What is a Research Ethics Statement and Why does it Matter?

Saif M. Mohammad
Senior Research Scientist, National Research Council Canada

Saif.Mohammad@nrc-cnrc.gc.ca  @SaifMMohammad

EMNLP 2020 Plenary Panel Discussion:

Emily M. Bender, Rosie Campbell, Allan Dafoe, Pascale Fung, Meg Mitchell, Saif M. Mohammad
As NLP and ML systems become more ubiquitous, their broad societal impacts are receiving more scrutiny than ever before. Several high-profile instances have highlighted how technology will often lead to more adverse outcomes for those that are already marginalized.

**What part do we play in this as researchers?**
What are the hidden assumptions in our research? What are the unsaid implications of our choices? Whose voices are we amplifying? (and whose we are not?) Are we perpetuating and amplifying inequities or are we striking at the barriers to opportunity?

Answers are often complex and multifaceted. **Ethics statements** can help navigate research choices, communicate implications.
Authors will be allowed extra space after the 8th page for a broader impact statement or other discussion of ethics. The NAACL review form will include a section addressing these issues and papers flagged for ethical concerns by reviewers or ACs will be further reviewed by an ethics committee. Note that an ethical considerations section is not required, but papers working with sensitive data or on sensitive tasks that do not discuss these issues will not be accepted. Conversely, the mere inclusion of an ethical considerations section does not guarantee acceptance. In addition to acceptance or rejection, papers may receive a conditional acceptance recommendation. Camera-ready versions of papers designated as conditional accept will be re-reviewed by the ethics committee to determine whether the concerns have been adequately addressed. Please read the ethics FAQ for more guidance on some problems to look out for and key concerns to consider relative to the code of ethics.
Ethics Policy

Authors are required to honour the ethical code set out in the ACL Code of Ethics.

The consideration of the ethical impact of our research, use of data, and potential applications of our work has always been an important consideration, and as artificial intelligence is becoming more mainstream, these issues are increasingly pertinent. We ask that all authors read the code, and ensure that their work is conformant to this code. Where a paper may raise ethical issues, we ask that you include in the paper an explicit discussion of these issues, which will be taken into account in the review process. We reserve the right to reject papers on ethical grounds, where the authors are judged to have operated counter to the code of ethics, or have inadequately addressed legitimate ethical concerns with their work.
what goes into Ethics Statements (ethical considerations)

... not “Appendix material”
... not just “good to have”
... not something we have never seen before
what goes into Ethics Statements (ethical considerations)

... central to our work
... things we have always seen in good work
(usually sprinkled across various sections of a paper)
What is a good place to talk about ethical considerations?

- Introduction/Motivation → Impact Statement
- Related Work
- Data → Data Statement
- Methodology
- Evaluation → Bias mitigation, Ethics focused shared task (e.g. SemEval 2018 Task 1)
- Experiments
- Error Analysis → Traditionally, a place where people have listed ethical considerations
- Limitations
- Conclusions
- Future Work

Ethics Statements can bring together the ethical considerations stated in the paper in a cohesive narrative, and elaborate on them.
Ethical Considerations: Introduction/Motivation

- what questions are we asking?
  - no hurry to get to the solution
  - let's understand the question better
    - bring out the nuances and complexities
    - what assumptions are we making?
- why should we care about this question/problem/task?
- who is impacted by this problem?
- who is impacted by this work?
- who is left out?
- what problems are we not tackling?
- are these choices maintaining structural inequities or questioning them?
Ethical Considerations: Related Work

Whose voices are we amplifying?

NLP is actively encroaching on other fields:
- humanities
- psychology
- culture studies
- social sciences
- public health

We must give researchers from these fields a voice. Learn from them. Situate our work in their literature. Collaborate with them.

That does not mean they have all the answers.

Your NLP background gives you unique perspective.

Bring it to bear by collaborate with all stake holders (including the people affected). Avoid the trap of AI/ML/NLP solutionism.
Ethical Considerations: Related Work

Whose voices are we amplifying?

NLP is actively encroaching on other fields:

- humanities
- psychology
- culture studies
- social sciences
- public health

**COI:** The incentives for fast science act against the careful and thoughtful pace of slow (truly interdisciplinary) science that engages with all stakeholders right from the start.

