

# Yes, We Care (more)!

## Results of the 2021 Ethics and Natural Language Processing Survey

<b>Luciana Benotti</b> Universidad Nacional de Córdoba Argentina	<b>Mark Drezde</b> Johns Hopkins University USA	<b>Karën Fort*</b> Sorbonne Université/Loria France
<b>Pascale Fung</b> HK Uni. Science & Technology Japan	<b>Dirk Hovy</b> Bocconi University Italy	<b>Jin-Dong Kim</b> DBCLS, ROIS Japan
<b>Min-Yen Kan*</b> National University of Singapore Singapore	<b>Malvina Nissim</b> University of Groningen Netherlands	<b>Yulia Tsvetkov*</b> University of Washington USA

\*Contact authors and chairs: [acl-ethics-chairs@inria.fr](mailto:acl-ethics-chairs@inria.fr)

### Abstract

This report presents the results of the ACL Ethics Committee’s stakeholder survey on ethics. Over the course of a three-month period between December 2021 to February 2022, a total of 263 respondents attempted the survey. To keep the survey simple and easy-to-complete, nine short questions were asked of the anonymous participants. The results of the survey indicate that the computational linguistics and natural language processing community cares deeply about ethics, and believe the wider community is not well-informed of ethical concerns. Stakeholders prioritized the needs for appropriate reviewing and authoring guidelines, training and ethics education opportunities and resources, along with a wider dialog with the general public.

## 1 Introduction

Realising the growing importance of computational linguistics and natural language processing (CL/NLP), the Association of Computational Linguistics (ACL), a major professional organization for the advancement of the field, created an ethics committee to take charge of overseeing the role of ethics. Due to the sensitive nature of the topic, and its various interpretations, the ACL executive sought to have holistic gender and geographic representation within the committee. The resultant ACL Ethics Committee (henceforth, AEC) was formed with a co-chair and two additional at-large members from each of three primary regions of membership (Americas; Europe, Middle East, Africa; Asia and Oceania). The committee was

also gender-balanced, with five female members, and four males. The chairs were appointed for five-year terms to allow for sufficient time to execute policies, and three-year terms for the remaining six at-large members.

During one of the inaugural meetings, the AEC chairs decided to set a survey of its stakeholders as a key priority. The survey serves 1) to understand the (then current) position of stakeholders with respect to ethics and ethical responsibility, and 2) to prioritize the agenda that the AEC should follow to fulfill the needs of the stakeholders. This report serves as a summary of raw inputs as interpreted by the AEC<sup>1</sup>.

## 2 Prior Surveys

The results of two similar surveys were run previously in 2015, were published at LREC by [Fort and Coullault \(2016\)](#). These surveys — one at a national (France) level of 102 completed responses, and one at an international level with 200 completed responses — showed similar findings: that the self-selected survey participants did find ethics in CL/NLP of growing concern, that other parties (authorities) are unaware of the limitations of CL/NLP work, and that there is a need for both ethics training and education, as well as clear guidance for authors, reviewers and event organizers.

The current survey draws on this prior experience. We reprise and simplify the original questions, so that indicative longitudinal analyses is possible. The current survey additionally serves

<sup>1</sup>This version: (draft) v230714 — 14 Jul 2023. Previous versions: v230705 (initial).

to ask the community of its ethics needs, such that the AEC can set its priorities accordingly to the mandate granted by the survey results.

### 3 The Survey

The survey contains nine items, of which five are multiple choice or response questions that were accompanied by a free text field for elaboration. The survey starts with broader questions on the general perception of ethics in CL/NLP, and narrows in on specific items. We used the Framaforms<sup>2</sup> polling framework, a free tool that allows for the easy creation of anonymous online surveys. We reproduce the exact wording of each question in the listing below.

1. Please elaborate, based on your experiences, if some specific aspects of ethical reviewing in recent conferences were useful / not useful / missing.

Examples of functions that past ethical conference committees have done include: Reviewer ethical guidelines, Author ethical guidelines, Ethical reviewing feedback, Reading list for ethical material, Having an ethics checklist, among others.

