Participatory Design for Positive Impact: Behind the Scenes of Three NLP Projects

Marianne Wilson[1] and **David M. Howcroft**[2*] and **Ioannis Konstas**[3]
and **Dimitra Gkatzia**[1] and **Gavin Abercrombie**[3]
[1]Edinburgh Napier University {m.wilson2, d.gkatzia}@napier.ac.uk
[2]University of Aberdeen david.howcroft@abdn.ac.uk
[3]Heriot-Watt University {g.abercrombie, i.konstas}@hw.ac.uk

Abstract

Researchers in Natural Language Processing (NLP) are increasingly adopting participatory design (PD) principles to better achieve positive outcomes for stakeholders. This paper evaluates two PD perspectives proposed by Delgado et al. (2023) and Caselli et al. (2021) as interpretive and planning tools for NLP research. We reflect on our experiences adopting PD practices in three NLP projects that aim to create positive impact for different communities, and that span different domains and stages of NLP research. We assess how our projects align with PD goals and use these perspectives to identify the benefits and challenges of PD in NLP research. Our findings suggest that, while Caselli et al. (2021) and Delgado et al. (2023) provide valuable guidance, their application in research can be hindered by existing NLP practices, funding structures, and limited access to stakeholders. We propose that researchers adapt their PD praxis to the circumstances of specific projects and communities, using them as flexible guides rather than rigid prescriptions.

1 Introduction

Participatory design (PD) is an approach to design and development that actively involves stakeholders, such as end users, customers, and citizens in the design process to ensure the resulting products meet their needs and reflect their values (Spinuzzi, 2005). PD engages with stakeholders using various methods, including surveys, focus groups, interviews, and workshops. While the Human-Computer Interaction (HCI) community has employed PD for decades, it remains less common in the Natural Language Processing (NLP) community, likely due to the field's fast-paced development and focus on leaderboards. However, there is a growing recognition within NLP research of the need to move beyond purely technical approaches in order to create positive outcomes for stakeholders: researchers and practitioners are encouraged to engage in participatory and co-design approaches to better understand and address the needs of communities affected by NLP (Costanza-Chock, 2020; Parker and Ruths, 2023). This reflects a broader perspective in which 'positive impact' is defined by the communities these systems aim to serve, rather than by research agendas, and highlights the importance of evaluating NLP systems on how well they meet objectives defined in collaboration with communities, rather than relying solely on performance metrics.

The adoption of PD in NLP has been suggested as a means to alleviate contemporary issues with NLP technology, namely bias, fairness, and usability (Caselli et al., 2021). By involving diverse stakeholders in the design process, PD supports the development of NLP systems that are more inclusive and representative of various user groups, thus fostering equitable outcomes. Additionally, incorporating stakeholder input early and iteratively can benefit developers by improving the suitability of NLP applications for real world use cases. Caselli et al. (2021) suggested nine *principles* to guide the adoption of PD in NLP systems. These focus on community-based practice, fictional design scenarios, and enhanced reflexivity throughout data collection, annotation, deployment, and evaluation. More recently, Delgado et al. (2023) proposed a *framework* for evaluating PD of AI systems. This focuses on various dimensions of participation that range from consultation to ownership in terms of design goals, scope, and methods. An overview of Delgado et al.'s *principles* and Caselli et al.'s *framework* are provided in Tables 1 and 2.

In this paper, we use these *principles* and *framework* as lenses to reflect on three research projects which did not all necessarily aim to be maximally participatory from the outset. The projects present intricate challenges, varying risk levels and degrees

[*] Contribution conducted at Edinburgh Napier University.

Proceedings of the Fourth Workshop on NLP for Positive Impact (NLP4PI), pages 252–263

	Consult	Include	Collaborate	Own
Participation Goal	Why is Participation Needed?			
	To improve the user experience	To better align AI with stakeholders' preferences and values	To deliberate about system features	To shape the systems scope and purpose
Participation Scope	What is on the table?			
	User interface of the system	Underlying datasets (e.g. identification, curation, annotation)	Overall design of system (e.g. task specification, model features)	Whether and why the system should be built
	Who is involved?			
	Stakeholders recruited by the project team for discrete feedback	Stakeholders recruited by the project team for domain expertise	Stakeholders designated by the community collaborate in design	Stakeholders designated by community play a central role across the project lifecycle
Form of participation	What form does participation take?			
	Giving input of design ideas via questionnaires and interviews	Group discussions with project team	Ongoing collaborative protyping and decision-making	Reflexively deciding on the participatory approach

Table 1: PD for AI framework reproduced from Delgado et al. (2023)