We can benefit from both slow and fast science.

We must give researchers from these fields a voice. Learn from them. Situate our work in their literature. Collaborate with them.

**That does not mean they have all the answers.**

Your NLP background gives you unique perspective.

Bring it to bear by collaborate with all stakeholders (including the people affected). Avoid the trap of AI/ML/NLP solutionism.
Ethical Considerations: Data

A whole panoply of considerations. Here is one:

For annotations, is there a “right” answer and a “wrong”?

- Yes: domain experts annotate the data
- No: we want to know how people perceive this word, phrase, sentence, etc.
  - large number of annotators
  - seek appropriate demographic information (respectfully and ethically)

How should we aggregate the information?

- Acknowledge the limitations of majority vote aggregation
- Danger of saying that the views of a certain demographic is the norm or standard
- Acknowledge that we are missing out on some/many voices
- Saying all voices are correct has its own problems
  - How to address and manage inappropriate biases?
Ethical Considerations: Design and Methodology

People have a bias towards large numbers.
AI amplifies the bias to large numbers.
-- Jutta Treviranus (Expert on inclusive design)
We Count! https://www.youtube.com/watch?v=OAXmCAqZqRk

Pareto principle, 80/20 rule, Zipf's law, power law distribution

Multi-variate scatter plot of needs of a set of people. Source: Jutta Treviranus.
Yes, design is political. Because design is labor, and your labor is political. Where you choose to expend your labor is a political act. Who we omit from those solutions is a political act.

A designer does not believe in edge cases.

This job isn’t about creating bullhorns for fascists and others who’d use their power to denigrate others. It’s about making sure those who are threatened by the inhumane have the better bullhorns.

This job isn’t about building tools that hand our data to the corporations of Silicon Valley. It’s about building tools to keep that data from them.

-- Mike Monteiro
Ethics Statement: Tips

- **Acknowledge**
  - the pain of those affected
  - your biases and conflicts of interest
  - how you have / have not involved various stake holders
  - you cannot capture everything

- **Invite feedback and things to add**
  - blog posts, open review, preprints
  - live document

- **Space limitations**
  - put it on the project webpage
NLP Scholar Project

Aspects of Analysis
- The analyses presented in *The State of NLP Literature* post cover only some aspects of the literature. Prior work has explored other aspects such as citation link analysis, co-author networks, influence, types of citations, etc. Yet, several interesting questions remain unexplored.

Accessing Information about the Papers
- Google does not provide an API to extract information about the papers. Martina-Martin et al. (2018) and others have pointed out that this is likely because of agreements with publishing companies that have scientific literature behind paywalls. The ACL Anthology is in the public domain and free to access. We extracted citation information from Google Scholar profiles of people who published in the ACL Anthology. This is explicitly allowed by their requests and exclusion statements, and how past work has studied Google Scholar.

Citation Information from Google Scholar
- Google Scholar is used widely in research. However, it has received criticism regarding the amount of citations, reducing academic worth to citations and h-index, etc. (see criticisms of the Citation System and Google Scholar in particular. How Has Google Scholar Changed Academia? & reasons why Google Scholar isn't as great as you think it is).

- Some papers published in a particular field of science have a mass of citations, and it is likely that the work done in the past is under-represented (authors who left academia or retired may be less likely to create a Google Scholar profile). Nonetheless, we do not expect this to significantly impact the information drawn from the analyses presented, as we do have citation information for over 35,000 papers.

Aligning Information in AA and Google Scholar is Tricky
- They do not have a common paper id or author id.
- Occasionally two papers have the same title.
- The same author may use different names in different articles.
- Multiple authors might have the same name.

We use the paper title and publication year combination as the unique identifier for a paper. However, there are some papers of papers that share the same title and year of publication. These are omitted.

New Papers are Constantly Added to AA.

The current instantiation of NLP Scholar is based on the papers in AA as of June 2019. We will update NLP Scholar with new AA information periodically.

Papers Receive More Citations with Time.

The current instantiation of NLP Scholar is based on the citation papers received as of June 2019. We will update NLP Scholar with new citations information periodically.

