Lost in Deliberation: Making Democracy Understandable

Astier, Cristina*; Khallaf, Nouran; Barriuso, Octavio; Mazzanti, Claudia;
Sayman, Volkan; Bott, Stefan; Sharoff, Serge; Saggion, Horacio

Universitat Pompeu Fabra, Basque Country University, University of Leeds,
Cibervoluntarios, ActionAid, Nexus

Abstract. Drawing on theories of deliberative democracy and Systemic-Functional Linguistics, this study explores how institutional language constitutes a barrier to democracy. Complex texts exclude marginalized groups, including individuals with intellectual and cognitive disabilities (ICD), older persons, and migrants, from participation in deliberative democratic processes. The inclusion of marginalized groups in deliberation poses different challenges to deliberative democratic theory. There is an alleged tension between quality deliberation and autonomy concerns with deliberative innovations for inclusion. We aim at clarifying this tension and argue that the joint efforts of AI and deliberative theory offer a promising avenue for inclusive deliberation. The use of AI to enhance democracy provides tools to overcome barriers for deliberation, making participatory and deliberative processes more inclusive. We argue that AI could contribute to deliberative innovations increasing opportunities for inclusion by providing accessible information. However, these innovations are not value-neutral, the use of Large Language Models (LLMs) for deliberation raises ethical concerns. We focus on algorithmic biases, disinformation, and manipulation threats and claim that our target groups are especially vulnerable to ethical concerns due to their position against a background of structural injustice. Then, based on quality-controlled human annotated datasets, we present a typology of simplification strategies and develop a classifier to detect linguistic complexity and an LLM-based text simplification system which enable a more inclusive participation. We conclude that participatory AI systems must be transparent, interpretable, and co-developed with communities to uphold democratic values and promote social justice.

Keywords: Artificial Intelligence · Deliberative Innovations · Inclusion · Democratic Values · Large Language Models

1 Introduction

This study investigates the contribution of AI in the form of Large Language Models (LLMs) to generate deliberative innovations that enable more inclusive deliberative processes by addressing the linguistic barriers that prevent marginalized communities from engaging in deliberative democracy. It analyses whether

LLMs create the conditions for citizen deliberation and the potential threats that their use could generate. Our work is situated within the broader framework of deliberative democracy, which emphasizes mutual understanding and collective decision making among a diversity of stakeholders [2]. Deliberative democratic theory is well equipped to first identify the barriers and challenges to participation and deliberation that marginalized and traditionally underrepresented people might face in these spaces; second, it provides the theoretical tools to generate innovative strategies, enhancing democracy through an ethical use of AI tools, to both overcome those barriers and improve the quality and inclusiveness of deliberation.

The theory of deliberative democracy claims that public decision making, in order to be politically legitimate, must be the result of an open and ongoing process of public deliberation in which citizens, especially those potentially affected by the decision, must engage on the basis of freedom and political equality [8]. Additionally, the deliberative element of this theory, the principle of argumentation [10], states that decision making processes are organized to maximize the logic of argumentation, i.e., their capacity to promote a consensus-oriented public dialogue based on reasons and arguments, rather than a confrontational logic based on the power of particular-interests-based negotiation or divisive voting. According to this principle, stakeholders participating in the deliberative process must be able to understand the issues at hand, contribute their perspectives, and evaluate arguments in a shared communicative space implementing good deliberative practices. However, in practice, civic language, especially in institutional communications, often relies on highly formalized and complex linguistic structures that limit accessibility. This presents a challenge for groups such as individuals with ICD, older persons or migrants who are not familiar with legal or bureaucratic discourse. These populations are routinely excluded from deliberative participation not by direct or explicit discrimination but by the inaccessibility of the language through which deliberation occurs. Conversely, even when such individuals do contribute their opinions or lived experience to the deliberative process, their voices are more likely to be ignored because their contributions can be misinterpreted in favor of formally educated participants with higher literacy or familiarity with bureaucratic processes. In addition, this practice is one of the reasons why deliberative processes might be seen as elitists, indirectly excluding certain groups of deliberators, thus diminishing the epistemological value of deliberation. The linguistic form of an argument of a stakeholder can influence whether it is recognized as valid, rational, or even question whether it complies with the principle of argumentation.

