Cartography of Natural Language Processing for Social Good: Definitions, Statistics and White Spots

Paula Fortuna[1], Laura Pérez-Mayos[1], Ahmed AbuRa'ed[1],
Juan Soler-Company[1], Leo Wanner[2,1]
[1]NLP Group, Pompeu Fabra University,
[2]Catalan Institute for Research and Advanced Studies (ICREA)
paula.fortuna|laura.perezm|ahmed.aburaed|juan.soler|leo.wanner@upf.edu

Abstract

The range of works that can be considered as developing NLP for social good (NLP4SG) is enormous. While many of them target the identification of hate speech or fake news, there are others that address, e.g., text simplification to alleviate consequences of dyslexia, or coaching strategies to fight depression. However, so far, there is no clear picture of what areas are targeted by NLP4SG, who are the actors, which are the main scenarios and what are the topics that have been left aside. In order to obtain a clearer view in this respect, we first propose a working definition of NLP4SG and identify some primary aspects that are crucial for NLP4SG, including, e.g., tackled areas, ethics, privacy and bias. Then, we draw upon a corpus of around 50,000 articles downloaded from the ACL Anthology. Based on a list of keywords retrieved from the literature and revised in view of the task, we retrieve from this corpus articles that can be considered to be on NLP4SG according to our definition and analyze them. The result of the analysis is a map of the current NLP4SG research and insights concerning the white spots on this map.

1 Introduction

Measuring the social impact of NLP is not a trivial task. A priori, the range of works that can be considered as developing NLP for social good (NLP4SG) is enormous. It goes from more theoretical works (Cowls et al., 2021), language resources (Midrigan Ciochina et al., 2020; El-Haj et al., 2015) and models (Devlin et al., 2019) to concrete technologies of which many target the identification of hate speech (Fortuna et al., 2021) or fake news (Shu et al., 2017). But there are also others that address, e.g., text simplification or paraphrasing, which can be used to alleviate consequences of dyslexia (Rello et al., 2015), conversational agents for mental health treatment (Gaffney et al., 2019),

or eLearning applications, which support students with specific learning disabilities (Bjekić et al., 2014).

In general, many NLP technologies can be used for good but also for bad; at a larger scale, they may affect the lives of many people, and it is difficult to predict in the first place all the potential positive or negative sides resulting from the application of these technologies. In order to discard at this stage uncontrolled "collateral" positive or negative technology influence, we can assume that social good does not come as a side effect when researching certain fields and developing technologies. Even more: if we do not address directly, measure and intentionally promote and control social good, we can cause more harm than good. Therefore, it is of paramount importance to define what we mean when we say "NLP for Social Good", what aspects of peoples' lives are improved by NLP4SG and how, and what suitable strategies are to promote and measure the impact of technological solutions related to NLP. However, so far, there is no clear picture of what areas are targeted by NLP4SG, who are the actors, which are the main scenarios and what are the topics that have been left aside. In this paper, we discuss what NLP for social good (NLP4SG) is, and how we can promote the development of more socially positive technologies. The contribution of this paper is twofold:

(i) we offer a working definition of NLP4SG and related concepts that can serve as a first orientation in the field;

(ii) we provide an analysis of the current state and the tendencies of the research on NLP4SG.

The remainder of this paper is structured as follows. Section 2 defines NLP4SG and introduces some other central aspects of it – the applications, collaboration, and ethics. Section 3 details the

data, methodology, and results of our evaluation of the social impact in the NLP field. Section 4 elaborates on how to improve the current state of affairs. Section 5 addresses the limitations and ethical concerns, and Section 6, finally, summarizes the implications of our work and draws some conclusions.

2 Defining NLP4SG

Before we set out to provide an overview of the NLP4SG research and explore its characteristics, we need to define what we mean by NLP4SG. Let us start by analysing what is "social good". In the context of social science, Barak (2020) proposes a conceptual "social good" model according to which there are three elements needed to promote social good: innovative technologies, social good domains, and engaging unconventional systems of change, which in this work we also refer to as "collaborations". In the following subsections, we focus on each of these dimensions and dig into other NLP4SG related aspects.

