AI/ML Ethics Workshop

Workshop 1 – July 2022

Workshop goals

- Appreciate current AGU research ethics policy
- Review the current state on AI/ML ethics in research
- Anticipate AI/ML ethics stakeholder “pulse” survey data
- Review selected case examples of AI/ML research with ethical implications
- Establish AI/ML ethics working groups
- Conduct a “pre-mortem” to anticipate what could possibly go wrong
- Anticipate next steps for Workshop 2

Workshop design – Day 1 (ET US times)

12:00 Welcome and Overview

12:15 Overview of current AGU research ethics policy

- Billy Williams, Executive Vice President of Diversity, Equity and Inclusion, AGU

12:30 Lighting talks on current state of AI/ML ethics in research

- Thomas Donaldson, University of Pennsylvania
- David Gagne, University Corporation for Atmospheric Research

1:15 Anticipating results from AI/ML ethics stakeholder "pulse" survey data

1:30 Dialogue on gaps and issues

1:45 Break

2:00 Brainstorming themes for working groups, potentially including:

- Working Group 1: *Transparency/Reporting*
- Working Group 2: *Replicability/ Explainability*
- Working Group 3: *Risk/Bias/Impacts*
- Working Group 4: *Outreach/ Training/Leading Practices:*
- Working Group 5: *TBD*

2:30 Breakout groups by themes

3:30 Working Group reports

4:00 Adjourn

Opening talks

Billy Williams, Executive Vice President, DEI, AGU

David Gagne, University Corporation for Atmospheric Research

Thomas Donaldson, University of Pennsylvania

Anticipating stakeholder “pulse” survey results

Establishing/implementing ethical standards

- Establishing **clear ethical standards and guidelines** for the use of AI/ML in research.
- Researchers having **sufficient knowledge of what AI/ML algorithms are designed to do** in research.
- **Implementing/ensuring compliance** with ethical standards and guidelines for the use of AI/ML in research.
- **Researchers being able to use AI/ML in any way they find appropriate**, without being limited by any ethical standards or guidelines.

Interested parties/stakeholders

- **Knowing who are the interested parties** likely to be impacted by the use of AI/ML in research.
- **Interested parties associated with research involving AI/ML having sufficient knowledge and input** into what the algorithms are designed to do.

Explainability/replicability

- **Ensuring explainability/ interpretability** when AI/ML is used in research.
- **Ensuring replicability** in science when AI/ML is used in research.

Anticipating stakeholder "pulse" survey results (cont.)

Potential Bias, Risk, and Harm

- Having/developing tools and methods to **audit AI/ML results for potential biases.**
- Having/developing tools and methods to **assess the risks** when it comes to the use of AI/ML in research.
- **Clarifying who is responsible for any harm** that results from recommendations or findings based on the use of AI/ML.
- Guidance on the use of AI/ML directly or indirectly with **sovereign data in tribal communities and/or with respect to vulnerable populations**.

Workforce development

- **Teaching students (undergraduate and graduate)** about the ethics of AI/ML when used in research.
- AI/ML not just automating human tasks but **augmenting/extending human capabilities**.

Implications for science

- Increasing understanding of how AI/ML are **changing power dynamics in society.**
- Increasing understanding of how AI/ML are **transforming research** in science, engineering, the humanities, and other domains.

Respondents to date (n=60)

Role	%	n
Researcher who uses AI/ML in research	38.3%	23
Researcher who does not use AI/ML in research, but is knowledgeable about the technologies	21.7%	13
Researcher who does not use AI/ML in research and is not knowledgeable about the technologies	10.0%	6
Research Computing and Data Professional	11.7%	7
Student (graduate or undergraduate)	13.3%	8
Administrator/leader in university	6.7%	4
Administrator/leader in government	5.0%	3
Administrator/leader in government contractor	3.3%	2
Administrator/leader in commercial organization	5.0%	3
Administrator/leader in not-for-profit organization	1.7%	1
Other - Write In	20.0%	12

Potential working group topics

- **Working Group 1: Transparency/Reporting:** Transparency/reporting on uncertainties with AI/ML ethics in research
- **Working Group 2: Replicability/Explainability:** Ensuring replicability/ explainability with AI/ML ethics in research
- **Working Group 3: Risk/Bias/Impacts:** Identifying risks, bias, intended and unintended consequences with AI/ML ethics in research
- **Working Group 4: Outreach/Training/Leading Practices:** Collecting innovations in training, professional development, and outreach at all career stages