This challenge is particularly pressing in the context of building institutional trust and democratic legitimacy. As democratic institutions confront growing public skepticism and political polarization, they aim to increase participatory engagement. For example, the European Commission has recently established the C*enter for Participatory and Deliberative Democracy*. Yet, if citizens cannot understand or engage with the content due to linguistic barriers they leave those individuals behind, excluding then decision-making processes. Consequently, these mechanisms risk becoming symbolic rather than substantive, and a lost opportunity to provide substantive equality in the form of deliberating as equals rather than formal equality. Our research draws inspiration from Jürgen Habermas' concept of the "ideal speech situation" [2], the "systemic turn" in deliberative democratic theory [23], normative analyses of potential AI threats to democracy, democratic values, and deliberation [6],[20], as well as from recent advances in Large Language Models [31] which both can help develop deliberative innovations by simplifying complex texts enabling deliberation and formulating arguments from marginalized stakeholders. These concerns raise two overarching research questions for our investigation:

1. How does complex institutional language undermine inclusive participation in democratic deliberation?
2. What role can LLMs play in mitigating these linguistic barriers without compromising the integrity and deliberative quality of democratic discourse?

To address these questions, we adopt a multidisciplinary approach that combines normative analyses of the challenges of inclusion for deliberation, AI-enhanced deliberative innovations and their corresponding ethical concerns, with the technological solution based on Systemic-Functional Linguistics (SFL), LLMs, AI interpretability research, and participatory design methods. All in all, our aim is to use ethical AI to foster the inclusion of people with disabilities and those experiencing language barriers in deliberative processes by enhancing their autonomy via their capacity to participate and deliberate. Our work is framed within the iDEM project (Innovative and Inclusive Spaces for Deliberation and Participation). This European funded project aims at overcoming the linguistic barriers that limit the participation in deliberative spaces of people with limited skills in reading, writing, or understanding complex language required for deliberative and participatory processes. By adopting a user-centred approach and collaborating with organizations which represent people with intellectual disabilities, we ensure maximum impact at promoting a more accessible, inclusive, and thus, egalitarian and unbiased democracy. We develop human language technology by fine-tuning models to automatically detect and classify text complexity, then simplify it accordingly.

2 Artificial Intelligence and Deliberation: Friends or Foes

Deliberative democracy is not without problems. From participatory democracy to deliberative democracy, four of the main challenges faced by deliberators are: first, the asymmetries among their cognitive resources, second, the exclusion of marginalized groups, third, the quality of dialogical social interactions, and fourth, polarized dynamics [19]. These problems are especially salient in the case of marginalized groups and individuals with ICD and limited linguistic skills who experience substantive exclusion. According to iDEM research, these individuals face a total of 14 barriers to democratic participation, which affect how information is understood and arguments discussed. However, at the core of deliberative theory, the principle of inclusion states that the democratic element in ‘deliberative democracy’ requires that all those potentially affected by public decisions have a significant role in deliberative decision making processes, both institutional and non-institutional. In this context, this section has a twofold objective: first, it aims at examining whether there is an internal tension, regarding the inclusion of people with ICD, between quality deliberation and autonomy concerns with deliberative innovations for inclusion; second, it assesses whether AI tools can overcome some of these challenges and improve the inclusion of marginalized groups in democratic deliberation. All in all, this section aims at shedding light about the question of whether AI in the form of LLMs could contribute to the conditions that make deliberation for people with ICD possible.