2.1 Social good and NLP technologies

In order to address how NLP technologies can contribute to social good, we draw upon existing research in the broader area of Artificial Intelligence (AI), which intersects NLP problems and methodologies. AI for social good (AI4SG) has recently gained attraction. Floridi et al. (2020) define "social good" in the context of AI. We apply this definition to NLP by replacing 'AI' by 'NLP' and consider NLP4SG as:

> "Design, development, and deployment of **NLP** systems in ways that (i) prevent, mitigate or resolve problems adversely affecting human life and/or the well being of the natural world, and/or (ii) enable socially preferable and/or environmentally sustainable developments."

In what follows, we review the domains and the contexts in which NLP4SG is carried out.

2.2 Applications for Social Good

In research and ethics, the definition of social good focused so far on its use in application areas that generally have a direct positive impact on the society. Several lists of such areas have been worked with. For instance, Shi et al. (2020) highlights agriculture, education, environment sustainability,

healthcare, combating information manipulation, social care and urban planning, public safety and transportation; Floridi et al. (2020) focuses on healthcare, education, equality, climate change, and environmental protection; and Hager et al. (2019) deals with justice, economic development, workforce development, public safety, policing, education, public health, transportation, and public welfare. In the analysis presented in this paper, we draw upon Shi et al. (2020) to compose an adapted list of NLP4SG areas, keeping agriculture, education, environmental sustainability, healthcare, public safety and transportation. We exclude "social care and urban planning", as they may refer to different aspects, and we rephrase "combating information manipulation" as "media corrupted communication" because we want to include not only fake news, but also abusive language. Finally, to tackle specific NLP health-related issues, we extend the list by "language disorders". Consider the first column of Table 1 for the list of areas that we take into account.

2.3 Collaborations for social good

Tomašev et al. (2020) details how AI4SG projects should be approached as a collaborative effort in bringing communities together in order to carefully assess the complexities of designing AI systems. Community involvement assures integration and inclusiveness, and it brings more information to the decision on the design of a technology, including knowledge about the contexts in which design decisions are going to have an impact. Furthermore, community involvement adds other perspectives to the design since researchers alone cannot anticipate all the needs of the users and all the possible usages of a technology. Along the same lines, we propose that NLP4SG needs the collaboration of users, activists, minorities, grassroots movements, businesses, non-governmental organizations (NGOs), and social entrepreneurs to achieve a social positive technological development.

2.4 NLP and Ethics

To achieve a positive impact, technological solutions need to adhere to ethical principles, e.g., guidelines provided by the European Commission, the Organisation for Economic Co-operation and Development, or the Montreal Declaration for Responsible AI (Tomašev et al., 2020). Naturally, this also applies to NLP. Technology based on human data can be potentially harmful, and the presence

of ethics in NLP is therefore much needed. There are three primary topics that frequently underlie ethical issues in NLP research: privacy, bias and dual use (Bender et al., 2020).

Privacy tackles how to protect the privacy of data authors used in the training or evaluation of NLP systems. It has been more widely discussed, e.g., in (Hovy and Spruit, 2016).

Dual use anticipates how a developed technology could be repurposed for negative applications and thus helps design systems such that they do not cause harm; cf., e.g., (Bender et al., 2020).

Bias is about understanding how over- and under-sampling of different populations will affect datasets and models that are built using these datasets. Potential solutions include building less biased datasets, debiasing trained models and matching appropriate training data to a given use case (Bender et al., 2020).

3 Evaluation of social good in NLP

Evaluating the current state of NLP for social good is a crucial step towards the identification of the gaps and promotion of a more impactful technology development. For this purpose, we build upon the *NLP Scholar Dataset* (Mohammad, 2020) and analyse existent features together with new classifications on social good aspects. In what follows, we describe in detail the data and the procedure of our analysis. We make the code available to the community[1].

3.1 Data

The *NLP Scholar Dataset* provides access to more than 50k instances from both ACL Anthology (AA) and Google Scholar (GS), and includes authors' names, year of publication, venue of publication, etc. We use the version of this dataset from June 2020 (Mohammad, 2020). The dataset includes some entries that are not really papers (e.g., forewords, prefaces, programs, schedules, indexes, invited talks, appendices, etc.). After discarding them, we are left with 52,288 papers. Regarding the available paper descriptors, we use: Title, Year, Authors, NS paper type, NS paper venue and GS citations. This data is enriched with some other fields introduced in the next subsection.