1. PD is about consensus and conflict	• PD entails a process of mutual learning between researchers and community • PD adopts a variety of research and design methods (workshops, participants observation, cards, ...)
2. Design is an inherently disordered and unfinished process	• Use-before-use: tool's use is envisioned before the tool is actually implemented • Design-after-design: tool's design isn't exhausted with delivery, but will be modified by the users' appropriation, use, and feedback
3. Communities are often not completely determined a priori	• Communities are not a unitary whole, but can get formed within and through the design process
4. Data and communities are not separate things	• The shift from language as data to language as people: language data are produced by human speakers • Communities should be involved in the different stages of the NLP pipeline
5. Community involvement is not scraping	• Collaboration with a community should imply ethical engagement practices based on respect, equity and reciprocity • Researchers should communicate to the community the usage of the collected data in a transparent and appropriate way
6. Never stop designing	• Community adaptation should be treated as a feature of an NLP system at the design stage
7. Text is a means rather than an end	• The linguistic output of NLP systems should serve people's needs rather than imitate people's production of language.
8. The thin red line between consent and intrusion	• Do not assume that community members are technology experts nor technologically illiterate • A community's refusal to collaboration is a risk that must be accepted
9. The need to combine research goals, funding, and concrete social political dynamics	• Designers and researchers as intermediaries between the interests of the different actors involved (project beneficiaries, investors, funding agencies, and other stakeholders' goals)

Table 2: Principles for PD in NLP reproduced from Caselli et al. (2021)

of participation: the first project focuses on developing language technology for a low-resource language focusing on museum artefact descriptions (SGGE), which presents fewer risks; the second project designs a chatbot for youth career support (CSC), with moderate risks and significant impact (CSC); and the third aims to design an annotation framework to tackle gender-based abuse online, involving harmful and potentially triggering content (ESO). Each project uses different methods to meet PD objectives: community expert engagement (Pillai et al., 2023), the Delphi method (Linstone and Turoff, 2011), and focus groups (Morgan, 1996).

The intention of this exercise is not to objectively report results for each project, which are available in other publications (Howcroft et al., 2023; Wilson et al., 2024, Forthcoming). Here, the aim is to explore the relationship between the *principles* and *framework* and the practicalities of using PD in NLP research by focusing on our experience as researchers. Incorporating our own inherent subjectivity into this reporting also furthers PD praxis by providing a concrete illustration of the process of researcher reflexivity that is central to PD.

This paper makes the following contributions: (1) providing methodological insights into three distinct NLP projects that incorporate PD in their design; (2) evaluating these projects with respect to Delgado et al.'s framework and Caselli et al.'s principles to further understanding of how these can be

applied in practice; and (3) supporting researchers interested in adopting these approaches in their own work by providing insights into researchers' experiences conducting participatory NLP research.

Our analysis indicates that the framework and principles should not be used as prescriptive templates. Instead, they are useful for consultation as researchers plan and evaluate their individual research projects. However, researchers should consider other context-specific factors when incorporating participatory design in their research, such as the nature of the communities they are working with, funders' requirements, project timescales, and the expertise and knowledge available to support their projects.

2 Background and related work

PD originally emerged in Scandinavia and has its origins in workplace democracy, political participation, and feminism (Gregory, 2003; Muller and Kuhn, 1993; Spinuzzi, 2005). Its core tenets are to provide a voice to people who lack expert design skills and to ensure designers remain accountable for the impacts on users and communities. This is achieved by approaches to engagement that aim to re-balance power relations between users and designers (Robertson and Simonsen, 2012).

While PD can be viewed both as a research methodology and a design approach (Spinuzzi, 2005), its implementation varies and may draw upon different methods to create *hybrid spaces* that bring together users & designers to challenge assumptions and for mutual learning & co-creation. Practitioners highlight the importance of local context and knowledge (Ehn, 1988); 'design as change', whereby possible futures, new tools, changed infrastructures, and interactions are envisioned (Gregory, 2003); and dissensus, with pluralities of conflicting positions inherent to the design process (Caselli et al., 2021; Keshavarz and Maze, 2013).

The shift to PD means going beyond extractive forms of research participation, where people are sources of data, labels or evaluation (Birhane et al., 2022), and as such represents an alternative approach to traditional user-centric design which does not allow for users to become full collaborators or own the direction of a project. Instead, a range of stakeholders should have an opportunity to shape the research process and outputs. This can include, a number of approaches, such as participatory research (i.e. citizen or open science (e.g. Nekoto et al., 2020; ECSA, 2015); participation in model deployment and use (surveyed in Wang et al., 2021); and participation in evaluation or feedback for NLP systems (e.g. Heuer and Buschek, 2021; Knoll et al., 2022). Delgado et al. (2023) and Caselli et al. (2021) reviewed existing research to develop a *framework* and *principles* for PD in NLP. In turn, this paper uses these as a lens to reflect on three projects, offering practical insights into the application of PD in NLP.

Framework Delgado et al. (2023) reviewed work that claims to engage in PD, considering the research goals, scope and methods, mapping these to the spectrum of participation modes: *consult - include - collaborate - own*. Their analysis found that most of the 59 publications engaged in *consultation* rather than *ownership*. They concluded that computer scientists need to do more than simply 'add diversity and stir' to achieve true PD.

Principles Caselli et al. (2021) developed nine principles for PD in the development of NLP systems. The principles integrate general PD principles and NLP practice, with an emphasis on researcher reflexivity. Summaries of the *framework* and *principles* are available in Tables 1 and 2.