2.1 Linguistic Barriers in Deliberative Democracy: a Plea for Inclusion

There are two main reasons why including everyone in general and people with ICD in particular in deliberative processes is critical for the legitimacy of democratic systems: intrinsic and instrumental. Intrinsic reasons are based on the democratic principles of political equality and fairness, solidarity, and respect for diversity. A democratic system is politically legitimate insofar as everyone is treated with equal consideration and respect and has an equal say in public decisions [5] [9]. Egalitarian treatment for the case of people with ICD means accommodation. Text simplification during the recruitment, deliberation, and aftermath phases allows people with ICD, who use simpler works and concepts, to understand the topic, express their views, and ultimately participate in the discussion with others. Clearer texts and assistance with expression relevantly contribute to mitigate communication barriers which, as developed in Section 3, are partly supported by structural inequalities. Instrumental reasons are based on collective intelligence theory, conceiving participation of all citizens, including those with ICD as a source of quality public decisions (understood as efficient, effective, and efficacious). According to the epistemic analysis, the inclusion of marginalized people in deliberation is crucial to strengthening the quality of public decisions, which is, in turn, essential for legitimate democratic systems.

Now, although there are intrinsic and instrumental reasons to include people with ICD in deliberative democratic processes, and the role of simplification directly contributes to overcome some of the barriers, making deliberation more egalitarian, certain specific deliberative innovations that foster inclusion have generated an alleged tension regarding the autonomy of deliberators.

In line with the principle of inclusion, institutionalized forms of deliberation are key to democratic legitimacy. In recent years, there has been a *systemic turn* in deliberative democratic theory. It claims that informal public deliberation should be approached as a coherent set of plural deliberations operating at different contexts with different rules and manifestations, aiming at contributing to societal discussion building ample consensus [23]. This systemic approach has opened the door to refinements of the concept of deliberation. This is the case of the concept of "hermeneutical or conceptual exclusion". According to this concept, marginalized groups are doubly excluded from deliberation: first, by inhibiting their ability to "express certain political claims" and, second, by reducing "the likelihood that their political claims will be easily assessable by the public at large." [1]. To overcome this form of exclusion and following the ethos of the *systemic turn*, deliberation that aims at including people with ICD should recognize the *interdependence* of individuals and institutions. However, the concept of *interdependence* between deliberators and institutions conflicts with one source of the value of democratic deliberation: the capacity of individuals to act as autonomous agents, rendering the outcome of public decisions legitimate. Recently, the discussion on the inclusion of individuals with ICD in deliberation has proposed innovative understandings of deliberation and the interaction between deliberators, this is the case of *collaborative speech* [26] [7]. This concept is based on human vulnerability and linguistic acquisition and aims at serving as a solution for people who otherwise are not able to articulate their needs, preferences, and demands. It proposes to adapt deliberative communication methods to ICD deliberators, rather than merely address their perceived deficits, fostering a collaborative environment that embraces uncertainty and messiness in communication. The literature has proposed deliberative innovations to enable the accommodation of participants regardless of their cognitive capacities [1]. In this context, to ensure that information is understood is especially relevant to ensure that *collaborative speech* is possible and empowers deliberators rather than diminishing their autonomy through misinterpretation, paternalists practices, or underestimate their contributions.

Rather than either weaken the deliberative process or undermine their autonomy due to required adaptations, the participation of marginalized and vulnerable people as well as people with ICD, migrants, and the elderly in democracy strengthens the quality of public decision making. The next section continues the exploration of democratic innovations and focuses on the use of AI in the form of LLMs, and the role of simplification in making deliberative processes inclusive. It argues that this technology could play a key role in contributing to secure the value of political equality, enabling deliberative innovations for inclusion in line with *collaborative speech*, and promoting autonomy.