2. *Multiple Choice.* Do you think it should be possible for a paper to be rejected solely on ethical grounds, through the process of ethical review?

Please elaborate.

3. *Multiple Choice.* Are there applications of NLP that should not be the topic of research?

Please elaborate below.

4. *Multiple Choice.* Many NLP algorithms and applications can be deployed off-the-shelf by non-experts. Do you think the stakeholders are aware of the limits or possible unintended harms of the tools we create?

If you have ideas on how we can improve public awareness of the risks of NLP technologies, please elaborate below.

5. *Multiple Response.* Many NLP algorithms and applications have dual use. Who do you consider to be accountable when an NLP technology is mis-used and/or causes harm?

Please elaborate below.

---

<sup>2</sup><https://framaforms.org/abc/en/>

6. What do you think are the most urgent tasks for the global \*CL ethics committee?

7. Have you participated in any class or training concerning (in any role) ethical research practices or ethics in NLP?

If Yes, please elaborate below.

8. Please provide your email address if you would like to be involved in future discussions on developing ethical guidelines / training / resources for the community.

Also, if you have any ideas on how to best involve the whole community (e.g. doing further surveys, volunteering as ethics reviewer, having dedicated tutorials), please elaborate below.

9. Any other comments you would like to share on the subject of ethics in NLP?

**Participation.** Our AEC members called for survey participation via personal social media and various mailing lists (LINGUIST, Corpora, and through the ACL's portal, which sends opt-in email notifications for CL/NLP events to its registered membership), over the duration of the survey run, with multiple reminders. After removing empty and duplicated response submissions, 254 unique responses were recorded. 83 participants left their email for further contact.

The number of survey respondents, compared to the previous 2015 international survey reported in (Fort and Couillault, 2016), are similar (cf. 200) However, since 2015, there has been a clear growth in membership in the CL/NLP community: 7,173<sup>3</sup> members in ACL as of 2021, compared with 2,536<sup>4</sup> members at the end of 2015. Our respondent pool therefore corresponds to a smaller proportion of the CL/NLP community. Taking ACL membership as an approximate proxy for CL/NLP stakeholders, the respondent rate thus covers  $254/7176 = 3.5\%$ . In contrast, the previous 2015 survey covered  $200/2536 = 7.3\%$  of the membership, so represents a wider sample of stakeholders (roughly double). Supporting evidence from the ACL Business Office reports attributes the growth in membership in growing interest in CLP/NLP but also to membership drives

---

<sup>3</sup>[https://www.aclweb.org/adminwiki/index.php/2022Q1\\_Reports:\\_Office](https://www.aclweb.org/adminwiki/index.php/2022Q1_Reports:_Office)

<sup>4</sup>[https://www.aclweb.org/adminwiki/index.php/2016Q1\\_Reports:\\_Office](https://www.aclweb.org/adminwiki/index.php/2016Q1_Reports:_Office)

Question	1†	2	3	4	5	6†	7	8	9†
# responding	144	253	253	253	251	141	252	98	64
%age responding	56.4%	99.6%	99.6%	99.6%	95.4%	53.6%	95.8%	37.3%	24.3%
# commenting	144	164	147	137	122	141	91	71	64
%age commenting	56.4%	62.4%	55.8%	52.1%	46.4%	53.6%	34.6%	27.0%	24.3%

Table 1: Response and commenting rates (per question;  $n = 254$ ). For free-text only questions (identified by †; i.e., Q1, 6 and 9), response and commenting are identical.

of the ACL in recruiting from a wider geographical demographic in emerging areas (Africa, Middle East, Asia — especially China).

Respondents to the survey were told that the survey was completely anonymous, although Question 8 allows the respondent to self-identify their response. We have archived the raw data, and duly plan to release the processed data<sup>5</sup> to interested researchers, upon suitable verification.