3 Three PD NLP projects

In this section, we provide an overview of the projects as context for the analysis using the *framework* and *principles* discussed in §4

3.1 Scottish Gaelic Generation for Exhibitions

The Scottish Gaelic Generation for Exhibitions (SGGE) project was a first step towards developing Scottish Gaelic chatbots for use in museums (Howcroft et al., 2023). This work was not meant to achieve full bidirectional collaboration between researchers and the community but was intended as a proof of concept for the research team to develop relevant skills and expertise for interfacing with the community while meeting the professional expectations of their research community. In this project, we collaborated with the National Museum of Scotland to identify exhibits which might be of special interest to speakers of Scottish Gaelic. We then recruited speakers of Gaelic to participate in text-based conversations about the exhibits, with one member of each pair of speakers playing the role of a museum guide or curator and the other a museum

Participation	SGGE	CSC	ESO
Goal	**Consultation**: improve user experience	**Own**: Shape system's scope and purpose	**Collaborate**: Ongoing focus groups & engagement
Scope: What	**Include**: Underlying datatsets	**Own**: Overall design of system	**Include/Collaborate**: Design of annotation schema
Scope: Who	**Include**: Domain expertise	**Include**: Domain expertise	**Include**: Domain expertise
Form	**Consultation**: input on data collection process and UX design	**Unclassified**: Delphi study (questionnaires for collaborative decision-making)	**Collaborate**: Ongoing focus groups & engagement

Principle	SGGE	CSC	ESO
1. Consensus and conflict	✗	✓	✔
2. Disordered and unfinished	✔	✔	✔
3. Communities not defined a priori	✔	✓	✓
4. Data and communities not separate	✔	✔	∼
5. Community involvement <> scraping	✔	✔	✔
6. Never stop designing	✗	✔	✔
7. Text is a means, not an end	∼	✗	✔
8. Consent v intrusion	✔	✔	∼
9. Balance competing goals	✔	✓	✓

Fully Aligned: ✔ Somewhat Aligned: ✓ Planned alignment: ∼
Not Applicable or Not Aligned: ✗

Table 3: Overview of the projects through the lenses of Delgado et al. (2023) and Caselli et al. (2021).

visitor, inquiring about an exhibit. The task was designed to make it possible to train chatbots to interact with museum visitors.

3.2 Career Support Chatbot

The Career Support Chatbot (CSC) project, in collaboration with Skills Development Scotland (SDS),[1] developed design & evaluation criteria for a chatbot to support their government-funded career support for young people. Given the complexity of career support, and potential impacts on socio-economic outcomes for individuals and society, it is crucial to ensure that interventions are designed in a way that maximizes positive impact. Therefore, a panel of SDS staff were invited to participate in a Delphi study (Linstone and Turoff, 2011), to identify an appropriate task, and to validate requirements for the chatbot (Wilson et al., 2024). Delphi studies use successive rounds of anonymous questionnaires to structure communication between a group in a way that builds consensus, while mitigating issues of groupthink and anchoring effects (Zartha Sossa et al., 2019). The requirements are articulated in the panel's own words, and therefore provide a strong foundation for further participation in chatbot design and evaluation.

3.3 Equally Safe Online (ESO)

Equally Safe Online (ESO)[2] has been developed in partnership with non-governmental organisations (NGOs) to tackle online gender-based violence (oGBV), initially by co-designing a taxonomy as the basis of annotation guidelines and to train classifiers. The proposal was developed in collaboration with three NGOs, and the team includes a researcher with both academic and NGO affiliations.[3] We have held workshops and focus groups in a mix of online/in-person and single-/multi-NGO settings, consisting of introductions to taxonomies, discussions of how participants encounter (o)GBV, and hands-on exercises to develop categories. As the project has progressed, more organisations have become involved. The output is a taxonomy, that is now being used in further PD work to create datasets and models, with the aim of developing solutions that support the NGOs to combat oGBV.

4 Analysis

We now apply Delgado et al.'s (2023) *framework* and Caselli et al.'s (2021) *principles* to structure our reflection on these projects. Italics are used to highlight where specific concepts from these sources are used. With respect to Delgado et al. (2023), we map the goal, scope, and form of participation of each project on the consultation to ownership spectrum. With regard to Caselli et al. (2021), we identify the extent to which each project's planning and execution exemplify the participatory design principles (1-3), principles for NLP tools (4-6), and principles of researcher reflexivity (7-9). See Table

[1] https://www.skillsdevelopmentscotland.co.uk/

[2] https://sites.google.com/view/equallysafeonline

[3] Additional participants have joined the co-design sessions. Details of all the organisations are presented in Appendix A.

3 for an overview of the position of each project with respect to Delgado et al.'s framework and the nine principles of Caselli et al.

4.1 Delgado et al.'s (2023) framework

The projects achieved different **participation goals** within the *framework*.

The research objective of SGGE was *consultation*, with participants' input expected to help improve user experiences. ESO's goal has been more wide-ranging, conceived as a co-design *collaboration* between researchers and participants. The goal of CSC, meanwhile, was to support domain experts to define the system task, defined as *ownership*.