Additional Working Group options:

- **Organizational/Society Guidance:** Guidance for organizations in establishing and administering AI/ML ethics policies, including codes of conduct, principles, and other categories
- **Participatory Methods/Domain Expertise:** Inclusive research design and conduct with AI/ML – ensuring voice for diverse communities, domain expertise, and context
- **Public Trust in Science:** Issues of believability in science with AI/ML

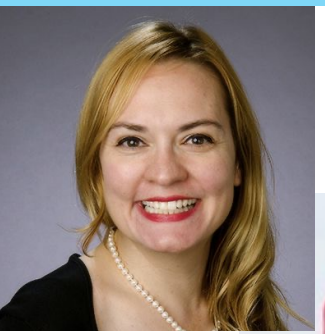
Working group process

- Introductions (30 seconds each) (7-10 min.)
- What is “in” for this topic? What is “not in” for this topic? (15-20 min.)
- Mission statement/vision for your working group (20-30 min.)

Workshop design – Day 2 (ET US times)

10:00	Welcome, Overview, and Check-in
10:30	Case example 1: Christine Kirkpatrick and Kevin Coakley, San Diego Supercomputing Center, with Discussion
10:50	Case example 2: Yuhan Douglas Rao, North Carolina Institute for Climate Studies, with Discussion
11:10	Case example 3: Micaela Parker, Academic Data Science Alliance, with Discussion
11:30	Ethical Language that is Interoperable and Extensible
12:00	Lunch Break
12:30	Working groups
1:30	Pre-mortem (what could possibly go wrong?)
1:45	Next steps prep for Workshop 2
2:00	Adjourn

Case examples



Case example 1: Christine Kirkpatrick and Kevin Coakley, San Diego Supercomputing Center

Case example 2: Yuhan Douglas Rao, North Carolina Institute for Climate Studies

Case example 3: Micaela Parker, Academic Data Science Alliance

Interoperable and extensible formatting

Formatting option:

- Elements of an **AI/ML Code of Ethics**
  - **Principles**
  - **Responsibilities**
  - **Definitions**
- Recommended **Leading Practices**
  - Consider organizing around 2-3 framing questions (FAQs) beginning with “How do we…”

    For example:
    - *How do we assess AI/ML training data for potential biases?*
    - *How do we protect against bias when there is no “ground truth” data set for an AI/ML model?*
    - *How do we ensure AI/ML ethics alignment across professional societies, government agencies, commercial organizations, private foundations, universities, and other key stakeholders?*
- **Use Cases/Illustrative Examples**

Working group process

Working Group 1: Transparency/Reporting
Working Group 2: Replicability/Explainability
Working Group 3: Risk/Bias/Impacts
Working Group 4: Outreach/Training/Leading Practices
Working Group 5: Organizational/Society Guidance
Working Group 6: Participatory Methods/Domain Expertise
Working Group 7: Public Trust in Science

Process:

- Brainstorming on potential elements of recommended language (30-40 min.)
- “Testing” potential recommended language against case examples (10-15 min.)
- Organizing work between workshop 2 and workshop 2 (7-10 min.)

Workshop 2

Day 1:

10:00	Welcome, Overview, and Check-in
10:30	AI/ML ethics stakeholder “pulse” survey results
11:00	Working Groups meetings in breakout rooms – finalizing proposed language
12:00	Lunch Break
12:30	Working Group 1 Draft Language
12:40	Working Group 2 Draft Language
12:50	Working Group 3 Draft Language
1:00	Working Group 4 Draft Language
1:10	Working Group 5 Draft Language
1:20	Working Group 6 Draft Language
1:30	Working Group 7 Draft Language
1:40	Integrative Comments
2:00	Adjourn

Day 2:

10:00	Welcome, Overview, and Check-in
10:30	Future of AI/ML and Ethical Implications
11:00	Implementation Considerations *Tracking Impact and Feedback on Progress, Including Leading and Lagging Indicators*
12:00	Lunch Break
12:30	Briefing of AGU, NASA, and Other Leadership, with Discussion
1:45	Concluding Comments
2:00	Adjourn