

An aerial photograph of a city, likely Washington D.C., showing a mix of modern and older buildings, streets, and green spaces. A large, dark green banner with white text is superimposed over the top half of the image. Below the banner, a white box with a red border contains text and a list of bullet points. The background image shows a prominent white dome building in the center, surrounded by other structures and a road with traffic.

Takeaways @ NAACL 2021

Program Co-Chairs:

Dilek Hakkani-Tur, Luke Zettlemoyer, Anna Rumshisky

- Balancing short vs. long papers [\[see Opening Remarks\]](#)
- Load-balancing across areas [\[see Opening Remarks\]](#)
- Takeaways from ethics review ←

Ethics Review Process

New for NAACL this year:

- Emily M. Bender and Karën Fort co-chaired. 36 committee members.
 - Created excellent new FAQs and training materials
- ~150 papers were flagged by reviewer or ACs, and sent to ethics committee for further review.
- Ethics chairs made recommendations to PC co-chairs, who made decisions after reviewing recommendations from Ethics chairs and reviewer comments.
- As a result of this process, 7 papers were accepted conditionally (one was ultimately rejected, 6 accepted)
- Authors received “conditional acceptance” notification that specified what conditions had to be fulfilled for the paper to be accepted into the conference.

Example: Conditional acceptance notification

Dear <author>:

On behalf of the NAACL-HLT 2021 Program Committee, we would like to inform you that the following submission has been conditionally accepted to appear at NAACL-HLT 2021, based on the completion of conditions specified by the ethics review committee, explained below:

<paper title>

Conditions for acceptance:

The authors should provide a more robust discussion of ethical considerations, including the questions of what the deployment of this technology would mean in the world. For example, what are the failure modes, who might be harmed if they fail? Is the technology equally effective across groups of people, or are marginalized populations more likely to hit the failure modes? What kinds of malicious use possibilities should regulators be aware of?

Breakdown of conditional acceptance decisions

Ethics chairs recommendations for flagged & reviewed papers

- 13 papers “conditional accept”
- 4 papers “reject on ethical grounds”

(some rejected on technical grounds)

Most of these these fell into one of three categories:

1. Insufficient thought given to possible implications of research
2. Potential violation of the terms of service and/or data owners' copyrights
3. Potentially unethical aspects of human subjects research (e.g. AMT compensation)

Challenging to make decisions

- **Insufficient thought given to possible implications of research**
 - In most cases, could addressed by a more thorough discussion in the Broader Impact section of the paper
 - But in some cases, one would need to re-think the research!
 - **Our takeaway:** this should not be done last-minute at publication time, but rather at the point when the project is designed and approved by an institutional review board

Challenging to make decisions

- Potential violation of the terms of service and/or data owners' copyrights
 - Ethics committee often pointed to one part of terms of service, authors pointed to other parts.
 - **Our takeaway:** this might be better handled by different organizations' legal teams, rather than decided on separately by each conference.

Challenging to make decisions

- **Potentially unethical aspects of human subjects research**
 - In some cases, ethics committee disagreed with the authors on whether research was ethical w.r.t. use of data, payment to annotators, etc. This was sometimes the case even when the authors stated that their research was approved by their institutional IRBs
 - It seemed in part due to the lack of opportunity for the authors to explain and clarify conditions of their research in context of ethics review, as it is often done during IRB protocol reviews
 - **Our takeaway:** this might be handled by institutional IRBs, rather than ethics committees at individual conferences.

IRB protocols vs. ethics review at conferences

- Institutional review boards (IRBs) are designed to oversee research with human subjects, and their primary purpose is to weigh costs and benefits of research to general population and to study participants. IRBs are typically composed of several people with diverse experience and training.
- IRB human subjects research protocols have to be set up BEFORE the project is executed. An IRB protocol may take weeks (and sometimes months) to develop, revise, review, and approve. How can ethics review at conferences hope to do it with the same rigour and in a well-regulated way, for multiple studies, in just a few weeks?
- Of course, IRBs sometimes do not understand full implications of AI research - but as with all other research that affects people, it's our job as a community to educate them!

Authors who do not have access to an IRB or legal team?

- Depending on the country and organization, authors' institutions may not have an institutional review board -- or it may not be governed by the same guidance
- Authors' institutions might not have a legal department
- What about unaffiliated authors?
- **Possible solution:** have ACL maintain a standing IRB that can assist authors in such cases
 - Costly, but worth it?
 - Less costly than have a team of volunteers review all flagged papers for every conference!
 - Perhaps, ACL rolling review can serve as a platform? Be clear that IRB approvals are required for any studies involving human subjects or artifacts, since they have the potential to harm people.

Discussion with Ethics Chairs

Ethics Chairs and Program Chairs discussed the process:

- We agreed that we need to have better mechanisms in place to ensure ethical oversight of NLP research and that we should encourage the practice of setting up IRB protocols for human subjects research.
- However, had slightly divergent opinions about defaulting to existing mechanisms for dealing with ethical and legal concerns.

(Please see the blog posts from the ethics chairs and the program chairs for more about the takeaways from the ethics review process)

Summary of Discussion with Ethics Chairs

- Ethics chairs argued that IRBs are answering the question about whether the research should be carried out, while ethics review is answering the question about whether this research should be published.
- PCs argued that it is essentially the same question, and that we do not need to maintain separate ethical standards for human subjects research, when one is already one maintained by the IRBs (which have the time and means to deliberate and are trained to weigh potential harms against possible societal benefits).
- However, we do need to work to make sure the standards are kept up to date with the current technology and potential for its misuse. Then the ethics review will only need to ensure that the proper process of research oversight was followed (which is feasible to do even in the short time span of a conference).