The **scope** of *what* participants' contribution encompassed also varied. For SGGE, this mapped to *include*: providing underlying data. While it proposed to *collaborate* with participants, ESO's scope has so far fallen between this and *include* on the *framework*. From the start, the researchers did not consider how to offer more 'on the table' than design of annotation schema and datasets, with no specific plans to involve participants in model design. However, researchers and participants are engaging in a *collaborative* effort to design the entire annotation module of the project, rather than merely using participants as annotators, as is the case in most NLP dataset creation efforts. Allowing communities to shape the research process is a defining feature of PD in other disciplines, that has limited representation within the *framework*. On this dimension, CSC again mapped to *ownership*, as career experts were able to define the overall objective of the system. However, ownership did not extend to potential system users.

In terms of the *who* of participation scope, all three projects are at the *include* level. SGGE recruited participants on the basis of their linguistic expertise. Participants in ESO, were defined to some extent at the proposal stage, when formal partner organisations were approached. We initially engaged only with the heads of the stakeholder organisations, though more junior members of the NGOs have since been able to join the co-design sessions. Similarly, CSC participation was defined by the careers service that collaboratively funded the project. However, staff from across the organisation were invited to participate to include a broad scope of experiences in the panel.

The **form** of participation also varied. For SGGE, input was primarily *consultation* given through participation in the experiments, providing conversations and summaries which could be used in future chatbot development, with some additional input given through email conversations about the task itself. In addition to the experimental participants, the projects' Gaelic specialist served as a representative of the community in the *collaboration* and *ownership* levels of the framework, helping to shape the task design and how best to include experimental participants in the study.

For ESO, participation has fallen somewhere between *include* (i.e. 'group discussions with the project team) and *collaborate* (i.e. 'ongoing collaborative prototyping and decision making'), with workshops and focus groups, sometimes ongoing with the same participants. Co-creation has involved collective prototyping, as the researchers work towards guided decision making by stakeholders, as co-design develops.

By applying the Delphi method, CSC aimed for *collaboration* to identify an appropriate career support task for automated intervention (i.e. the overall design of the system was on the table). However, this method does not clearly align with a single category in the framework: questionnaires collect the panel's views (*consultation*), but these are used to facilitate the panel's internal deliberations and decision-making and, combined with the iterative and responsive design of the questionnaires, mean that they function more like *collaboration*.

4.2 Caselli et al. (2021)'s Guiding Principles

The projects are not as easily compared against the *principles* as the Delgado et al. (2023) framework. Instead, we describe whether and how each principle applied to the projects, as shown in Table 3.

PD Principles (1-3) SGGE's process did not include a lot of *consensus and conflict* (1) as it pursued a primarily researcher-driven protocol. This was in part necessitated by the need to create a concrete proposal to secure funding for our engagement at this early stage of our research. The process, however, was (2) *inherently disordered and unfinished* as we needed to develop a new interface for our experimental design and iterate on that design as we engaged with the community, fixing bugs in our interface and updating the scheduling protocol to accommodate users. The communities of interest were indeed *not completely determined a priori*: while the researchers knew that there are a variety of speakers of Scottish Gaelic belonging to different speech communities based on geogra-

phy, heritage, education, age, etc, they cast a broad net during recruitment, resulting in a mix of self-selecting speakers who did not necessarily form a single community nor a representative sample of the population of Gaelic speakers.

CSC used a Delphi study approach. Delphi studies are commonly used as a method for building *consensus* (1) on a topic between the panel (Linstone and Turoff, 2011). Here, consensus does not mean measuring the majority opinion, or unanimity, but using the process to support the panel to construct an output that accurately reflects the nuanced views of the panel as a whole. CSC used design fictions (Dunne and Raby, 2013) in the first Delphi questionnaire, which were explicitly *use-before-use* (2) in that they described a range of imagined situations in which users could use a chatbot for different aspects of career support. Although the boundaries of the community were defined based on the collaborating organization, the community was not completely defined *a priori* (3) as the panel was assembled based on recommendations by gatekeepers within the organization, and required voluntary agreement to participate.

ESO has sought *consensus* from a variety of participants. Although encountering some *conflict* due to different backgrounds, this has not prevented broader agreements being reached. *Mutual learning* has occurred although has been lacking in both directions at times, as we have needed to explain technical background or collect more feedback (1). We have used two PD methods and a variety of activities to stimulate collaboration, adapting to different dynamics, with output altering as new voices have been incorporated and previous ideas adapted and altered (2, 3). We are discussing with participants *the tool's envisioned use before [...] implement[ation]*, but are not yet at the *design-after-design* stage, which may be problematic due to academic research funding structures.

4-6: NLP tools Looking to Caselli et al.'s (2021) reflections on NLP tools, SGGE aimed to explore language use, but also to produce linguistic (and non-linguistic) *data* which could be used to develop chatbots and summarization systems (4). While this inherently involved speakers from the Gaelic community(/ies), these participants were not involved in the selection or future refinements of the tasks. This research did not involve *scraping*, and participants were invited to take credit for their participation in the study and contribution to research on Gaelic (5). In this stage of prototyping, it was not yet appropriate to pursue further community adaptation in line with the principle to *never stop designing*, as we need to be mindful of participants' time and would need chatbots ready for testing to warrant further engagement (6).