3.2 Methodology

We enrich the available dataset with the abstracts of the papers and automatically annotate the NLP4SG-related variables. To validate our automatic annotation procedure, we extract 200 papers as validation set, gathering one opinion per paper with respect to the quality of the annotation.

Retrieving paper abstracts. For each instance (paper) of the dataset, we collect the pdf file of the paper, and extract its abstract using Grobid[2]. Then, we use Microsoft Academic Graph API[3] to complete the missing abstracts. In total, we have been able to retrieve the abstracts for 95.8% of the papers in our dataset.

Annotation as *explicit NLP4SG* For each considered NLP-application area, we compile a list of keywords. This allows us to match NLP publications with the obtained "keyword lexicon" and assess the positive impact in the field.

To come up with the keyword lexicon, we use a set of keywords from (Shi et al., 2020),[4] enriching it further with keywords extracted from the Wikipedia page for language disorders,[5] and with words extracted from the UN Sustainable Development Goals[6]. To filter the final keyword lists, two annotators, instructed with the definitions of NLP4SG from Section 2, reviewed the titles and abstracts of the papers retrieved by each keyword, discarding those with a high percentage of false positives. For instance, the "genetic" keyword is present in the *health* set of the original list from Shi et al. (2020). As this keyword retrieves a high percentage of papers referring only to genetic algorithms we opted to remove it.

The final keyword list is divided into two sets: *areas for social good* and *other dimensions of social good*; cf. Tables 1 and 2. *Areas for social good* keywords correspond to social good applications. As previously outlined in Section 2.2, the main areas are *Agriculture*, *Education*, *Environmental sustainability*, *Healthcare*, *Public safety*, *Social care*, *Transportation* and *Urban planning*.

---

[1]https://github.com/paulafortuna/NLP4SG_NLP4PI_paper

[2]https://github.com/kermitt2/grobid
[3]https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/
[4]https://github.com/csinguva/NLP4SocialGood/blob/master/keywords.py
[5]https://en.wikipedia.org/wiki/List_of_language_disorders
[6]https://sdgs.un.org/goals

Area	Example keywords
Agriculture	sustainable agriculture, farmer, vegetation, livestock
Education	tutor, pedagogy, tuition, textbook
Environmental sustainability	sustainability, wildlife, pollution, biodiversity, climate action
Healthcare	cancer, covid, autism, impairment
Media corrupted communication	fake news, polarization, politic, toxicity
Public safety	crime, police, safety, fraud, terrorism
Social care	gender gap, racism, migrants, social justice
Transportation	carpool, passenger, railroad, traffic
Urban planning	emergency, cost of living, low-income, sustainable cities
Language disorders	dyslexia, coprolalia, echolalia, glossolalia

Table 1: Social good areas' example keywords.

Other Dimensions	Example keywords
Ethics	ethical, bias, privacy, data statement
Social good	interpretability, accountability, social good, social impact
Systems of change and collaboration	NGO, activist, inclusive, financial cost

Table 2: Social good general example keywords.

To these main areas we add two areas of particular relevance to the NLP field, namely *Language disorders* and *Media corrupted communication*. To account for areas that are not explicitly related to applied research, we provide an alternative taxonomy that covers *Other dimensions of social good*: *Ethics*, *General social good* and *Systems of change and collaboration*. For the *other dimensions of social good* we add keywords in accordance with the definitions provided in Section 2

We automatically annotate the set of papers as *explicit NLP4SG* vs. *non-explicit NLP4SG* by using keyword matching. The term 'explicit' intends to highlight here that keyword matching is robust enough to capture only those papers that explicitly mention any of the NLP4SG keywords that we are looking for, and, therefore, it is possible that we misses papers that tackle NLP4SG in a more subtle manner. Papers of the dataset that are not tagged as 'explicit NLP4SG', i.e., that do not match any of the keywords, are tagged as 'non-explicit NLP4SG'.

The outcome of the automatic annotation task has been manually validated by a meta-annotator, who approved the assignment of the *explicit NLP4SG* tags in 95% of the times.