## 4 Methodology and Results Analysis

As all of our questions featured free-text for respondents to elaborate on their answer, it is critical to analyze the free text that accompanied such responses. To do this, we followed a three-phase protocol. In the first phase, one of the three AEC chairs performed a manual coding of the free-text comments for one question according to their own typology, which was constructed by first reviewing a small proportion ( $\sim 10\%$ ) of the comments. Subsequently, a second AEC member performed a coding of the same comments using the created taxonomy. A final phase required the two members to meet to reconcile their codings.

Table 1 gives overall statistics of the survey results. We see that over 95% of the respondents gave responses for the multiple choice or response questions, and that between 20–60% of respondents chose to elaborate on their categorical responses through comments (bottom row).

### 4.1 Question 1: Ethical Reviewing Processes: Past experiences and overall opinions

This first question aimed at collecting some background information from the survey participants with respect to their prior experiences with ethics reviewing. This was an open-response question of which 144 respondents completed. The responses

highlighted various positive and negative experiences with ethics reviewing, insights into potential gaps or obstacles in the ACL ethics policies at recent conferences, the need to define a clearer scope of functions for ethics committees, and to create a unified repository of ethics training materials, as well as suggestions for new policies to overcome current obstacles. Overall, the responses indicate a growing awareness of the importance of ethical considerations and the need for continued efforts to progress towards ethical reviewing policies and practices that are comprehensive and effective.

The majority of respondents express support for the need to conduct ethics reviewing: only 6% (10/150) respondents stated that ethics reviewing is not useful and should be abandoned; however, reported prior experiences are mixed, some respondents expressed their frustration or disagreement with existing policies and suggested directions for improvement which we summarize next.

23% of respondents (35/150) highlight the importance of *carefully defining the scope, methodological consistency, and evaluation consistency of ethics committees and their reviewing processes*. They argue that a lack of clear definitions can lead to inconsistent ethical evaluations and confusion in the community. They suggest that more rigorous definitions and processes that are well communicated can prevent pushback.

27% of responses highlight that *training materials, such as tutorials and readings, should be provided to NLP authors* to improve their understanding of ethical considerations in research. This can help authors write more comprehensive ethics sections in their papers, promoting transparency. Further, 19% argue for the usefulness of better *publicizing training materials and ethical guidelines based on the policies of the Association for Computational Linguistics (ACL) for authors, reviewers, and area chairs (ACs)*. Some argue that publicizing these guidelines can help establish a shared understanding of ethical expectations across the

<sup>5</sup>As we guaranteed participants that participation is fully anonymous, we need to scrub any identifying information in the free text responses, and to disassociate identity and email addresses (Q8) from the rest of the responses.

NLP community.

21% specifically discuss *ethics checklists*. While there is an overwhelming support for the need of ethics reviewing, the majority of respondents (22/32) argue against the mandatory checklists, viewing them as an attempt to unify a process of ethical judgment that is inherently not uniform. Some respondents highlight that ethics checklists are generic, do not cover all potential issues, and can come across as ethics-washing.

26% discussed their prior experiences with writing an Ethical Considerations section, with 21/40 stating that it is important to keep this section, rather than relying on generic checklists. Some respondents argued that the section should be mandatory. 19/40 argue for keeping the section optional, relying instead on ethics reviewers to flag papers with potential ethical concerns.

Among suggested future policies, most prominent themes include (1) the need for better handling of human subjects research, especially for authors and institutions that do not have any IRB access; (2) the need to incorporate a broader and more inclusive perspective on ethics than the current “WEIRD ethics” (Western, Educated, Industrialized, Rich and Democratic); (3) the need for an updated ACL Code of Ethics, rather than relying on the existing ACM Code of Ethics, and (4) the need to incorporate paper retraction policies.

#### 4.2 Question 2: Possible to reject a paper solely on ethical grounds

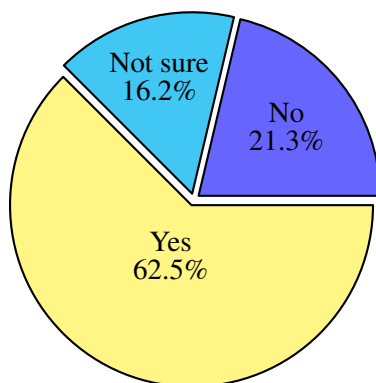


Figure 1: “Q2. Do you think it should be possible for a paper to be rejected solely on ethical grounds, through the process of ethical review?”