CSC involved *communities* in order to define the system purpose, rather than to facilitate the collection of *data* for use in chatbot development. When explicitly asked for anonymous feedback about participating in the study, participants mentioned benefits to their own praxis, independent of contribution to the research aims, suggesting *ethical engagement* (5). While the project eventually has a chatbot as its natural end product, the requirements are rooted in *community adaptation* (6).

In ESO, participants have so far been involved only in conceptualisation and taxonomy development and not in the design of automated modelling. We aim to do this in future parts of the project (4), although we initially wish to harness the expertise and limited time of our partners in developing the framework on which these systems will rely. We do not believe we are involved in community *scraping* (5), although we do have different incentives than participants (e.g. publishing papers vs. influencing policy). We *communicate ... the usage of the collected data* through information sheets, consent forms, and in-person explanations. We are trying to build *community adaptation* (6) into the design of the project, and treat small localised solutions as a benefit rather than a limitation (as argued in Abercrombie et al. (2023)).

7-9: Researchers' reflexivity For SGGE, these principles have become more relevant as the work on Gaelic continues beyond the scope of the original pilot. Our goals were collecting linguistic data for model development, working on a prototype, and beginning to develop connections with speakers that we could build on in the future, rather than releasing a 'completed' system which would require further input from the community to ensure its utility. This aligns with principle that *text is a means, not an end* (7). Our recruitment process met potential participants where they already were using social technologies, and in this way was minimally *intrusive*, though we indeed did not find as many participants as we had hoped (8). One of our collaborators is a member of the Gaelic community and an ethnologist, and we are indebted to him for his willingness to join our project. Indeed,

we would not have felt comfortable pursuing this effort without this guidance and support, helping us craft a funding proposal and our research efforts in a way that *balanced research goals with the goals of the the community* (9).

The Delphi method of CSC aligns with the principle of *serving the community needs* (7), in that the study was designed intentionally to allow community members to identify which need(s) they believed a chatbot could serve. Seeking voluntary participation, leveraging design fictions, and including optional free-response components in our surveys helped to ensure that participants could *contribute as much or as little as they wished, regardless of their knowledge of chatbot technology* (8). The fact that only 3 of the original panel of 23 dropped out during the research indicates that participation was not considered *intrusive*. Finally, the consensus-building elements of the Delphi study facilitated the researchers' *role as intermediaries* (9). The method is designed to support successful *navigation of competing priorities* between individual participants.

ESO was not involved in collecting *text*. Therefore, avoiding extractive methods has been focused on balancing *consent and intrusion*. We have not encountered *refusal to collaborate* as the stakeholder groups have agreed to be partners and/or participants. Where we have encountered groups who do not wish to engage, we have accepted this. *Balancing competing goals* of research, funding, with the complex social political dynamics of the domain and the communities (9) is the most challenging aspect of co-design. Due to funding dynamics and the project timeline many decisions had to be made on the topic and scope of the project before real stakeholder engagement could occur.

4.3 Reflections

Building on the insights gained from considering the projects in light of Caselli et al.'s *principles* and Delgado et al.'s *framework*, we reflect on our own PD practice.

SGGE Overall, this project represents the pilot-project stage of participatory research in NLP: we needed to learn more about the community and identify avenues of research where our existing expertise could be relevant to our first steps toward working with Scottish Gaelic communities. Our intention is to continue to build on these experiences with future grants with the aim of building collaboration to enhance gaelic speakers engagement with cultural heritage collections.

CSC The Delphi study allowed participants and researchers to collaboratively work towards outputs in a format that served the needs of the research and stakeholders. The use of asynchronous questionnaires and design fictions helped to reduce the burden placed on participants. The requirements are captured using language accessible to SDS staff to provide a solid foundation for future ownership by them when the research is complete. Preparatory work by the researchers to develop an understanding of career support was essential for meaningful participation. The aim was to design a study that would reduce the risk of unintended negative consequences of introducing a chatbot into a high-risk domain, within the limitations of the experts' available time and prior knowledge of chatbots.

ESO We did not begin with a clear picture of what PD would entail. This resulted in some challenges as the project evolved. Despite this, we aim to avoid using PD 'to provide legitimacy for preexisting plans' (Costanza-Chock, 2020). As ESO is ongoing, we can change our practice to e.g. collect more feedback and foster more collaboration.

Two years in, we do not yet have completed datasets or experiments. With under-resourced NGOs, it can be difficult to arrange sessions, which get postponed or cancelled due to other pressures. This has had negative effects on our ability to *never stop designing* (Caselli et al., 2021) with the same participants, leading some sessions to feel more like *consultation* than *collaboration* (Delgado et al., 2023). However, we are already seeing a more stakeholder-centred outcome compared with that of previous work in the same domain, which recruited participants to work to researcher-defined specifications (e.g. Cercas Curry et al., 2021). Where the annotation schema used in that project was based solely on previous NLP work, the specifications of the new taxonomy are driven by the participants.

5 Lessons Learnt

The PD framework and principles are useful—but different Both perspectives on PD provide different affordances. Delgado et al.'s (2023) framework is a useful lens for mapping a project's position within the broader participatory turn, while Caselli et al. (2021)'s principles are useful for researchers to reflect on their positionality

within a research project. Structuring our collective reflections on three unrelated research projects using Delgado et al.'s framework and Caselli et al.'s principles has yielded useful insights into the process of using PD in NLP that can be applied more generally.