3.3 Results and Discussion

It has been stated that the number of publications in NLP has been increasing over the last years (Mohammad, 2020). Our results confirm that this is also the case for *explicit NLP4SG* works (cf. Figure 1). Our results indicate that until 2010, the percentage of *explicit NLP4SG* papers per year was more constant (around 5%). The majority of the papers until 2010 is related to social good mostly because the research focused on some specific areas. More recently, this trend has been changing. During the last 10 years, not only is the percentage of *explicit NLP4SG* increasing, but the percentage of papers mentioning *other dimensions of social good* has been increasing as well; cf. Figure 1. The year with most *explicit NLP4SG* publications has been so far 2020, where more than 20% of the publications already mention social good-related terms or areas. This figure also shows that the percentage of NLP4SG publications referring to our NLP4SG *areas* is higher than the percentage of publications referring to *other dimensions* of NLP4SG, and only a minority of publications refers to both sets of terms at same time.

Figures 2 and 3 show the different *areas* and *other dimensions* of NLP4SG in terms of verified frequencies. *Healthcare* is the preferred area of investigation, followed by *social care*, *media corrupted communication* and *education*. *Public safety*, *transportation*, *urban planning*, *environmental sustainability* and *language disorders* are areas with less publications. Regarding other social good dimensions, we can state that the research has been focusing mostly on *ethical* issues, directly mentioning *general social good* and related concepts, but rarely referring to *systems for change and collaboration*.

The observed tendency over time and the corresponding detailed analysis show that NLP research is increasingly conscious about its implications for the society and begins to directly address these implications. Still, some particular aspects such as, e.g., collaboration with actors outside NLP, remain to be addressed. In addition, despite having increased considerably over the last years, the percentage of NLP4SG-related research can further be improved.

In order to buttress this claim, we compiled some telling numbers that contrast *explicit NLP4SG* with *non-explicit NLP4SG*; cf. Table 3. These numbers point to the lack of prominence of social good in
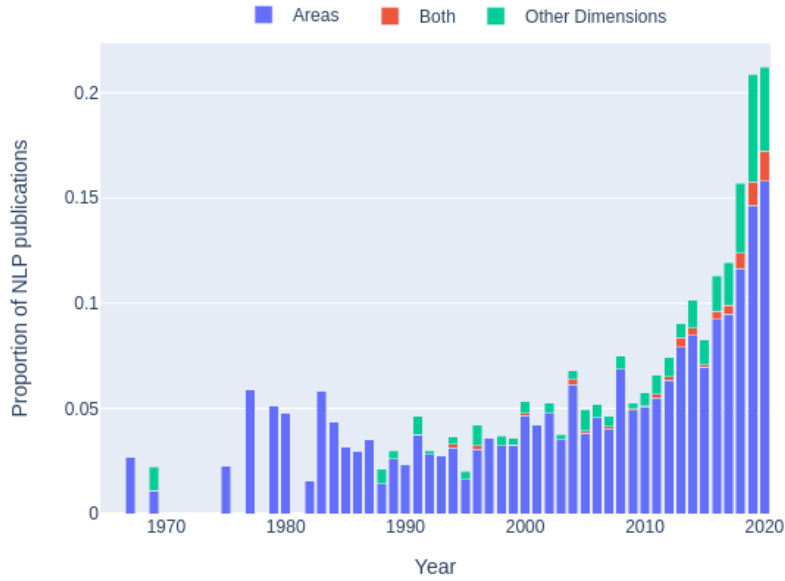
Figure 1: Social good publications per year in proportion to the total of NLP papers. Each bar represents the accumulation of papers that matched general keywords (green), areas (blue) or both (red).
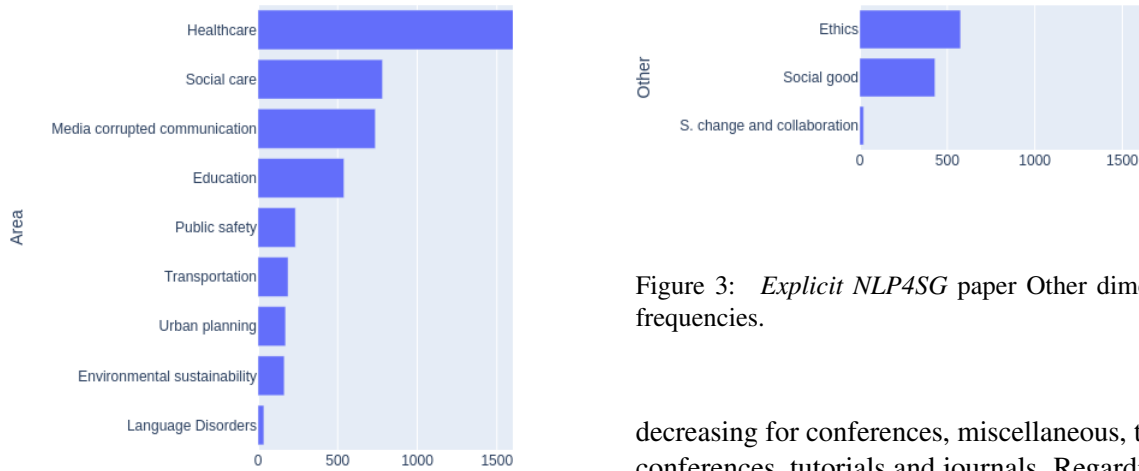