Respondents gave 254 multiple choice responses for Question 2, of which 159 were “Yes”, 54 “No” and 41 “Not sure” (Figure 1). 160 left

comments, explaining or nuancing their decision. From these comments, we were able to properly attribute 21 categorical responses of “No” or “Not sure” were in fact “Yes” appropriately qualified (e.g. “No,” with a comment of “It should not be rejected unless the ethical issue is extremely serious and obvious”, which strictly interpreted would mean Yes). This raises the proportion of “Yes” to more than 70% of respondents. Most comments were either about the conditions under which a paper could possibly be rejected solely on ethical grounds, examples of reasons to reject a paper, or reasons why it should not be possible to do so.

The most cited condition (32) to accept the possibility of rejection for ethical reasons is that there should be a clear decision-making process involving several peers, explicit rules and interactions with the authors before the decision is made. 22 respondents insisted in saying that this should be a rare case, reserved for extreme cases, 12 similarly said that it should be limited to egregious cases, without necessarily giving more precision. The most cited example of cases which could justify such rejection is unethical datasets (18 respondents), unethical application (14), including dual use with a much more often negative than positive outcome and research led in an unethical way or violating Institution Review Board (IRB) recommendations (13). Eight (8) respondents mentioned issues of authorship attribution, plagiarism or fraud. 17 persons insisted that this is very important for the community or, more precisely, that there is no use in having decided to apply a code of conduct if there are no sanctions.

By far the most cited reason to refuse this possibility is that this would be too subjective (21 respondents). Seven (7) people cite dual use as a reason not to reject papers on ethical grounds, four mention a fear of censorship and four suggest that this is not our job (i.e., we are unqualified to make this decision) or that we cannot afford it. Four (4) justify their refusal by saying that such research will be published anyway and suggest instead to flag the papers somehow or publish them with comments.

Seven (7) respondents explained that, to them, unethical papers are also (or mainly) poor papers scientifically. Interestingly, this was a comment from all groups of respondents (“Yes”, “No”, and “not sure”).



### 4.3 Question 3: Applications that should not be researched

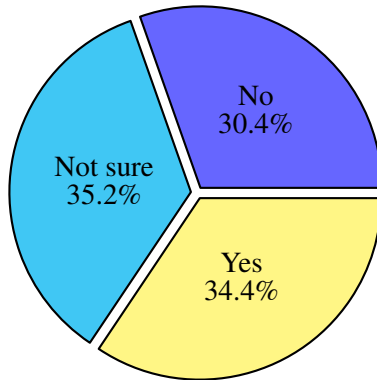


Figure 2: “Q3. Are there applications of NLP that should not be the topic of research?”

253 respondents answered Question 3, broken down as 87 “Yes”, 77 “No” and 89 “Not sure” (Figure 2). 137 of the respondents left comments. The analysis of the comments allows to recategorize 6 “No” or “Not sure” into “Yes” and one “Yes” in “No”. The comments also confirm that the community is divided on the subject, with lots of hesitations from both sides. 23 respondents insisted on the paramount importance of keeping the research open, 19 on the possible dual use of the applications (i.e. including positive use), while 23 clearly stated that applications with a negative impact on society or a minority should be banned, 25 mention predictive applications (in particular predictive policy or justice, predicting gender from texts, etc), “measuring” IQ from texts or applications with a direct impact on people’s lives. 10 would be in favor of not allowing surveillance applications, 4 mention military ones.

### 4.4 Question 4: Stakeholder Awareness

The majority of respondents agreed that many stakeholders are not fully aware of the limitations of the NLP technology and the potential harm that they could cause (Figure 3). It was also pointed out by a minority that exactly who constitutes a stakeholder and what harms are defined as was not well defined, but respondents were able to respond given this general prompt. What can be done to close this gap in awareness? Approximately half ( $130/254 = 51\%$ ) of the respondents elaborated in the free text field what they felt might help.