The projects had different goals, scales and stakeholder communities. Each focused on a different stage of the NLP pipeline: dataset collection, task definition, and annotation taxonomy development. Nonetheless, all three successfully engaged their respective communities of interest in a meaningful way throughout the research process. As Table 3 shows, the projects can be situated within the framework for participatory NLP research (Delgado et al., 2023) and demonstrate alignment with the majority of the relevant principles (Caselli et al., 2021).

Participatory research is valuable Across the different topics, goals and methods, all of our projects have benefited from the work that has gone into engaging with the respective communities. Speculative consideration of our projects in comparison to the likely outcomes of 'traditional' approaches supports this. For SGGE, the traditional approach would depend on crowdsourcing text, removing the opportunity to connect with the community and understand their connection to the cultural heritage artifacts being discussed. Whereas, PD resulted in a richer dataset, that reflects the communities' relationship with the dataset topic. For CSC, this would have meant defining the chatbot task based on a gap in the research literature, without any assurance that this was aligned with career service practice, ethics or that the outputs would be comprehensible to stakeholders. Instead, PD supported the creation of detailed criteria that can be used for design and extrinsic evaluation of chatbots in this context. For ESO, this would be taking or adapting an existing taxonomy (as in Cercas Curry et al., 2021) or developing one purely from theoretical work (e.g. Guest et al., 2021). Instead, by using PD, ESO has focused on the requirements of stakeholders rather than the convenience of the researchers.

Based on our experiences, we are unanimous about the positive impact that PD has had on our research. While the planned outputs for the benefit of these communities have not yet been realised, participants' formal and informal feedback and willingness to engage indicate that they also perceive value in the process of participating.

Connecting with participant communities is an ongoing process Building relationships with communities requires a shared understanding of each other's objectives, priorities, and processes. As our projects demonstrate, there are multiple ways to build these bridges, depending on the research conditions. For the CSC project, the collaborative funding model established the links with the community and created shared objectives. For SGGE and ESO, the research teams included people who were already members of the stakeholder communities. However, differences in the nature of the knowledge and experience needed from participants required different approaches to recruitment from each. SGGE advertised to recruit Gaelic speaking participants, who were motivated by their personal interest in contributing to the research and nominal monetary compensation. The specialist and sensitive nature of ESO's topic necessitated building relationships via established organisations, rather than recruiting individuals directly. This highlights the value of considering the potential plurality of stakeholder communities when designing participatory research, as opposed to conceptualising these as a 'unitary whole' (Caselli et al., 2021) that forms during the research.

PD looks different in different contexts Our reflections highlight the importance of embracing the inherent messiness of PD. As Delgado et al. (2023) emphasise, and our projects' varied positions on the framework supports, *ownership* is not intended as a universal target for researchers to aim for. Research design, participation goals and methods should be a product of the specific research objectives and context, rather than driven by a normative research agenda. To conduct participatory research with integrity, researchers should recognize that participants' time, knowledge, and experiences are as valuable as their own. This means adapting the research process to meet the specific, unique combination of community needs and research goals, rather than aiming to align with approved templates for 'good' participatory research. This requires researchers to develop their sense of methodological curiosity. Delgado et al.'s survey lists a wide range of methods that have been used in participatory AI design, many adapted from other disciplines. Selecting the most appropriate of these requires not only understanding the methods themselves, but also giving consideration to the researchers' skills and strengths, the research objectives and the

communities they are working with. Ideally, there would be scope to adapt or trial multiple methods within a single project, however, this is rarely feasible. Encouraging the collection and publication of feedback from participants about their experiences is a simple, but effective, contribution to PD practice in NLP.

Participatory research is not easy Nonetheless, we would also like to highlight some of the difficulties inherent to participatory research. In particular, it is important to emphasise that it can be slow. Even in situations where communities are eager to engage and have the resources to do so, there are inherent overheads involved. It is time-consuming to construct a shared language between researchers and stakeholders. But, this is crucial for meaningful collaboration that avoids exploiting participants or undercutting the value of their contributions. The conventions of traditional research outputs often obscure this aspect of PD. Research funding processes are also at odds with the nature of PD. Funders' requirements for detailed plans far in advance of actual engagements with communities means that, rather than 'accept the risk' (Caselli et al., 2021) that communities might refuse to collaborate, it is easier not to plan to engage. Even if communities have agreed to participate, academic funding timescales are lengthy. The delay between stakeholders agreeing to be involved and the research starting can jeopardise participation. Logistical obstacles to participation can unexpectedly delay recruitment and may require a high level of flexibility and organisational skills. The collection and analysis of the kind of data generated by participatory research may require specific skills. If these are not already available within the research team, then additional collaboration with external colleagues may be required, which adds an additional layer of complexity. Collectively, these issues can result in a prolonged journey between funding and publications for the researcher and any positive impact for the communities.