Figure 2: *Explicit NLP4SG* paper areas frequencies.

Figure 3: *Explicit NLP4SG* paper Other dimensions frequencies.

decreasing for conferences, miscellaneous, top-tier conferences, tutorials and journals. Regarding the particular venues, the non-SemEval shared task, RANLP, Workshops, student Research, and Demo lead the top five of venues with the highest percentage of *explicit NLP4SG* papers; cf. Figure 5.

4 Improving the current state of affairs

As shown in the previous section, the NLP field is recently more attentive to social good related issues. Nevertheless, we do believe that there are certain aspects that need further attention by the community. In what follows, we enumerate these aspects, along with some hints on how to address them.

the field. Our results show that *explicit NLP4SG* papers, accounting for 9.63% of the total, tend to have, in average, more authors per paper and less citations. Moreover, 24.02% of the authors have published at least one paper belonging to *explicit NLP4SG*. Shared tasks, workshops and system demonstration are the venues publishing *explicit NLP4SG*; cf. Figure 4. The percentages keep

23

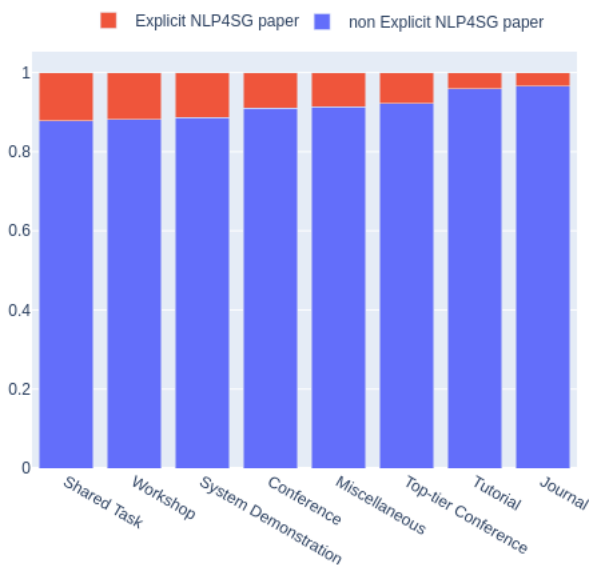| Metrics	Explicit NLP4SG	Non-explicit NLP4SG
Total number of papers in percentage	9.63%	90.37%
Average number of Google Scholar citations per paper	25.87	42.03
From the total authors publishing in NLP	24.02%	75.98%
Mean number of authors per paper	3.65	2.97

Table 3: *Explicit* vs. *Non-explicit NLP4SG* papers statistics.

Figure 4: Percentage of *explicit NLP4SG* publications per type of venue.

Social good areas with less research. Areas that we identified as producing less NLP for social good publications are, e.g., *language disorders*, *environmental sustainability*, *urban planning*, *transportation*, and *public safety*. While it is natural that areas that are less related to language receive less attention in NLP, e.g., transportation, they still offer room for many NLP-applications, which can be tackled with a positive impact.

The discussion on which percentage of the research in NLP is appropriate for the different areas still remains open – if it can be resolved at all.

More than social good areas. Although we follow previous research in an attempt to measure NLP4SG by matching keywords of certain areas (Shi et al., 2020), we must be cautious when look-

ing at the obtained results: while research in a certain area may imply social good, it may also imply social harm, depending on how a certain technology is going to be used (e.g., a fake news detector may be used to detect, but also to generate fake news). Another analysis over the same data may aid to interpret the achieved results and help to understand whether the approaches of the previous work to address social good areas lead to positive or negative outcomes. With this in mind, and for the sake of a broader analysis of research impact, we include into our consideration other NLP4SG dimensions such as ethics, social good terms and systems of change and collaboration. As a guideline for future research, we may conclude that providing data statements and terms of use for the developed technologies would help preventing potential misuses.