2.2 The Role of AI in Deliberative Democratic Innovations

The revision of deliberative forms of communication for the inclusion of people who cannot express in an argumentative manner, opens a promising avenue for the introduction of AI tools and in particular, LLMs in deliberative processes. In this sense, AI could play a transformative role in democratic processes. In participatory setting where understanding language and texts are key to provide the inclusion of people who experience language barriers, LLMs could simplify information into more easy to understand language [1] overcoming language-related barriers. Thus, AI could contribute to the non-exclusion of uneducated people from deliberative settings that have correspondingly been perceived as elitists, improving the legitimacy of political institutions, decision making processes, and public decisions. Additionally, LLMs could contribute to the efficiency of decision making processes fostering consensus between deliberators [31]. All in all, LLMs have the potential to improve deliberative quality, enabling *collaborative speech*, and a better assistance for the participation of individuals experiencing language barriers. However, at this point someone could ask: do algorithms design the right conditions for citizen deliberation?

In the literature, we can find different explorations of the application of AI to enhance democracy including the use of digital twins (DT) [21], AI for quality mass online deliberation [17], deliberation moderated by AI [31], and the use of simplification to encourage engagement in online deliberation [30].

First, the proposal of DT aims at overcoming the difficulties posed by real-world observations or laboratory experiments to the design of inclusive deliberation processes tailored to a specific set of purposes. DT are proposed as a regulatory sandbox, a dynamic computational modeling framework which serves as a more efficient alternative for designing specific deliberative processes, in particular to test procedural rules for deliberation. However, as the authors state, the efficiency of DT is highly conditioned by " the accuracy of behavioral assumptions, the quality of input data, and their capacity to generalize to real-world democratic practice." [21]. While the inherent complexity of human interactions and social systems poses a substantive barrier for an effective use of DT, it could be a useful technology if developed issue-specific. Second, recently, AI has been considered as a promising tool to enable quality deliberation both in the form of online mass deliberation and randomly selected mini-publics. In the case of mass participation (which is based on the idea of integration and could be summarized with the slogan "all minds in one room" [32]), the inclusion of AI tools could contribute to overcome the classical problem of exchanging reasons, arguments, and justifications, which usually only work within smaller groups, at a population scale. In this scenario, the main role for AI tools in the form of LLMs is to act as moderators, managing the complexity of data and moderating simultaneously different assemblies by providing members with a full range of views. Additionally, further roles for AI in mass deliberation could include translations, fact-checking, data clustering, and aggregation [17]. Specifically the role of moderation and aggregation in deliberation, have been at the center of a recent experiment called the Habermas Machine (HM). This experiment used a fine-tuned version of Google's DeepMind Chinchilla language model to process diverse and opposed views in an online deliberation, a virtual citizens' assembly, and to generate consensus. According to the experiment, AI moderation was more successful at achieving wider consensus than human moderation. A 56/100 of the participants, who were ignorant on whether the moderator was human or not, chose AI instead of human moderation rating them as clearer, more informative, and less biased [31]. The deliberation protocol implied the following steps: first, moderators wrote "group statements" capturing underlying common views, second, participants ranked the initial statement and the top-ranked statement was selected on the basis of aggregation, third, participants privately wrote critiques, ranked the revised statements, and wrote a final preference on the basis of aggregation. AI moderation in these examples relies on preference aggregation to achieve consensus. However, deliberative theory was developed as an alternative or complement to mere preference aggregation procedures such as electoral procedures which are based on majoritarian outcomes. Although both examples are aware of the need to consider minority views, the imperative of inclusion and deliberation in an egalitarian basis, they miss the deliberative argument [17]. Briefly, in group statements and revised group statements we do not know how was the process of someone who changed her mind, what she found as compelling reasons, were she thinks she was, maybe, mistaken, and finally, whether the mind-change was for the right reasons.

Considering the potential adoption of NLP tools in government for civic participation, Guridi et al. [12] investigated their adoption by carrying out interviews with politicians and civil servants. Despite the potential of the technology, since it can reduce work load for civil servants, it is found that its adoption in governments remains limited, with politicians arguing that these tools should guarantee legitimacy before they can be applied.

