
  - The scope of computing research per this report's recommendations
  - The assessment of computing research to include not only performance analysis and mathematical advances, but also evaluation of potential ethical issues and societal impacts
- Intended downstream impacts of recommendations
  - Provide model for other researchers, technology developers & deployers adopting the research
  - Ensure future computing professionals across industry, not just in research, are better equipped to address ethical and societal concerns

Ethical and societal impact considerations must be first-order concerns beyond as well as in research.