Ethics of NLP and Crowdsourcing

Massimo Zimmerman
University at Buffalo
massimoz@buffalo.edu

Abstract

We look at the potential problems with maintaining ethical approaches and outcomes concerning Natural Language Processing, namely the impacts that funding, designing, and outputting NLPs can have on anyone. If such ethics are respected, what ideas and resolutions can be made commonplace so that NLPs do not suffer from ethical quandaries that have occurred in other disciplines of machine/deep learning applications (e.g. COMPAS).

1 Introduction

As Janet Morris, an American novelist and advocate for non-lethal military technologies and expenditures, once said: ”No ethos, pursued without thought or mercy, is ethical”. This quote, in particular, encapsulates the essence of this literature review, as it establishes a moral basis for how everyone should judge the ethicality of their actions and decisions. While Natural Language Processing is a fascinating advent of AI technology, society needs to ensure that our endless pursuit of technological prowess within this discipline provides benefit to everyone, and these so-called advances do not jeopardize or harm anyone or any group in the process. In this paper, we explore the motivations for maintaining the ethicality of Natural Language Processors, as well as the impacts this technology has on not just the applications of NLP and the industry, but also the researchers, scientists, and workers (specifically crowdworkers).

Before examining the ethics associated with the development and deployment of Natural Language Processors, one must first acknowledge the purpose and motivations behind establishing a system of ethics. Unfortunately, there is no universally-accepted ethos that all people abide by, which can further complicate the dilemma of deciding what ethical notions researchers/developers and the associated NLPs they create are supposedly required to uphold. Thus, it is much simpler to, instead, identify the desirable outcomes set forth by promoting a system of ethics - to ensure NLPs ’do good’ and ’do no harm’ towards any persons or benign entities.

1.1 Dangers of Unethical NLPs

Now, what facets of Natural Language Processing could induce harmful consequences on society in general? Through the summation of all the research conducted on this particular topic, most sources align on the matters of **privacy**, **confidentiality**, and **transparency** being points of concern. While living human subjects are rarely utilized in developing NLPs (thus posing no physical harm to any individuals), there are numerous areas which still draw apprehension from researchers and the general public alike: the data provided through a parsed corpus; the disclosure of personal information relating to crowdworkers, annotators, etc.; and the lack of transparency surrounding these matters, respectively.

Yet, within all man-made systems exists a degree of *bias*, which is oftentimes ignored or not even made aware of, and Natural Language Processing is certainly no exception to this. Systemic bias takes many forms - whether it be **overfitting**, **overexposing**, or **overgeneralizing** any demographic of humanity - and addressing these concerns are not as straightforward as, say, ensuring confidentiality of personal information via mandated practices or well-encrypted software protecting a database. Factor in the impact that NLPs and other AI/Machine Learning technologies have on crowdsourcing, and the ethicality of an entire discipline becomes as chaotic as it can be.

This reality, however, does not stop researchers and other AI enthusiasts from pursuing real solutions to these issues revolving around systemic bias. The following sections will detail how exactly systemic bias, crowdsourcing, and other ethical can be addressed and what the future holds for providing a universally good applications of NLPs and, by extension, other AI technology.

2 NLP and Data

NLPs exist everywhere we look - recommendation systems, machine translation systems, healthcare, chatbots, and general question-answering systems (Prabhumoye et al., 2020) - and it's foolish to assume NLP technology will not become more prevalent in the near future, unlocking innumerable amounts of other applications and systems from which anyone can find benefit. While advancements in NLP will certainly effectuate upon all, it cannot come at the loss or deterioration of human life. Thus, when it comes to entrusting technology to carry out a task, the first question everyone asks is: will myself (and my data) be safe?

2.1 Safeguarding Data

Let's begin with the prospect of data.

The medical/clinical aspect of NLP is, perhaps, the perfect example to demonstrate the sensitivity of acquiring what little data NLP researchers are able to get their hands on, thanks to many legal and institutional statutes that exist to protect the privacy of one's bio-metric and other personal information (Šuster et al., 2017). Of course, hearing that it is quite difficult to acquire access to this data can be seen as a positive in the eyes of those concerned with the privacy and security of the information they share with a healthcare organization/clinic or medical group. However, then arises the issue of transparency - since it is so arduous for researchers to receive permissions to access any sort of data from most corporations, all researchers have afforded to them are small amounts of samples.