6 Conclusion

We do not highlight the challenges to deter potential participatory researchers, but to equip them with the knowledge needed to plan effective research. We find Delgado et al.'s (2023) framework to be useful for planning participatory research and understanding the landscape of PD in NLP. Caselli et al.'s (2021) principles provide valuable support for ongoing reflexivity. Both are helpful for improving our understanding of PD in practice. However, as Delgado et al. emphasise, PD research should not be evaluated based on its position on the *framework*. The goals, scope and form of PD research should be determined based on the specific research context.

Our practical advice to potential PD researchers takes a similarly pragmatic approach. Giving consideration to how PD could be incorporated into projects can benefit both researchers and stakeholders. In practice, this means taking time early in the research process to understand who might have an interest in the topic or be impacted by future applications of the research. Before finalising detailed research objectives or plans, we recommend engaging with those people, even informally, to explore opportunities for them to contribute to shaping the goals, scope, or form of the research. This should also take into account what is feasible within the social, political, and financial context of the research. As PD often requires flexible timelines and resources, this should be factored into funding proposals to ensure that the shared goals can be realized. Critical engagement with a wide range of research methods is also key. To this end, we encourage NLP researchers to be open to selecting and adapting methods to suit the specific context. Peer review and writing practices that normalize reporting feedback from researchers and participants about their experience of the research can help build our understanding of which methods are suited to different settings.

Compared to the usual research cycle, where a problem is explored from conception to publication (and then considered complete and dropped), PD is time-consuming. It requires a broad range of skills from the research team, and requires ongoing engagement from participating stakeholders. It is an uncertain process, requiring researchers to let go of ownership and see where the co-design process leads. However, the benefits of undertaking PD are outputs that are closely aligned with stakeholders' needs and that reflect their priorities and language. This means that the research is more likely to result in positive impacts for the communities it aims to serve.

Limitations

This examination of participatory design in NLP is limited by our own positionality as researchers at

established institutions in a wealthy nation with access to grant funding: the challenges we have faced are inherently different from those that researchers from other backgrounds might face. Moreover, being limited to assessing our own experiences means that there is room for self-assessment bias in our analyses and that the breadth of our study is limited. We believe that this is an acceptable trade-off for the depth of familiarity with the work discussed, which cannot be replicated by meta-analyses or survey papers. Building on Delgado et al. (2023) and the projects presented here, future work would benefit from a comprehensive survey of NLP research to evaluate how effectively participatory design methodologies can be applied when working with diverse communities and research goals.

Ethical Considerations

The projects described in this paper were all approved by the Institutional Review Boards of the respective universities (details withheld to preserve anonymity until acceptance).

All participants provided informed consent and were able to withdraw at any time. As the subject matter of ESO is particularly sensitive, to ensure participant welfare, we followed the guidelines of Kirk et al. (2022), by briefing participants before sessions, and limiting exposure to harmful content as far as possible.

For partner organisations of ESO, participants' involvement represented payment in kind as their contributions to the project. Other organisations that participated received payment.

For CSC, the risk of undue influence from the collaborating organization on the research is managed by ensuring that an independent award body is responsible for administering the project and funding.

Acknowledgements

We would like to thank the CLAN reading group at University of Aberdeen and Tommaso Caselli for their valuable feedback on drafts of this paper, as well as the Workshop Programme and Organising committees and previous reviewers from ARR. We would also like to express our gratitude to everyone who has contributed their time and expertise to each of the research projects.

Ioannis Konstas and Gavin Abercrombie were supported by the EPSRC project 'Equally Safe Online' (EP/W025493/1).

Marianne Wilson's PhD Studentship is funded by the Scottish Graduate School of Social Sciences, in collaboration with Skills Development Scotland. She is supervised by David Brazier, Dimitra Gkatzia and Pete Robertson.

David M. Howcroft and Dimitra Gkatzia's contributions were supported by EPSRC project 'NLG for low-resource domains' (EP/T024917/1), and data collection for SGGE was supported by the Creative Informatics small grant 'Scottish Gaelic Generation for Exhibitions'.

References

Gavin Abercrombie, Aiqi Jiang, Poppy Gerrard-abbott, Ioannis Konstas, and Verena Rieser. 2023. Resources for automated identification of online gender-based violence: A systematic review. In *The 7th Workshop on Online Abuse and Harms (WOAH)*, pages 170–186, Toronto, Canada. Association for Computational Linguistics.

Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '22, New York, NY, USA. Association for Computing Machinery.

Tommaso Caselli, Roberto Cibin, Costanza Conforti, Enrique Encinas, and Maurizio Teli. 2021. Guiding principles for participatory design-inspired natural language processing. In *Proceedings of the 1st Workshop on NLP for Positive Impact*, pages 27–35, Online. Association for Computational Linguistics.

Amanda Cercas Curry, Gavin Abercrombie, and Verena Rieser. 2021. ConvAbuse: Data, analysis, and benchmarks for nuanced abuse detection in conversational AI. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7388–7403, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Sasha Costanza-Chock. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. MIT Press.

Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The participatory turn in AI design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO '23, New York, NY, USA. Association for Computing Machinery.

Anthony Dunne and Fiona Raby. 2013. *Speculative everything: design, fiction, and social dreaming*. The MIT Press, Cambridge, Massachusetts.