An analysis of respondents’ comments we feel reinforce the urgent need for a push for awareness

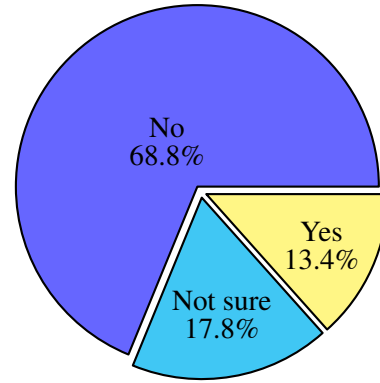


Figure 3: “Q4. Do you think the stakeholders are aware of the limits or possible unintended harms of the tools we create?”

among many stakeholders, starting with our own community and reaching all the way towards the general public. We describe major categories of comments, in which a single respondent’s comment may count towards multiple categories (multiple labels). Most comments were associated with at least one category ( $138/130 = 1.1$  categories per comment).

Four categories were mentioned the most frequent, each consisting of approximately one-fifth of all comments:

1. **Stating explicit limitations, intended uses and dual-use harms in publications (30).**

Respondents first took our own community’s artefacts to task; recommending that publications be given explicit space to document the limitations of the work and its intended use. Where significant negative repercussions may occur, respondents suggested that dual-use scenarios should be highlighted. Limitations with evaluation metrics were also highlighted as an area of limitation, as accounting for bias, misuse and energy efficiency deserve special mention.

Model cards (Mitchell et al., 2019) were suggested by some (7) as a standard accompanying artifact with the possibility of developing a set of warning badges that could be displayed alongside with open-source code repositories, analogous to a nutritional label on foods and warning advisories on tobacco products. Ethics checklists were also recommended by some as means of convenience and also as a means of priming readers towards general awareness of ethical issues.

Respondents were mixed on whether this should be a requirement for all papers, but many respondents felt that it should be required for any paper that may be controversial, either as advised by a reviewer or self-determined by the authoring team. Furthermore, there was general consensus that such text should not could towards the length limit of the article, as it was seen by many as an essential aspect.

2. **Public Outreach (29).** Secondly, respondents suggested that the community needs to work towards better contact and maintaining ties to the general public, perhaps through media intermediaries. Suggested outreach channels took many forms, from ones targeting the NLP community itself to ones that targeted the general public, mass media, and industry practitioners.

For awareness within NLP community, respondents felt that a dedicated workshops or sessions would serve to help to underscore the importance of these issues, as some opined that researchers themselves are often unaware of the scope of the risks their works may entail when brought to production. Shared activities, akin to the security area's capture the flag (CTF) or science slam were mentioned as means to bring more awareness to the community.

For general public outreach, respondents highlighted current and planned activities, but largely agreed more could be done. Proposed activities included an ACL-sponsored newsletter, audio and video podcasts, discussion on social media platforms such as Twitter or Instagram, or general interest talks could help spur public awareness. Appropriate science communicators, akin to astronomy's Carl Sagan and Neil deGrasse Tyson, were also raised as important vehicles. The emergence of such activities and spokespersons may be encouraged by the community through sanctioned public outreach awards.

3. **Publications and changes to publication models (25).** Respondents mentioned that the community's own publications may also close the awareness gap, dovetailing with general public outreach. Special tracks,

venues for explicit discussion of ethics, limitations, retrospective analyses, negative results were all raised as possibilities. Ethics reviews themselves were taken as a helpful vehicle, inasmuch as their presence (and publication) would encourage much-needed dialogue within the community. It was opined that discussion and argumentation may be worth more than the actual communicated stances or outcomes, as these may assist community members in their ethics sense-making process.

4. **Clearer Use of Language (23).** Respondents suggested the community can express the scientific results and outcomes of the science in clearer terms. This recommendation was directed largely at the perceived uptick in artificial intelligence hype, where respondents felt that industrial and media players have a disproportionate responsibility to curb their use of inflated and hyped claims to call attention to their research or influence. It was also felt that the ethics community could better define what harms and stakeholders are, such that the gap in awareness can be more precisely qualified. This issue intersects the above two issues, as clearer use of language affects both the primary scientific discourse in publications but also secondary reports from press releases and summaries that can be both purposefully or unwittingly paraphrased, sensationalized and taken out-of-context by news media and outputs that may have interests unaligned to those of the NLP/CL community.