2.2 Bias from Withholding Data

With all of the nuisances surrounding obtaining consent and secure access of sensitive data aside, now the issue of **sampling bias** comes into play. With a scarce number of available samples at a researcher's disposal, from only a select region or corporation, the complication of overgeneralizing a population creates even more sampling bias (Šuster et al., 2017) (Kaplan et al., 2014). As aforementioned, the problems relating to bias oftentimes are not easy to remedy, and in the argument of protecting an individual versus ceding some potential bias/inaccuracies in NLP applications, the issue of which route is more 'ethical' becomes almost impossible to answer.

Suster et al. (2017) and Kaplan et al. (2014) are not the only source to recognize this debacle between transparency and privacy.

Leidner and Plachouras (2017) discuss a particular notion coined by other researchers (Thieltges et al., 2016) known as the "devil's triangle", relating to the question of how to strike a balance between transparency and the accuracy/robustness of any chatbot/NLP classifier. To further complicate the matter, the question arose of how to create a chatbot that can ethically represent or interact with a human, i.e. provide comfort/compassion for issues, make the ethically correct choice on the behalf of a human life, etc.

One can imagine now how data (or the lack thereof) can lead to immense difficulties with properly training NLPs to handle the various tasks once conducted by humans, let alone creating an application that is ethically positive. Now, imagine if there were some NLP designs that can promote more ethical practices in the lives of humans. Luckily for us, quite a few NLP implementations already do exist.

2.3 Benefits of Ethical NLPs

In a digital world, data can just about be anything imaginable - scientific, geographical, bio-metrical, financial, even something as simple as words themselves.

Thus, a NLP can successfully understand and even respond/interact with humans, in the hundreds of human-made languages, each spanning millions upon millions of unique words, it is quite impressive!

Yet, as we know, when a human simply reiterates the words of someone else, and then claims it to be their own, that leads to a dilemma known as *plagiarism.*

Universities and other scholarly institutions place a major emphasis on combating and even punishing acts of plagiarism, whether done intentionally or accidentally. The damage that plagiarism can invoke on both the plagiarized words and the plagiarizer's credibility is immense. In response to this, many turn to NLP tools (such as Grammarly) which have been programmed to fact-check and detect plagiarism in any piece of writing it processes. Even normally training NLPs on source material can lead to inconsistencies and potential plagiarism cases being detected that would not have been otherwise.

While plagiarism is just one manner in which NLP techniques can conduct an ethically-positive task for humans at an incredibly low risk, the combined research efforts of various NLP advocates and employers not only allow for more ethical NLPs to be deployed, but also helps demonstrate inaccuracies, limitations, and any potential biases, in current NLP tools and techniques.

3 What Exactly are the 'Ethics of NLP'

Without a doubt, the most controversial and conflicted question this paper will shed light upon is, plainly, what are the so-called ethics that NLP applications and the researchers which curate them are to abide by? Many sources do not agree on a definitive answer - though, has the ethicality of anything ever been unanimously agreed upon?

Rather than trying to rank the strength of 'ethicality' amongst the many researchers and publications on NLPs, we will instead compare and contrast the major arguments on the correct ethical approaches, in the hopes of synergizing a well-formulated approach to how ethical practices of NLPs and their employers can be maintained. Otherwise, a true ethical approach to NLP would only be continuously questioned and reformed as time went on.

3.1 Ethical Concerns

In 2017, the Association for Computational Linguistics hosted their first workshop on Ethics in Natural Language Processing, in an effort to bring together researchers and advocates from all facets of NLP to discuss the concerns related to the impacts of NLPs. In the summary of the *Proceedings of the First ACL Workshop* (2017) , the wide range of topics discussed through the various paper submission focused primarily on, "overgeneralization, ... privacy protection, bias in NLP models,... " and more. While this proceeding was the first of its kind, the recognition of the potential ethical concerns for NLPs (which, sadly, still exist today) showcases the foresight of the collective researcher community to properly identify the issues surrounding this very topic.

