Introducing NLP for Social Good (NLP4SG)

Co-authors: Zhijing Jin (AI PhD at Max Planck Institute & ETH), Luise Wöhlke (CS undergrad at University of Edinburgh), Dan Lahav (AI researcher at IBM Research), Peter Jacobs (MA student at Uni Jena)

We are excited to introduce a new initiative to spread effective altruism ideas in an important subfield of AI, the field of *natural language processing* (NLP). Research in this field focuses on language-related AI technologies. Our initiative, called NLP for Social Good (NLP4SG), aims to help leverage these technologies towards real-world social needs.

nlp4sg.vercel.app

Executive summary:

- The NLP for Social Good (NLP4SG) initiative aims to realize better resource-allocation in NLP to optimally leverage powerful NLP technologies for impact. To achieve this, we believe we have to bring real-world pressing problems and NLP research closer to each other. EA functions as a guide for our efforts.
- The initiative started in early 2021, and involves 150+ interested NLP/AI researchers around the world.
- In Aug 2021, we held the 1st workshop on NLP for Positive Impact at the global top NLP conference, the Annual Meeting of the Association for Computational Linguistics (ACL 2021). Our next workshop will take place in Nov 2022.
- The approach our initiative takes is twofold: (1) **Helping researchers** have an impact by promoting EA-aligned research agendas and providing resources on having an impact in NLP. (2) **Helping philanthropic/charitable organizations** by connecting them with NLP technologies & researcher networks that could benefit their cause.
- The main impact of NLP4SG research comes from efficiency improvements for work on important causes and generating insights on large textual datasets (for research).

- Our work deals with NLP tools pertaining to "narrow AI"[1], and thus is not hugely relevant to "strong AI" safety. However, we openly address any lingering safety and other concerns in the "Potential Concerns" section.
- We can use your help! We are looking for: (1) EAs with experience in cause prioritization to compile a list of top NLP4SG causes, (2) EAs with expertise in a cause area to introduce NLP researchers to real and pressing problems & together identify ways NLP can help, (3) pointers to datasets, and (4) volunteers for the initiative.

Acknowledgements: We thank Sella Nevo, Omer Nevo and Edo Arad for their review of the post and constructive feedback. Any mistakes are our own.

Motivation

Recent years have seen rapid advances in applied NLP technology. Today, services used by billions of users such as Google Search, Google Translate, Alexa, news recommendations, Facebook, and Twitter are powered by NLP. Considering these commercial successes, we hope for a symmetric surge in NLP applications for positive impact. For example, NLP has been used to build an information management tool for medical professionals, which helps in the fight against Covid. However, we don't see nearly enough of such impactful uses of NLP, leaving ample low-hanging fruit unplucked. We want to rectify this asymmetry, and allocate appropriate resources to the study of impactful NLP technologies.

The Initiative in Brief

In early 2021, we started to work on NLP for Social Good, and launched the NLP for Social Good initiative. We are a group of EAs and non-EAs in NLP led by Zhijing Jin (AI PhD at Max Planck Institute & ETH) and Dan Lahav (AI researcher at IBM Research) (see "Who We Are").

The **vision** of the NLP for Social Good initiative is a world where the power of transformative technologies is leveraged for the greatest social good.

The **mission** of the NLP for Social Good initiative is to optimize resource-allocation towards ways NLP can help solve the world's most pressing problems, identified using reason and evidence.

---

[1] Narrow AI, which encompasses AI tools that do one task really well, stands in opposition to strong AI, also known as artificial general intelligence (or AGI). From here on, we'll use the terms "*narrow NLP*," referring to language technologies that solve some specific task, and "*strong NLP*," which aims to develop language agents with general, human-level intelligence. NLP4SG is concerned more with narrow NLP than strong NLP.

The **approach** we take to achieve our mission is twofold:
(1) **Helping researchers** have an impact by promoting EA-aligned research agendas and providing resources on having an impact in NLP.
(2) **Helping philanthropic/charitable organizations** by connecting them with NLP technologies & researcher networks that could benefit their cause.

NLP4SG has been received well by the NLP community, and we've had talks at the most important international NLP conference, ACL 2021, including the keynotes by Prof Rada Mihalcea, Prof Chris Pott, Prof Alex Cristia, and a dedicated panel discussion.

How Exactly Can NLP Tools Help Us Do More Good?

Very broadly, you can say that NLP4SG provides NLP tools based on existing technologies to increase the efficiency and quality of work on important causes, as well as providing insights that are not achievable without such tools. The word "narrow" here means that the tools are specific to one task, i.e. they aren't *generally intelligent*. This means that broad concerns about "strong AI" safety are somewhat less relevant to NLP4SG, but we'll discuss this at more length in the "Potential Concerns" section.