Unfortunately, many respondents felt that this recommendation was not easily executed, specified, nor readily enforceable. Similar to the comments on public outreach, it was felt that encouraging strong, positive examples by official recognition might be a possible realization.

Several other categories of comments made up roughly 10% or less of the remaining comments:

- **Formal Course Education (15).** Some also brought up the need for formal training of the next generation of the community's scholars. Tertiary education as well as professional training needs to feature discussion and activities to promote awareness.

- **Public Examples (9).** Having public examples of limitations, biases, and ethical quandaries were also brought up as good example case studies and templates for the spectrum of possible ethical issues that could occur in contemporary and future NLP/CL research. Having a repository of such cases could better help the community's researchers draw parallels from their work to such public case studies, such that historical problems are understood and not repeated.
- **Legal standards and penalties (7).** Respondents also brought up legislative issues, surrounding technology use. Standard operating procedures for both developer and end-user agreements and licensing for NLP technology use, as well as ones for whistleblowing on unethical deployment of NLP systems, were raised as needing development to help spur responsible and mature deployment of NLP.

#### 4.5 Question 5: Moral Responsibilities

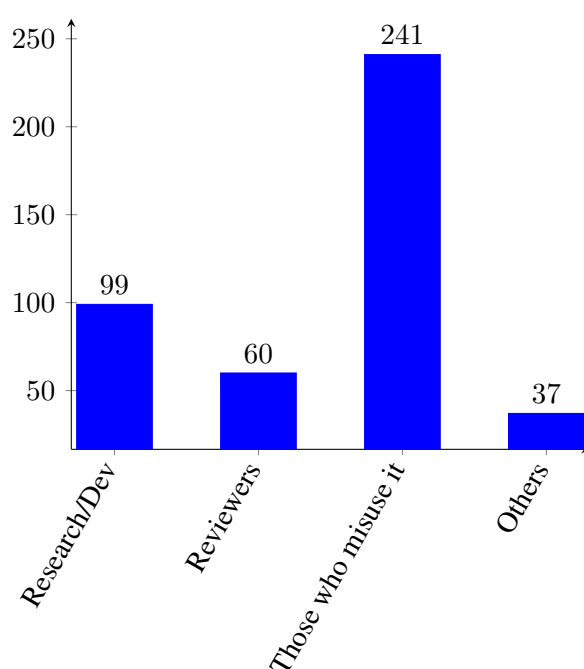


Figure 4: “Q5. Who do you consider to be accountable when an NLP technology is misused and/or causes harm?”

98% (251) of our survey respondents answered this multiple response question (Figure 4) breaks the statistics for the four checkboxes down in the question.

It was clear that most respondents felt that answering the question was impossible to simplify down to a binary yes/no responsibility; most respondents' comments agreed that the responsibility is rather a question of spectrum. As such we believe that the numbers below should be inferred less as a categorical and rather as a indicative of an underlying distribution, where certain players have a higher sense of responsibility.

Nevertheless, analyzing this multiple-response question, we find that the specific combination of checkboxes yield a good narrative analyses of the  $117/263 = 44\%$  of associated comments:

- **Those who misuse it.** Most respondents (242; 97%) felt that the end-user has the obligation of utilizing the technology researched and developed by the community in a responsible manner. Half of them (135) felt they alone should guarantee and shoulder the legal responsibilities of NLP use. This option was mostly construed as having legal responsibilities, which would necessarily be context- and legal jurisdiction-dependent.
- **Researchers and Developers.** Many respondents (99; 40%) felt that researchers and developers have a responsibility in ensuring the proper documentation, scope of use, risk evaluation, and responsible use conditions are clearly outlined. Such responsibilities were envisioned as part of the normative social contract, rather than as a clear, obligatory legal responsibility with punitive repercussions. Most of the respondents (96 / 99 = 97%) who chose this option also chose the previous option (*Those who misuse it*).
- **Reviewers.** Some respondents (60; 24%) felt that reviewers are also responsible for technology misuse. Most of them were inclusive respondents (20 chose all the options, 36 chose all but *Others*). Respondents felt that reviewers' responsibility were simplified; to raise alerts for works that may have unaccounted ethical aspects or scenarios that may apply; these should be referred to expert opinion. An ethics committee (to which referrals are destined) should be charged with a higher level of responsibility for ethics endorsement before submissions are allowed to be published.