Fast forward to 2018, where the *Proceedings of the Second ACL Workshop* (2018) convened to address the exact same topics, on, "overgeneralization, ... privacy protection, bias in NLP models,... " and more. Obviously, while no one is to expect great change to occur in a year's span, the replication of ethical concerns between the First and Second Proceedings is definitely worth noting.

Perhaps the most eye opening concerns of them all are not what the NLP community believes to be a priority matter, but what these researchers believe the public is even made aware of. Fort and Couillault (2016) conducted a series of surveys with both French and 'International' researchers involving the participants' opinions on the state of ethicality surrounding their work and the public's knowledge of said operations. Some of the questions included:

1. "Have you ever refused a project due to ethical issues?"

2. "Do you think the public is aware of the limits and possibilities of the tools we create?"

3. "Do you think the authorities are aware of the limits and possibilities of the tools we create?"

4. "Do you consider yourself responsible for the usages imagined from the applications/algorithms you create?

5. "In your projects, do you consider the licensing and distribution of your language data?"

For the questions posed above in Fort and Couillault (2016), the responses were pretty mixed. The International research community's majority responses were a ”No” (53 percent) for Question 1, a ”No” (91 percent) for Question 2, a ”No” (78.5 percent) for Question 3, a ”Yes” (52.5 percent) for Question 4, and a ”Yes” (84 percent) for Question 5.

While there were virtually zero unethical applications of NLPs exposed and dissected upon, to see the majority of researchers say they have never refused any project due to ethical concerns showcases either the former (which is great), or that researchers are not particularly concerned with the ethicality of their NLP-related work (not so great). As for Question 2 and 3, a resounding number of researchers agreeing that the public and the authorities are not made aware of the potential that NLP applications possess is quite disturbing, for an ethical concern that is not publicly known is an ethical concern that can very well fly under the radar for a very uncomfortably long time. Finally, the responses to Question 4 and 5 detail a sense of responsibility by researchers involved with NLP development, which is very honorable and showcases both the knowledge and sense of responsibility these participants share for their work.

While there may be a universal similarity or majority consensus in ethical concerns, the subsequent ethical propositions facilitated by the former are far from equivalent.

3.2 Ethical Proposition

We return to the quintessential question of this paper: ”what *are* the ethics associated with Natural Language Processing”?

In order to properly brainstorm which ethical propositions should be included in the system of ethics surrounding NLP, we will attempt to synthesize the common and integral standards mentioned by the multitudes of researchers who have examined this topic prior.

Leidner and Plachouras (2017), for example, elaborated on set of seven principles for *privacy by design*, created by Ann Cavoukian in 2009, which they generalized to be their ethic standards in their paper. They read as follows:

1. Proactive not reactive
2. Ethical as the default setting
3. Ethics embedded into the process
4. End-to-end ethics
5. Visibility and transparency
6. Respect for user values

These are very loose standards, albeit provide more than enough details into the mindset of how Leidner and Plachouras (2017) view the state of ethics in NLP. Rather than list ethical concerns or devise exclusive solutions to single issues, these two authors conceived a general ethics system (with inspiration from Ann Cavoukian) by which can apply to all applications or aspects of Natural Language Processing.

Of course, this is one of many ethical guidelines set forth by separate authors. Another paper advocated that there only be **two** standards by which all other ethical guidelines be developed or applied to Natural Language Processing (Prabhumoye et al., 2020).

The first standard essentially states that an ethical guideline should aid in deciding which topics or concerns are worth investigating. In other words, a guideline needs to be future-proof and proactive in nature, or at the very least not be ambiguous so as to leave much interpretation into whether a topic has ethical concerns or not.

The second standard relates to a guideline then providing insight into deciding on the question of *how* to address an ethical concern. A guideline, once again, needs to help answer the problem/concern. A guideline that only leads to more questions is not a suitable guideline, according to Prabhumoye et. al (2020).