Regarding the use of simplification to promote engagement in online deliberation, Stodden and Nguyen [30] explore whether making texts simpler can encourage more people to engage in online democratic processes, especially those with literacy challenges. They found that although simplification does not directly influence the intention to use e-participation, it does not have a negative impact while simplified proposals are preferred for in e-participation. Different barriers experimented by individuals who lack linguistic skills including economic barriers, digital literacy, time, and social barriers intersect and are reinforced by structural inequalities. Text simplification contributes to mitigate the exclusion of individuals experimenting power and structural asymmetries by providing unbiased and clear information. Persons with ICD and limited linguistic skills will be benefited by simplification as this tool allows both them and facilitators as well as policymakers to first, translate complex information on the topic to clearer unbiased information and second, better design deliberative processes to include these persons in an equal footing. Clearer information makes clearer arguments and ultimately better discussions. Understanding information and enabling the use of simple concepts and sentences opens the door for the participation of people who are normally excluded, rendering deliberative spaces egalitarian and inclusive, this is to say, truly democratic. All in all, text simplification could be a useful tool to promote inclusion and different ways of communication at deliberative settings, fostering deliberative innovation framed as part of the systemic turn in deliberation and overcoming barriers experienced by individuals with limited communication and linguistic skills. In addition to the opportunities and limitations discussed here, the next section focuses on the ethical concerns that the use of LLMs for deliberation may generate.

3 Ethical Concerns with the use of LLMs in Democratic Deliberation

Opportunities and limitations of the use of AI tools to enhance and overcome feasibility constraints within deliberative practices generate different ethical concerns. This section explores three main ethical concerns that are particularly salient when people with ICD make use of AI tools in deliberative and participatory democratic processes. These concerns are: algoritmic biases, disinformation, and manipulation. Finally, the case at stake requires a brief introduction of a broader concern which sheds light to the general discussion: the social position of marginalized and vulnerable people within institutions. This structural concern is captured by the concept of structural inequality.

According to a possible value-neutral view, biases and discrimination have a functional characteristic, they are used in natural language for inductive purposes. Accordingly, they are used in natural language processing to predict outcomes. LLMs with embedded biases represent a challenge for deliberation due to their anti-deliberative effects. Biases are both embedded in the social world and in LLMs, and, derivatively, there is no zero risk of biases in AI systems. Particularly, given the widespread of political bias in LLMs, Behrendt et al. [3] propose a user-centric evaluation method to measure this perceived bias. They demonstrate that prompting LLMs for neutrality can mitigate some of this perception.

Disinformation is widely recognized as a threat to democracy [20]. It has anti-deliberative effects [20] as it undermines and erodes the epistemic value and potential contributions in deliberative processes [6]. While anti-deliberative communication could be very clear when it appears in the form of insults, fallacies, not letting others talk, or not listening to others, other forms of anti-deliberative practices and communication include disinformation. Disinformation in a deliberation could also appear as unintended and more nuanced. However, its effects on deliberation quality are pervasive. People who have the means to identify false information and anti-deliberative forms of communication are better placed to confront derivative problems including manipulation. Manipulation occurs when individual "A influences B in the direction of believing *p* or doing *d* by influencing B to believe *p* or do *d* by means of a flawed process of reasoning, which A knows works because B is unaware of the flaw in the process or is unable to rectify it. There must be some conflict of purpose here between A and B such that without the flaw, B would go in a different direction that goes against A's purpose." [6] Now, someone could say that manipulation is a highly relevant worry for the case at stake. Given that the most common form of information is still in the form of text, biases embedded in LLMs used to simplified texts could generate sexist, racist, or ableist information, generating anti-deliberative effects and contributing to flawed reasoning, and thus, an unintended form of manipulation. Situations of manipulation at a deliberative setting erode the epistemic potential of contributions, perpetuate inequalities between deliberators, reduces participation, resulting in less political equality [6].