An example of a narrow NLP tool that improves efficiency would be a tool that produces text summaries for a knowledge worker to save them time. This is effective if the knowledge worker is working on a pressing cause, e.g. a public EA communicator needing to read lots of material on, say, biorisk.

A simple heuristic you can apply to figure out whether NLP4SG can help with a problem is "*Does it involve text?*". Example NLP tasks include summarization, transcription, translation, topic classification, search, etc. These tasks span a variety of mediums involving written text (reports, emails, reviewing existing literature, reviewing applications, processing textual data like surveys, etc.) or speech (meetings, presentations, talks, etc.).

Thus, the opportunities NLP4SG holds for office and knowledge work, including research, are ample. Many problems can be automated or augmented to save resources and improve the quality of results. An example of an NLP tool for researchers is the Elicit research assistant, which some EAs use already. Another example of an NLP tool that does good would be an automated helpdesk service via text messaging for pregnant women in South Africa, with the hope of reducing maternal mortality. (We haven't looked into how cost-effective this particular example is, so please view this more as a proof of concept than an ideal EA intervention.)

To break NLP4SG down further, we can say there are roughly three types of efficiency-boosting tools. Ones that
(1) convert between different "forms" of language: Text-to-speech, speech-to-text, and language translation.
(2) generate text for some specification (e.g. question-answering tools).
(3) extract information from text. Optionally do this *many* times and try finding insights in the resulting dataset of many pieces of information (e.g. extract topics from tweets, then find topics that are trending).

Summarization would be an example of combining (3) (extract information from text) and (2) (generate textual summary from information).

This is all under the heading of efficiency improvements, but (3) has some hidden powers. Often, it makes tasks so much more efficient, that it practically enables entire projects that were infeasible before. Who has time to analyze 1 trillion tweets for example? Well, now this can be done in the scope of a single NLP paper. The results can then be used to quantitatively answer previously intractable questions—in the case of this paper, *How does public opinion influence Covid policy?*. More real examples of this general approach are:

- Extract medical symptoms from clinical notes > Answer *Who responds well to this heart disease treatment?*[2]
- Extract topics from patents > Answer *What drives innovation?*[3]

So NLP4SG can make tasks involving text more efficient, e.g. for knowledge workers and important causes, and even so efficient that we can find insights in huge amounts of textual data.

In **practice**, NLP4SG generally proceeds in two stages:
(1) NLP researchers tailor narrow NLP tools for the given application (e.g., automated podcast transcription).
(2) Experts of the given application lead the deployment and iterative improvement using user feedback (e.g., the copy writer who would usually write the transcripts).

Collaboration between NLP researchers and experts on applications is therefore crucial. The NLP4SG initiative places a big focus on improving this collaboration.

We note that this section only provides example applications. As mentioned above, NLP4SG has a broader approach and is concerned with other issues, such as directing NLP research to be more EA aligned.

---

[2] An Artificial Intelligence Algorithm to Identify Documented Symptoms in Patients with Heart Failure who Received Cardiac Resynchronization Therapy, Leiter et al.
[3] Topic based classification and pattern identification in patents, Venugopalan et al.

The Initiative in Depth

Who We Are

We are a group of full-time NLP researchers aiming to have a positive impact on the world through our work. Several of us are influenced by EA ideas and/or identify as EAs, which distinctly manifests in our initiative's mission and work. All organizational work is on a volunteer basis.

The NLP4SG initiative is led by Zhijing Jin (AI PhD at Max Planck Institute & ETH) and Dan Lahav (AI researcher at IBM Research) as well as supported by Prof. Rada Mihalcea (University of Michigan), Prof. Mrinmaya Sachan (ETH), Brian Tse (Policy affiliate at GovAI, Oxford, and founder of Confordia Consulting, China), Prof. Bernhard Schölkopf (Max Planck Institute & ETH), Dr. Joel Tetreault (Senior Director of Research at Dataminr), and Geeticka Chauhan (PhD at MIT). Our current volunteers are Luise Wöhlke, Peter Jacobs, and Flavio Schneider.

What We Do

As we said earlier, the approach our initiative takes is twofold:
(1) **Helping researchers** have an impact by promoting EA-aligned research agendas and providing resources on having an impact in NLP.
(2) **Helping philanthropic/charitable organizations** by connecting them with NLP technologies & researcher networks that could benefit their cause.