3.3 Which Ethical System To Choose

The contrast in theories regarding ethical propositions between select papers and authors is a clear demonstration of how split the researcher community is on what ethical system to govern NLPs, let alone inferring upon the centuries-old disputes that deontologists and adherents of philosophy have on which ethical system is the best to govern all of humanity. Despite their differences, at least quite a few papers and their respective authors agree that some form of ethical standards should supervise the field of Natural Language Processing.

Conversely, some authors such as Hovy and Spruit (2016) questioned the notion whether a discussion on ethics is even relevant for Natural Language Processing. They deduced that, due to the extremely low discourse in the realm of ethics in NLP, along with stating how NLP research (besides the work of annotators) does not require human subjects at all, NLP development ”has not obviated a need for ethical considerations” (Hovy and Spruit, 2016). However, this logic of thought is very dangerous, as all it takes is for a proactive mindset to become reactive one before a slew of injustices, which would’ve never occurred prior, all of a sudden materialize and wreck havoc.

Leidner and Plachouras (2017) said it best when they stated, ”[w]e cannot even aspire to give a survey of centuries of moral philosophy in a few sentences”, to which we are then expected to be able to tie a near infinite pool of knowledge into a simple system of ethics for a discipline as specific as Natural Language Processing. The former would be assuming that there exists a single school of thought with all of the answers/solutions to life, which simply is not the case. Alternatively, we as a society should pool our collective knowledge and concerns together to address both present and future concerns with the ethicality of our NLP applications, utilizing the past as a guideline, not as a solution.

4 Crowdsourcing

For all of the discussion on the ethics of monitoring the impacts of NLP applications, there was hardly any concern about the ethical consequences on people, as it was believed that human subjects were not viable with NLP training. That was, until **crowdsourcing** boomed.

To summarize, the rise of crowdsourcing came about as employers transitioned from contracting expert annotators or linguistics students in favor of crowdworkers, or as some call ”Turkers” from platforms such as Amazon Mechanical Turk, Tencent Questionnaire, Figure Eight, and so on. The reasons for this were all natural: lower costs, convenience, speed, and even scalability (Shmueli et al., 2021).

Supposedly, according to Shmueli et. al (2021), the ”general consensus of the literature” would see the utilization of crowdsourcing as a non-ethical concern, so long as crowdworkers are sufficiently compensated for their work. MTurk and other platforms seemingly concur, as crowdsourcing has only expanded in NLP research and in other machine learning and AI-related disciplines. Of course, the prospect of money does not simply solve every, or many in fact, ethical concerns revolving around treatment of people or the power imbalance between an employer and employee, in this case a researcher and a ’Turker’/crowdworker.

Yet, the issues with the ethicality of crowdsourcing does not lie in these workers being granted more opportunities than more experienced researchers with qualifications or even in the payout that Turkers receive (which is considered to be very ’fair’).

So then, where does the issue lie?

4.1 Ethical By Design(?)

Consider the following: Shmueli et. al (2021) determined, through a review of all 703 crowdsourced, accepted papers through ACL, EMNLP, and NAACL over the last 6 years, that only 14 papers even mentioned a review of their research throguh an IRB - Institutional Review Board. In other words, only 2 percent of all crowdsourced NLP research has even been analyzed or examined for proper ethical conduct/treatment of the crowdworkers and the data they were tasked with annotating, labeling, producing, etc.

The apparent lack of concern there is for providing a thorough evaluation of the supposed work environment/jobs tasked onto Turkers and other crowdsourced worker should already be a dangerous sign of negligence. Yet, some would argue that since NLPs do not directly train on human participants, like how humans would test a drug on mice, that there can be no such ethical concern involving a person. However, this precedent is far from the truth, and the next sections will explain how.

4.2 The Rights of Crowdworkers

Without completely jettisoning into the history of ethical conduct for human subjects in the last century, it is safe to say that research ethics and the notion of protecting human subjects has evolved immensely, taking shape in the form of the *Belmont Report* (for the Protection of Human Subjects of Biomedical and Research, 1978) and other mandates/practices developed by either national efforts in the US or through the rise and achievement of institutional review boards.