Other social good dimensions. When we look at the *explicit NLP4SG*, papers from the considered *areas* are more frequent than papers related to *other dimensions*, and it is only in recent years that *other dimensions*-related papers are increasing in number and have more weight. We believe that research in NLP would benefit from a wider discussion on social good dimensions such as ethics, positive impact and collaborations. In particular, questions such as how the development of NLP applications may involve end-users and include knowledge about their context of use require more attention.

Social good should not be the researcher's enemy. We show that *explicit NLP4SG* publications tend to have less citations in average and are published in smaller venues. The reduced number of authors of explicit NLP4SG papers suggests that there is a smaller NLP4SG community within the larger NLP community. We believe that it is urgent to actively encourage research to tackle social good areas, and, in particular, also to promote the formation of an interconnected social good community across different academic disciplines. Pushing towards this objective will benefit both the field and the society that is impacted by the technology that we produce.

5 Limitations and Ethical Concerns

As mentioned in Section 3.2, our keyword matching approach to the identification of NLP publications as being relevant to NLP4SG is robust and
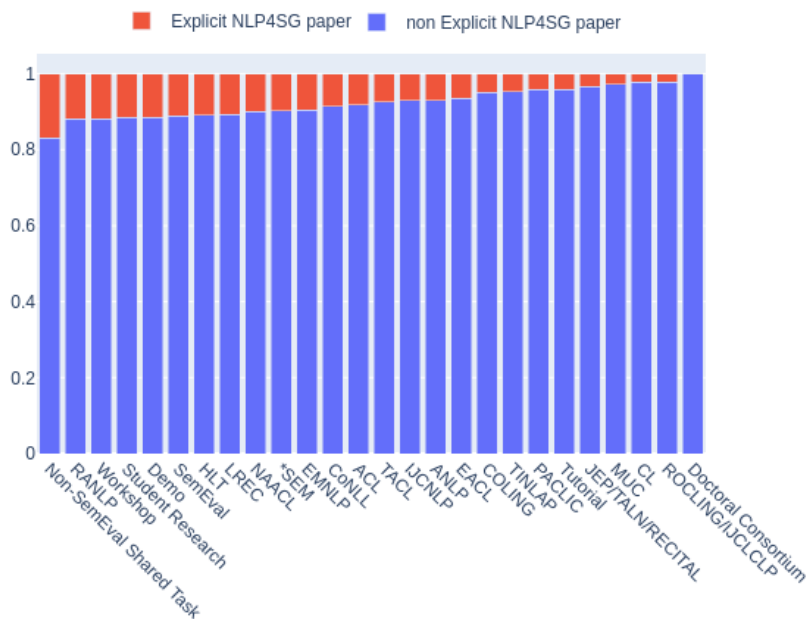
Figure 5: Percentage of *explicit NLP4SG* publications per venue.

performative enough only for a subset of the publications, namely those that contain one of the keywords that we are looking for. The term *explicit NLP4SG* intends to highlight this limitation. It is possible (or even likely) that it misses some papers tackling NLP4SG in a more subtle manner. This means that the results presented in this paper serve as a lower bound baseline.

As far as the collection of the used keyword is concerned, we started by using an initial sample of terms specifically conceived for Artificial Intelligence. We tried then to add some NLP related expressions and remove terms that were bringing misleading results. However, in the course of the presented analysis it became clear that a more systematic method could have revealed more social good NLP related terms. Furthermore, when discussing definitions of social good, we should bear in mind that what is considered to be a "positive impact" depends on the context and set of values. For instance, ethical concerns and guidelines are different according to different countries (Hovy and Spruit, 2016; Berberich et al., 2020) and are not absent of social and political interests (Washington and Kuo, 2020). As a consequence, we must acknowledge the limitation of our analysis in this regard since we follow an Eurocentric perspective and focus only on ACL publications.