Two worrisome phenomena in the NLP community motivate this:
1 - Good will, bad execution. The good news is, the NLP community has a strong wish to promote social good. However, NLP researchers are seldom also experts in matters of social good, and lack many of the tools EAs use. Prioritization is an example of this. Most existing NLP4SG papers concentrate on topics such as online hate speech detection, translating languages with little training data, and reducing gender and racial bias in language models. This is incredibly important and we deeply value this work. However, we'd wish for proportionally many papers on higher impact issues. See Jin et al. (2021) for a more comprehensive review on how the NLP community connects to social good.
2 - Lacking coordination. During discussions with researchers at NLP conferences, we observed that many academic efforts
(1) are disconnected from one another across different labs (so there are reinvented wheels, and enduring obstacles that could be solved collectively),
(2) do not collaborate with experts on real applications, or
(3) cannot find people to actually deploy the technology (so many good papers have no impact, and the industry reinvents the wheel later on).

We tackle this by **helping researchers** to improve NLP4SG execution & coordination in the field and **helping impactful organizations** to improve coordination between them and NLP4SG. We thus hope to create a vibrant ecosystem that supports existing NLP4SG projects and allows many more such projects to be created.

EA ideas are a great guide in doing this. Cause prioritization and cost-effectiveness analysis for example are core EA ideas that can be directly translated to NLP4SG to help researchers. We think by building a vibrant community like EA, we'll be able to help more NLP researchers and more impactful organizations.

Our Roadmap

Stage 1) Launching the NLP4SG community [Completed in 2021]
Stage 2) Collectively finding goals and making a plan [Ongoing]
Stage 3) Applying NLP4SG [Ongoing]

- Subgoal-1: Better aligning NLP4SG research agendas with EA principles
- Subgoal-2: Promoting cross-lab collaborations
- Subgoal-3: Promoting NLP-cause area collaborations
- Subgoal-4: Fostering and growing an active community

Stage 1) Launching the NLP4SG community [Completed in 2021]

- Start thoughtful, goal-oriented discussions on NLP4SG.
- Involve social scientists, political scientists, experts on important causes, and many NLP researchers to share perspectives on NLP4SG and brainstorm practical things to do.
- Establish the NLP4SG initiative.
- **Our deliverables:**
    - We established the NLP4SG initiative (nlp4sg.vercel.app).
    - We laid out our guiding principles in the position paper "How Good Is NLP? A Sober Look at NLP Tasks through the Lens of Social Impact" (Jin et al., 2021).
    - We set up a slack channel of 150+ NLP researchers around the world.
    - We started a Twitter account (@nlp4sg) to broadcast progress in NLP4SG to the research community.

Stage 2) Collectively finding goals and making a plan [Ongoing]

- We decided our strategy should be rooted in *what is realistic for the existing NLP community to steer towards* (based on existing philosophy and social dynamics).
- We engaged 100+ NLP researchers in discussions on topics such as which goals we should pursue, how to evaluate progress, funding opportunities to

support NLP4SG, and whether to establish separate guidelines for reviewers to judge the social value of a paper in addition to the technical value.

- **Our deliverables:**
    - We collected survey responses of 70 NLP researchers (with ~30% professors) on their expectations for NLP4SG.
    - We analyzed NLP papers published in 2020 to map out the space of NLP4SG research topics and their popularity (Figure 4 of Jin et al. (2021)).
    - We organized the first discussion session engaging ~70-100 NLP researchers at ACL 2021 (the global top NLP conference).
    - We initiated the annual workshop on NLP for Positive Impact co-occurring with top NLP conferences such as ACL 2021 and EMNLP 2022. Our most recent workshop involved keynotes and panel discussions by leading NLP researchers in the community as well as Prof. Baobao Zhang, who contributed expertise on AI governance.

Stage 3) NLP4SG in action [Ongoing]

- Subgoal-1: Better align NLP4SG research agendas with EA principles. Help close the gap between current NLP research topics and actual social needs.
- Subgoal-2: Promote cross-lab collaborations. Connect lab efforts on similar research topics.
- Subgoal-3: Promote NLP-cause area collaborations. Connect NLP researchers and experts on important causes/charitable organizations.
- Subgoal-4: Foster and grow an active community. Increase the number of involved researchers and followers on social media. Gain more supporters in EA and NLP research.
- **Our deliverables:**
    - We curated an extensive list of NLP4SG papers (to save literature review efforts, and avoid reinventing wheels).
    - We built a visual navigation tool for browsing the landscape of existing NLP4SG work.
    - We have more ongoing work on computationally analyzing how aligned existing research is with social needs. Stay tuned.
    - (In the long run, we hope to build a website analogous to GiveWell, but giving advice to *researchers*, on how best to contribute given their skills and time. We welcome all kinds of help as described in the section "How You Can Help Us".)