Additionally, the *Belmont Report* is centered around 3 basic ethical principles, all of which can be applied to the case of human subjects - **respect for persons**, **beneficence**, and **justice** (for the Protection of Human Subjects of Biomedical and Research, 1978). A later adaptation of this report became the basis for a rule known as the *Final Rule* (Food et al., 2018), which states that a human subject is simply an individual with which an investigator, researcher, etc. conducts research involving:

- the human subject's personal information, and later uses or analyzes it; or
- uses, analyzes, or even generates identifiable private information relating to the human subject

Surely, this would categorize a crowdworker as a human subject and, thus, encourage an IRB to process more crowdsourced research. In fact, *Final Rule* mandates that research involving human subject(s) be subjected to review via an IRB us2018code, .

Final Rule even dictates that a researcher is responsible for upholding the Belmont Report principles even if the human subjects merely interact with the researchers, which would include the usage of MTurk, Amazon Mechanical Turk, and other analogous websites through which researchers can hire a crowdworker (Shmueli et al., 2021).

Despite this unethical lack of protections, there are far worse issues with crowdsourcing that are not commonly addressed.

4.3 Unethical Practices of Crowdsourcing

In today's society, it has grown increasingly apparent that certain words, headlines, images, and so on can invoke harm upon its readers. This notion is no different for annotators, whether they be a researcher or a crowdworker. Trigger warnings have become prevalent in papers, movies, and other mediums to protect viewers from observing a psychologically harmful piece, and crowdworkers should receive the same concern for their health.

As many crowdsourcing websites oftentimes connect the crowdworker with just the job to-be-completed, thus excluding any relationship between the employer and the worker, the crowdworkers are left liable for whatever damage is done to the project or their own health. With no sense of accountability or safeguards provided from crowdsourcing websites, crowdworkers can very easily suffer psychological or reputational damage through annotating or other NLP-related tasks they take up, and have no legal avenues to reverse it. This is a serious matter, as no man or woman should feel as though their work is causing harm to themselves.

Another unethical practice is the potential exposure of sensitive information on said workers through the production or completion of the jobs. For example, workers can very well reveal privy information about themselves when creating subjective labels for an NLP, or even be tracked by websites such as MTurk for the completion time it takes these workers to complete tasks and then displayed for potential researchers/employers (Shmueli et. al, 2021). If the former is considered serious, researchers are also able to obtain personal information through the crowdsourcing websites, which usually allow for filters to hire workers by their age, financial situation, gender, employment status. This, for example, is prevalent on Amazon Mechanical Turk, and it demonstrates yet another ethical lapse of judgement in protecting private data/information of not what is being processed into an NLP, but of the workers themselves.

If the risk of being harmed by a job and having sensitive information drawn without one's consent is not bad enough, there is a very concerning issue with crowdsourcing websites inherently providing work to vulnerable populations, which can then be assigned and involved with NLP research. This is especially apparent with MTurk, where research into the platform concluded with statistics showing a great portion of MTurk's employees residing in developing countries, meaning the increased risk of hiring an economically-disadvantaged persons. Additionally, it is difficult to prove whether crowdworkers on MTurk and other platforms are also of age, do not have impaired decision-making capacity, or are also educationally disadvantaged (Shmueli et. al, 2021). This creates an immense power imbalance scenario where crowdworkers, who perhaps may simply be trying to find any work available to them, can be unknowingly taken advantage of by a researcher from a well-off country/institution, all of which further exasperates the need for ethical intervention into the condition and status of crowdworkers.

Without a doubt, the field of crowdsourcing carries immense ethical concerns and malpractices, of which the crowdworkers and researchers alike may not even be aware of.

5 Going Forward

The premise of ethical change does not begin with a rejection of previous efforts/philosophies, but rather a promise to uphold a standard of morally-good actions, no matter how much the ethicality of things evolve, for everyone.