ECSA. 2015. Ten principles of citizen science.

Pelle Ehn. 1988. *Work-oriented design of computer artifacts*. Ph.D. thesis, Umeå University.

Judith Gregory. 2003. Scandinavian approaches to participatory design. *International Journal of Engineering*, pages 62–74.

Ella Guest, Bertie Vidgen, Alexandros Mittos, Nishanth Sastry, Gareth Tyson, and Helen Margetts. 2021. An expert annotated dataset for the detection of online misogyny. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1336–1350, Online. Association for Computational Linguistics.

Hendrik Heuer and Daniel Buschek. 2021. Methods for the design and evaluation of HCI+NLP systems. In *Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, pages 28–33, Online. Association for Computational Linguistics.

David M. Howcroft, William Lamb, Anna Groundwater, and Dimitra Gkatzia. 2023. Building a dual dataset of text- and image-grounded conversations and summarisation in gàidhlig (Scottish Gaelic). In *Proceedings of the 16th International Natural Language Generation Conference*, pages 443–448, Prague, Czechia. Association for Computational Linguistics.

Mahmoud Keshavarz and Ramia Maze. 2013. Design and dissensus: Framing and staging participation in design research. *Design Philosophy Papers*, 11(1):7–29.

Hannah Kirk, Abeba Birhane, Bertie Vidgen, and Leon Derczynski. 2022. Handling and presenting harmful text in NLP research. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 497–510, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Tom Knoll, Francesco Moramarco, Alex Papadopoulos Korfiatis, Rachel Young, Claudia Ruffini, Mark Perera, Christian Perstl, Ehud Reiter, Anya Belz, and Aleksandar Savkov. 2022. User-driven research of medical note generation software. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 385–394, Seattle, United States. Association for Computational Linguistics.

Harold A. Linstone and Murray Turoff. 2011. Delphi: A brief look backward and forward. *Technological Forecasting and Social Change*, 78(9):1712–1719. The Delphi technique: Past, present, and future prospects.

David L Morgan. 1996. Focus groups. *Annual review of sociology*, 22(1):129–152.

Michael J Muller and Sarah Kuhn. 1993. Participatory design. *Communications of the ACM*, 36(6):24–28.

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Muhammad, Salomon Kabongo Kabenamualu, Salomey Osei, Freshia Sackey, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa Berhe, Mofetoluwa Adeyemi, Masabata Mokgesi-Selinga, Lawrence Okegbemi, Laura Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkadir Dangana, Herman Kamper, Hady Elsahar, Goodness Duru, Ghollah Kioko, Murhabazi Espoir, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Chinenye Emezue, Bonaventure F. P. Dossou, Blessing Sibanda, Blessing Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2144–2160, Online. Association for Computational Linguistics.

Sara Parker and Derek Ruths. 2023. Is hate speech detection the solution the world wants? *Proceedings of the National Academy of Sciences*, 120(10):e2209384120.

Malvika Pillai, Ashley C Griffin, Clair A Kronk, and Terika McCall. 2023. Toward community-based natural language processing (CBNLP): Cocreating with communities. *Journal of Medical Internet Research*, 25:e48498.

Toni Robertson and Jesper Simonsen. 2012. Challenges and Opportunities in Contemporary Participatory Design. *Design Issues*, 28(3):3–9.

Clay Spinuzzi. 2005. The methodology of participatory design. *Technical communication*, 52(2):163–174.

Zijie J. Wang, Dongjin Choi, Shenyu Xu, and Diyi Yang. 2021. Putting humans in the natural language processing loop: A survey. In *Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, pages 47–52, Online. Association for Computational Linguistics.

Marianne Wilson, David Brazier, Dimitra Gkatzia, and Peter Robertson. 2024. Participatory design with domain experts: A delphi study for a career support chatbot. In *ACM Conversational User Interfaces 2024 (CUI '24)*. Association for Computing Machinery (ACM).

Jhon Wilder Zartha Sossa, William Halal, and Raul Hernandez Zarta. 2019. Delphi method: analysis of rounds, stakeholder and statistical indicators. *Foresight*, 21(5):525–544. Publisher: Emerald Group Holdings Ltd.

A ESO Partner and Stakeholder Organisations

Partner organisations:

- EmilyTest: a Scotland-based charity focused on educating students about GBV https://www.emilytest.org/
- Glitch: a charity that tackles online abuse https://glitchcharity.co.uk/
- End Violence Against Women: https://www.endviolenceagainstwomen.org.uk/

Other participating organisations:

- Amina - The Muslim Women's Resource Centre https://mwrc.org.uk/
- The Compass Centre https://www.compasscentre.org/
- Edinburgh Rape Crisis Centre https://www.ercc.scot/
- Edinburgh Women's Aid https://edinwomensaid.co.uk/
- Our Streets Now https://www.ourstreetsnow.org/
- Respect. Men's Advice Line https://mensadviceline.org.uk/
- Revenge Porn Helpline https://revengepornhelpline.org.uk/
- Scottish Women's Convention https://www.scottishwomensconvention.org/
- Suzy Lamplugh Trust https://www.suzylamplugh.org/
- Young Scot https://young.scot/