Potential Concerns

AI safety

- NLP4SG focuses on using narrow NLP tools for social good applications. It seems highly unlikely that strong NLP safety is a major concern for several reasons:

    1. We focus on applying **existing** tools to important causes. This mostly plays out in creating adaptations of current technologies, rather than groundbreaking capability advancements that may promote AGI.
    2. The initiative is led by people in and around the EA community, who take EA ideas, including AI risks, very seriously. We believe in promptly and transparently communicating concerns and problems. We also welcome involvement from the AI safety community in order to steer clear of any potential future problems.
    3. In our role to coordinate research efforts and advise researchers on impactful research agendas, we are uniquely well-placed to *identify* safety concerns in the field early and steer people into useful, safe NLP4SG research. Thus we aim to net *reduce* safety risks from NLP research overall.
- That being said, we do not take safety concerns lightly, and we would like to stress again that we welcome the involvement of more safety experts.

Narrow NLP Concerns

- Bias: Machine learning-based NLP models do not necessarily end up with similar accuracy for people of different demographics. Models might produce unexpected errors for minority groups or non-English speakers. To address this, there is an increasing number of research papers on mitigating bias in NLP models.
- Generic issues with the deployment of new technologies
    - Potential stakeholders & misuse: An important question is who will use the technology. It remains a political and social challenge to constrain big players such as governments and large companies from the misuse of NLP, such as for surveillance.
    - Data privacy: The current NLP community requires strong justification if user data is involved in a study, but being cautious about data privacy still can't hurt.
- The NLP and NLP4SG communities are undertaking great research efforts to improve on these shortcomings (for example via forming ethical committees). In the future, we'd also like to start thinking about prioritizing between these efforts better, for optimal resource allocation.

We are very interested in discussing potential concerns about NLP4SG. If you have any thoughts, please get in touch!

Lessons Learned

1. Fostering mutual understanding:
    a. **Starting too fast:** When introducing EA to NLP through our initial paper in Jan 2021, we received mixed reviews. We've learned to think

about how to start from the NLP community's existing understanding of social good, and then gradually integrate EA ideas.

b. **Surveys and open discussions are helpful:** We've found the survey we conducted on how NLP researchers prioritize helpful to understand their philosophy. The same goes for the open discussions with NLP researchers we facilitated.

c. **More challenges are ahead:** There are more differences in thinking between the EA and NLP communities that we will need to attend to. For example, it might be difficult to understand for EAs why NLP researchers prioritize issues such as tools for languages spoken by only a couple hundred of people over applications with larger scale. And it might be difficult for NLP people to adopt the notion of *effectiveness*, since it is difficult to derive good estimates of what the impact of a new technology will be.

2. **Aiming to shift from top-down to bottom-up organization:** Currently the direction NLP4SG takes is mostly determined by a few organizers, but we would like this to shift more towards bottom-up collective decision-making. Large discussions involving many NLP researchers have been fruitful for us. For example, we had a great discussion session at ACL with 50-100 participants on four key topics: how to build an NLP4SG community, new or speculative NLP4SG agendas and prioritization, NLP for climate change, and NLP for misinformation and healthcare. We further want to have a more active slack channel and more peer-to-peer discussion opportunities.

How You Can Help Us

We really appreciate the following types of support (ways to contact us at the bottom):

- **Experience in cause prioritization:** to compile a list of top NLP4SG causes.
- **Expertise in a cause area:** to introduce NLP researchers to real and pressing problems & together identify ways NLP can help.
    - In our upcoming workshop in Nov 2022, we'll have a session of lightning talks to pitch pressing problems and/or NLP4SG work. If you are interested, please send us a 2-min recording as a Google drive file link to nlp4sg.info@gmail.com. We will organize a session to show the recorded lightning talks. Please note that we will prioritize recordings that are informative and match the audience's interests.
    - Feel free to recommend good candidates to us to invite to our upcoming workshop. We are looking for panelists/keynote speakers who can introduce the needs of cause areas to NLP researchers. For example, people working on the frontier of education in poor regions can potentially introduce what type of text or speech assistants might be helpful and what are the budget constraints for such devices.

- **Help with datasets:** NLP research is powered by datasets. We are very grateful for pointers to good datasets that NLP4SG research can use, or parties who can help build a dataset.
- **Volunteers:** We need help maintaining the NLP4SG website and running our mailing list. We are also looking for people to help with research, e.g. by providing NLP modeling expertise, writing data annotations, or drafting research agendas.

  We'll prioritize volunteers with experience in Computer Science.
- If you are interested in keeping up with our work, you can also follow our Twitter accounts: @nlp4sg, @ZhijingJin, and @LahavDan.

If you would like to stay in touch, recommend courses of action, get updates, endorse this initiative, offer funding, collaborate with us, or support this initiative in some way, please fill in this Google Form [link]. It might take us a while to get back to you, but we will be very grateful for your support! You are also very welcome to get in touch via email with jinzhi@umich.edu (Zhijing) and danlahav@gmail.com (Dan), or write directly to nlp4sg.info@gmail.com.