For the crowdworkers involved with lending whatever time or skills they have to completing tasks crucial to NLP research and development, a mandate to require either Institutional Review Boards or the researchers themselves to review any and all risks associated with the crowdsourced tasks involved with their researcher would significantly reduce the said risks as well as increase awareness of crowdsourcing and the ethicality surrounding it (Shmueli et. al, 2021).

As for the general ethics of Natural Language Processing, we will continue to observe ethical development and applications of NLPs as long as we remain vigilante in adhering to ethical standards such as the Belmont Report's principles (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978), which are echoed in many other papers surrounding ethics. Leidner and Plachouras (2017); Fort and Couillault (2015); Thieltges, Schmidt and Hegelich, (2016) - all of these writers, as well as the Proceedings of the First and Second ACL Workshops (2017, 2018), sought to uphold these principles by examining where cases of inhumane treatment, systemic bias, and breaches in the trust and security of private information exist. Essentially, to promote the best ethics for NLP is to make ourselves aware of any potential unethical practices and remedy them.

With a collective and committed effort to root out injustice, the ethicality of NLP will remain both pure and resolute as researchers and employers seek to tap into this brilliant technology.

6 Conclusion

Natural Language Processing, and the hard work of the NLP community as well as crowdworkers, have facilitated a bright future for linguistic applications to aid and improve the livelihood of everyone. Whether it be email filters, translation software, smart assistance, recommedation systems, chatbots and text summarization for a plethora of industries including medical/clinical, and so on, the value that NLP technology carries to an industry or to a better quality of life is almost immeasurable.

Yet as with all great innovations, the good they bring must not be to the benefit of a few and to the harm of many. Ethical analysis of NLP technology, even when there was no unethical cases to be examined, still determined that systemic bias and the risk of divulging sensitive and personal information was large. Additionally, the rise of crowdsourcing - a ever-so-growing service due to the lower costs, convenience, and speed provided by crowdworkers (Shmueli et. al, 2021) - has been perhaps the most unethical practice associated with NLP and other AI-related technology to date.

Despite this, we as a community of researchers, scholars, and ethically-abiding citizens, can demand justice through our own awareness of unethical practices and promote change through our collective ideas and institutions. We can ensure that the world remains proactive in the battle to develop and create ethically-positive NLPs, while also keeping society safeguarded from any potential malpractices involved in the process.

References

Mark Alfano, Dirk Hovy, Margaret Mitchell, and Michael Strube. 2018. Proceedings of the second acl workshop on ethics in natural language processing. In *Proceedings of the Second ACL Workshop on Ethics in Natural Language Processing*.

US Food, Drug Administration, et al. 2018. Code of federal regulations (cfr). *Title*, 21:21.

United States. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1978. *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*, volume 2. The Commission.

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*.

Dirk Hovy and Shannon L Spruit. 2016. The social impact of natural language processing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 591–598.

Dirk Hovy, Shannon L Spruit, Margaret Mitchell, Emily M Bender, Michael Strube, and Hanna Wallach. 2017. Proceedings of the first acl workshop on ethics in natural language processing. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*.

Robert M Kaplan, David A Chambers, and Russell E Glasgow. 2014. Big data and large sample size: a cautionary note on the potential for bias. *Clinical and translational science*, 7(4):342–346.

Jochen L Leidner and Vassilis Plachouras. 2017. Ethical by design: Ethics best practices for natural language processing. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 30–40.

Shrimai Prabhumoye, Brendon Boldt, Ruslan Salakhutdinov, and Alan W Black. 2020. Case study: Deontological ethics in nlp. *arXiv preprint arXiv:2010.04658*.

Boaz Shmueli, Jan Fell, Soumya Ray, and Lun-Wei Ku. 2021. Beyond fair pay: Ethical implications of nlp crowdsourcing. *arXiv preprint arXiv:2104.10097*.

Simon Šuster, Stéphan Tulkens, and Walter Daelemans. 2017. A short review of ethical challenges in clinical natural language processing. *arXiv preprint arXiv:1703.10090*.

Andree Thieltges, Florian Schmidt, and Simon Hegelich. 2016. The devil's triangle: Ethical considerations on developing bot detection methods. In *2016 AAAI Spring Symposium Series*.