- **Others.** Among others, many respondents highlighted responsibility by two other stakeholders: corporations and governments. *Corporations*<sup>6</sup> (7) were felt as especially needing an urgent and appropriate set of checks and balances. The rapid progress in NLP has set off many ventures — both from new spinoffs and mature tech giants — to profiteer from gains and spin up marketing efforts to attract clients and investors. Some of these efforts play counter against the important, long-term agenda of realistically calibrating societal awareness of AI and NLP’s capabilities and limitations. *Governments*<sup>7</sup> (7) were also called upon by respondents to take on a key responsibility of setting up appropriate legislation for end-user agreements, licensing and accreditation; and judicial processes for determining misuse and the scope and liability of punitive measures.

On a final note, it was felt that the question was phrased to cast responsibility and blame but that the importance of civic discourse and societal understanding and responsibility should be a central tenet of responsibility. Actors are sometimes ignorant rather than malignant, in that harms were not intended, and as such promoting and sustaining public interest and discourse as well as changes in societal norms were seen as key indicators of progress.

#### 4.6 Question 6: What do you think are the most urgent tasks for the global \*CL ethics committee?

We received 132 answers to this open-ended textual response question. In our annotation, we identified the following major categories:

- **Policies (67)** notes: - ACL code of ethics; include cross-cultural considerations - better-defined ethics review scope, and regulatory white paper defining the basic ethical standards of the NLP community - clearer guidelines - rebuttal/revision/appeal opportunities in review process, - retraction policy, - enforcement policies, - implications of violating ethical issues (e.g. for what happens to researchers/members of ACL/research institu-

<sup>6</sup>Expressed as *companies, employers, managers, executives, funders*, in the elaborated responses.

<sup>7</sup>Expressed as *governments, politicians, policymakers, legislators, lawmakers*, in the elaborated responses.

tions who repeatedly violate ACL ethics principles?) - avoid censorship - transparency, reviewing statistics

- **Training Materials (38)** notes: - ethics training curriculum - online training materials
- **Outreach (28)** - facilitating discourse about ethics in the community, allow for community feedback for defining ethical norms - a dedicated forum for discussion, dedicated tracks ("a forum where ethics discussions can be had openly so the community has a open sense of what others think, away from twitter, annual track at a journal or conference") - "working with affinity groups (e.g. Queer in AI, Black in AI, Latinx in AI, North Africans in NLP" - including/consulting with ethics/morality experts by ethics committee , policy experts
- **Large Language Models (18)** notes - chatbots, privacy - IRB, data licensing - regulations for corporate research
- **Data and Resources (7)**. issues with data/resources

#### 4.7 Question 7: Ethics training

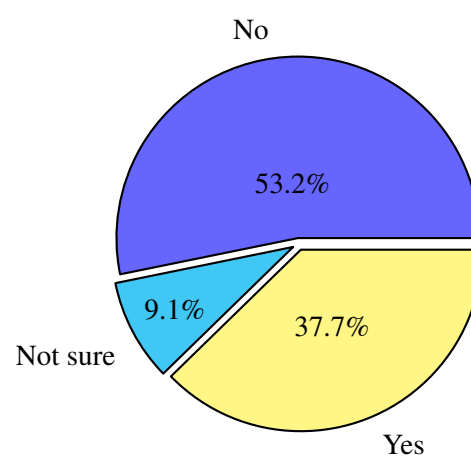


Figure 5: “Q7. Have you participated in any class or training concerning (in any role) ethical research practices or ethics in NLP?”