Rich get Richer

Visualizations in NLP scholar present papers with more citations more prominently than papers with fewer citations. This can have the effect of making more cited papers even more cited. (This is not true for Google Scholar, which ranks papers by relevance and citation counts. Citations are one (somewhat noisy) indicator of the amount of impact a paper has had. However, it is not an ideal measure. Papers get cited for a number of other reasons as well, and it is entirely possible that some of papers of interest might be those that are less cited. There are several ways in which NLP Scholar can cast light on less cited papers too. Here are some examples:

- By showing the papers on a timeline, one can easily track papers that influenced a high-citation paper in an area.
- When searching for papers in an area, one can compare citations of papers within that area. This provides a target paper in a more appropriate context. For example, a target paper may not have had hundreds of citations, but one can see that within the area of research, it is one of the most highly cited papers.
- The Languages visualizations highlight work in various languages.

Search based on Words in Titles
- Even though there is an association between terms and areas of research, the association can be strong for some terms. I use the association as an imperfect source of information about areas of research. This information may be combined with other sources of information to drive more robust conclusions. Planned future work on allowing searches for terms in abstracts and whole papers, as well as finding documents related to a query term based on word embedding based document representations will allow the current limitations.

Demographics
- Data is often a representation of people (Zook 2017). This is certainly the case here and we acknowledge that the use of such data has the potential to harm individuals. This work has several limitations, and some have ethical considerations in terms of who is left out. Future, while the methods used are not new, their use metrics reflection.
- Analysis focused on women and men leaves out non-binary people. Not disaggregating cis and trans people means that the statistics are largely reflective of the more populous cis-dich. We hope future work will explore gender gaps between non-binary — trans — cis. Similarly, tracking the skew in authors of diverse income experiences, and abilities is also crucial. This work does address these but hopefully more work on those will follow.

The use of female- and male gender associated names to infer population level statistics for women and men, can reinforce harmful stereotypes and is exclusionary to people that do not have such names, to people from some cultures where names are not as strongly associated with gender, and trans people who have not been able to change their name.

Since the names dataset used is far American children there is lower representation of names from other nationalities. However, many names are common in more than one country, and the large immigrant populations in the US means that there still exist substantial coverage of names from around the world.

- Chinese names (especially in the romanized form) are not good indicators of gender. Thus the method presented here disregards most Chinese names, and the results of the analysis do not apply to Chinese researchers with Chinese names.

- Some might argue that names partially address the gender inclusiveness guidelines listed in (Keys 2018): names can be changed to indicate (or not indicate) gender, people can choose to keep their birth name or change it, and the name, more so than appearance, can be independent of physiology. However, changing names can be quite difficult. Also, names do not capture gender fluidity or contextual gender.
Conflict of Interest

The Grey Hoodie Project: Big Tobacco, Big Tech, and the threat on academic integrity

Mohamed Abdalla
msa@cs.toronto.edu
Centre for Ethics, University of Toronto

ABSTRACT

As governmental bodies rely on academics’ expert advice to shape policy regarding Artificial Intelligence, it is important that these academics not have conflicts of interests that may cloud or bias their judgement. Our work explores how Big Tech is actively distorting the academic landscape to suit its needs. By comparing the...

Parallels between Big Tech and Big Tobacco: histories, actions

- appearing to be embattled, hiring (researchers) and funding (universities, conferences),
- taking ownership of ethics, influencing research questions

Tobacco Free Initiative (TFI)

Article 5.3 of the WHO Framework Convention on Tobacco Control

In setting and implementing their public health policies with respect to tobacco control, Parties shall act to protect these policies from commercial and other vested interests of the tobacco industry in accordance with national law.
Resources

- **Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science.** Emily M. Bender and Batya Friedman.
- **Datasheets for datasets.** Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford.
- **Some of my ethics statements:**
  - Practical and Ethical Considerations in the Effective use of Emotion and Sentiment Lexicons.
  - NLP Scholar Project: https://medium.com/@nlpscholar/about-nlp-scholar-62cb3b0f4488

**Contact:**  
✉️ Saif.Mohammad@nrc-cnrc.gc.ca  
🐦 @SaifMMohammad