6 Conclusions

The goals of this paper have been to help to draw a clearer picture of what NLP4SG is and where we stand in the current state of NLP. We established working definitions of NLP4SG and identified some aspects that are crucial for the analysis of NLP publications with respect to their relevance to NLP4SG, namely *technologies*, *areas*, *collaborations*. NLP-specific *ethical* aspects formed another perspective of our analysis. We drew upon the ACL Anthology corpus and annotated papers in terms of *explicit* vs. *non-explicit NLP4SG* to show a clearer view of the evolution of the field. We identified social good-relevant NLP areas with less research, as well as other social good dimensions that are important to address, and proposed a non-exhaustive list of aspects that need further attention by the community.

The results of the research in NLP have a huge impact on the whole society, and we strongly believe that it is urgent for the community to potentiate and encourage research that not only includes ethical consideration, but also actively addresses social good.

Acknowledgments

25

References

Michàlle E. Mor Barak. 2020. The practice and science of social good: Emerging paths to positive social impact. *Research on Social Work Practice*, 30(2):139–150.

Emily M. Bender, Dirk Hovy, and Alexandra Schofield. 2020. Integrating ethics into the NLP curriculum. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts, ACL 2020, Online, July 5, 2020*, pages 6–9. Association for Computational Linguistics.

Nicolas Berberich, Toyoaki Nishida, and Shoko Suzuki. 2020. Harmonizing artificial intelligence for social good. *Philosophy & Technology*, 33(4):613–638.

Dragana Bjekić, Svetlana Obradović, Milica Vučetić, and Milevica Bojović. 2014. E-teacher in inclusive e-education for students with specific learning disabilities. *Procedia – Social and Behavioral Sciences*, 128:128–133.

Josh Cowls, Andreas Tsamados, Mariarosaria Taddeo, and Luciano Floridi. 2021. A definition, benchmark and database of ai for social good initiatives. *Nature Machine Intelligence*, 3(2):111–115.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Mahmoud El-Haj, Udo Kruschwitz, and Chris Fox. 2015. Creating language resources for under-resourced languages: methodologies, and experiments with arabic. *Lang. Resour. Evaluation*, 49(3):549–580.

Luciano Floridi, Josh Cowls, Thomas C. King, and Mariarosaria Taddeo. 2020. How to design AI for social good: Seven essential factors. *Sci. Eng. Ethics*, 26(3):1771–1796.

Paula Fortuna, Juan Soler Company, and Leo Wanner. 2021. How well do hate speech, toxicity, abusive and offensive language classification models generalize across datasets? *Inf. Process. Manag.*, 58(3):102524.

Hannah Gaffney, Warren Mansell, and Sara Tai. 2019. Conversational agents in the treatment of mental health problems: Mixed-method systematic review. *JMIR Mental Health*, 6(10):e14166.

Gregory D. Hager, Ann W. Drobnis, Fei Fang, Rayid Ghani, Amy Greenwald, Terah Lyons, David C. Parkes, Jason Schultz, Suchi Saria, Stephen F. Smith, and Milind Tambe. 2019. Artificial intelligence for social good. *CoRR*, abs/1901.05406.

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, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers*. The Association for Computer Linguistics.

Ludmila Midrigan Ciochina, Victoria Boyd, Lucila Sanchez-Ortega, Diana Malancea_Malac, Doina Midrigan, and David P. Corina. 2020. Resources in underrepresented languages: Building a representative Romanian corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3291–3296, Marseille, France. European Language Resources Association.

Saif M. Mohammad. 2020. NLP scholar: A dataset for examining the state of NLP research. In *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, pages 868–877. European Language Resources Association.

Luz Rello, Miguel Ballesteros, and Jeffrey P. Bigham. 2015. A spellchecker for dyslexia. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, ASSETS 2015, Lisbon, Portugal, October 26-28, 2015*, pages 39–47. ACM.

Zheyuan Ryan Shi, Claire Wang, and Fei Fang. 2020. Artificial intelligence for social good: A survey. *CoRR*, abs/2001.01818.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *SIGKDD Explor.*, 19(1):22–36.

Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle CM Belgrave, Daphne Ezer, Fanny Cachat van der Haert, Frank Mugisha, et al. 2020. Ai for social good: unlocking the opportunity for positive impact. *Nature Communications*, 11(1):1–6.

Anne L. Washington and Rachel S. Kuo. 2020. Whose side are ethics codes on?: power, responsibility and the social good. In *FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27-30, 2020*, pages 230–240. ACM.