**4.8 Question 8: Also, if you have any ideas on how to best involve the whole community (e.g. doing further surveys, volunteering as ethics reviewer, having dedicated tutorials), please elaborate below.**

**4.9 Question 9: Any other comments you would like to share on the subject of ethics in NLP?**

We received 55 open-ended responses to the question, which we categorized into three categories: criticism, suggestions, and positive remarks. We provide some examples below (comments are pasted verbatim).

#### **Criticism.**

- Fear of censorship:
  - The ethical reviewing process should be more balanced, and should not be too stringent. We should avoid rejections based on ethical concerns, unless research contains egregiously unethical violations.
  - Ethics should be used as a tool to make people think, not to exclude/ shame or patronise.
  - I do think we need to educate both our community and end users about ethical issues associated with NLP technology. I appreciate that this is important. However, and this is a big however! I have serious concerns about our community potentially being overly judgmental and heavy-handed. We should acknowledge that our society is in the infancy of understanding ethical issues associated with AI, and there is no clear consensus on what the boundaries should be. I think that we need to RESPECT that very reasonable people can have different views on these boundaries and not enforce an overly strict view that might be held by a small set of people (e.g., those on the ethics committee at any point in time, or current \*ACL leadership, or simply those who speak up the loudest). So I'd advocate for a policy that only penalizes the most egregious cases, where (say) > 90% of people would agree. And it is VERY important that the criteria are documented and

spelled out carefully for both reviewers and authors, so that everyone is on the same page.

- The last few years have seen many changes in the submission process for research papers at \*ACL conferences w.r.t ethical issues. It has meanwhile become extremely time-consuming for authors to take these new measures into account.
- Please keep the additional workload for authors at an acceptable level. Also: please do not change the "ethics"-forms and "ethics"-criteria at every (second) conference.

- Ethics relativism – comments pertaining different ethical frameworks and cultural differences:
  - I appreciate the effort in ethics for NLP but feel like it has a very western standard currently and is not transparent about its practices in ACL conferences
  - I do not think it is ethical to limit the scope of ethics down to only the few popular ones at the moment.
  - It is unclear how to handle the situation that different people have different attitude towards morality, including ignoring it altogether. There isn't the "ethics" - different moral-philosophical positions lead to very different things that are permissible or not.

#### **Suggestions.**

- Training is needed:
  - As much as I don't like them, I think that a yearly mandatory online 30-45 mins course is the way to go. It will be a reminder of the \*principles\* we should all have in our mind when developing AI tech. It can be better than making you check some check boxes every time you submit a paper (after all work is completed and there is no way you will say "ah.. this dataset is probably biased - so I won't submit my paper")
- Suggestions of specific educational materials:

- See CNPEN opinion no. 3 on the ethics of chatbots <https://www.ccne-ethique.fr/fr/actualites/cnpn-agents-conversationnels-enjeux-dethique>

**Positive remarks.** (the majority of responses)

- I see many attacks on Twitter against those who speak out. Please don't feel discouraged, I'm sure many people like me are silent but very grateful. It's a wonderful feeling when we think to ourselves that something is not right and we find a paper or even a tweet that discusses our worries, so we see that there are experienced researchers thinking about and exposing these problems.
- Sincere thanks for what you are doing.
- Thanks for putting this together; super important questions :)

## 5 Conclusion

### Acknowledgements

The AEC would like to thank all of the respondents to the survey. This report and the mandate of the AEC is not possible without all of the time invested by our stakeholders to push for better policies and foresight germinated from within our own community.

## 6 References

### References

- Karën Fort and Alain Couillault. 2016. [Yes, We Care! Results of the Ethics and Natural Language Processing Surveys](#). In *international Language Resources and Evaluation Conference (LREC) 2016*, Proceedings of the international Language Resources and Evaluation Conference (LREC) 2016, Portoroz, Slovenia.
- Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. [Model cards for model reporting](#). In *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT\* '19*, page 220–229, New York, NY, USA. Association for Computing Machinery.