Unpacking the Expressed Consequences of AI Research in Broader Impact Statements

Priyanka Nanayakkara, Jessica Hullman, Nicholas Diakopoulos
Northwestern University

@priyakalot
@JessicaHullman
@ndiakopoulos

Abstract

The computer science research community and the broader public have become increasingly aware of negative consequences of algorithmic systems. In response, the top-tier Neural Information Processing Systems (NeurIPS) conference for machine learning and artificial intelligence research required that authors include a statement of broader impact to reflect on potential positive and negative consequences of their work. We present the results of a qualitative thematic analysis of a sample of statements written for the 2020 conference, and offer perspectives on how the statement can be implemented in future iterations to better align with potential goals.

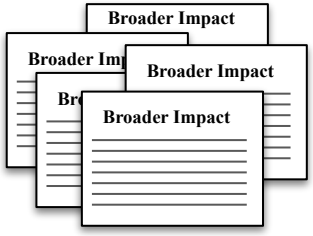
NeurIPS 2020 Broader Impact Statement

Authors were given the following official guidance about the broader impact statement (and limited further guidance):

"In order to provide a balanced perspective, authors are required to include a statement of the potential broader impact of their work, including its ethical aspects and future societal consequences. Authors should take care to discuss both positive and negative outcomes."- NeurIPS Call for Papers (highlights added)

Methods

We conducted a thematic analysis of broader impact statements. After analyzing 300 statements, we reached saturation.



We read through statements, pulling out relevant quotes to form conceptual clusters. Throughout this process, we updated groupings (and created new groups) as necessary. Additionally, we wrote memos to further understand and develop themes. All three authors met periodically to discuss groupings and to make updates as needed.

Results

We find the following themes/dimensions in our sample. These themes provide a conceptual framework for further thinking about broader impact statements.

Impacts

- Expression of Impacts
  - Valence
  - Orientation
  - Specificity
  - Uncertainty
  - Composition
- Types of Impacts
  - Privacy
  - Labor
  - Environment
  - Media
  - Bias
  - Efficiency
  - Generalizability
  - Democratization
  - Robustness & Reliability
  - Accuracy
  - Interpretability
- Relationship to Theoretical Work
- Who is Impacted?
- Timeframe of Impacts

Author Recommendations

- Outcomes
  - Safe & Effective Use of AI
  - Ensure "Fair" Outcomes
  - Protect Privacy
  - Reduce Environmental Impact
- Who is Responsible?

Additional details about the above themes (e.g., specific clusters of statements that fall into sub-themes) can be found in the full paper.

Discussion & Conclusion

Depending on the intended outcomes of the broader impact statement, our findings can inform future guidelines.

1. If the goal is to encourage reflexivity, it could be helpful to provide authors with guidance for how to map technical contributions to societal consequences.
2. To help researchers determine whether they need to change research directions in light of potential negative consequences, it could be helpful to encourage researchers to write broader impact statements early in the research process.
3. Finally, if the goal is to help minimize recklessness and negligence, it could be beneficial to provide researchers with guidance around specific ethical issues in their area of work to help them better identify/minimize risks.