Finally, these concerns merit special analysis when they are raised against a background of structural injustice. Structural injustice refers to situations in which individuals are exposed both to institutional and social practices that create and reproduce social positions that are directly related to advantages and disadvantages within a larger-scale sphere (or structure) of social relations [33]. People with ICD experience disadvantages and barriers to deliberation that are heavily conditioned by their gender, socio-demographic group, class, and type of disability. Structural inequalities can make the case for structural injustices when those characteristics generates situations which are wrong, including direct and indirect discrimination of individuals experimenting these inequalities. This structural element affects institutional procedures of decision making and institutional participation and deliberation. Text simplification mitigating structural inequalities diminishes situations of structural injustice. Persons sharing this context, are directly benefited by the ethical use text simplification, as clearer information improves the quality of the deliberation and sheds light over possible biases that otherwise could remind unnoticed.

4 Language and Language Technology for deliberation

4.1 Language Functions and Social Stratification

From the viewpoint of Systemic-Functional Linguistics, language use is not uniform but varies according to specific instantiation of the general parameters of the context of culture in the current context of situation and the relevant communicative needs, which are, in turn, realized in the form of text [13]. This variation operates on two key dimensions: across language users and across language uses. Variation across language users reflects their sociodemographic identities, including age, education, and dialect. Meanwhile, variation across language use reflects different registers which govern sociolinguistic norms, such as academic writing, political argumentation, or parent-child communication.

Such linguistic variation has implications for social inclusion. Language varieties operate as a gatekeeping mechanism that demarcates in-groups from out-groups. Prestigious language varieties often require adherence to linguistic norms that are inaccessible to individuals lacking formal training. For example, the process of term formation in deliberation discourse transforms *people who represent the mothers in hospitals* into *Maternity services user representatives*, see the discussion of term formation in sciences [14]. This creates compact but opaque terminology that encodes complex meaning without explicit syntactic cues: for example, in the everyday expression *people* is the grammatical subject to *represent* with the direct object of *mothers* (‘who represents the mothers?’), while *hospitals* explicitly encodes the location. On the other hand, the corresponding term merely contains a nested noun phrase structure with unclear relations, which need to be unpacked by the reader:

$((((maternity)(services))(user))(representatives))$

Terminology formation poses challenges for newcomers, who must not only learn domain-specific vocabulary but also internalize the logic and assumptions embedded in such formulations.

In this way, variation makes human communication naturally stratified, so that often the voices of those fluent in dominant language varieties become privileged. This undermines the ideal of an inclusive deliberative space where all citizens can participate equally. Social stratification of this kind raises the demand for *interlingual* translation to enable democratic deliberation across wider sections of society.

4.2 LLMs in Support of Deliberative Democracy

LLMs represent a novel development in our linguistic landscape. They are trained on massive corpora of human-produced text, e.g., the GPT-3 model was trained on approximately 500 billion words, which corresponds to roughly 56 thousand years of human reading [28]. Thus LLMs can learn internal representations of linguistic phenomena across a wide range of registers. This gives them the abilities to explain complex concepts or to translate stories of lived experience into proper arguments exemplified by such stories.

However, LLMs diverge fundamentally from humans in that they lack direct access to the communicative intentions behind texts. While human-produced texts used for training reflect real communicative needs, LLMs can only approximate these needs based on textual patterns. Therefore, the LLMs can be described as “stochastic parrots” [4], which leads to such undesirable properties of their outputs as biases and hallucinations. The LLMs can also threaten democratic processes by offering the possibility of flooding discourse with automatically generated messages [16] and flawed reasoning generating unintended manipulation. Given these capabilities and limitations, the question arises: how can LLMs be harnessed to improve democratic deliberation? The specific context of our project has the aim of developing LLM-based tools for participants with intellectual disabilities, elderly, and migrants. The project involved constructing a typology of linguistic barriers and fine-tuning a general-purpose LLM to suggest simplification strategies for the specific sentences, such as omission, compression, and explanation. Importantly, our approach emphasizes explainability, so users can understand the rationale behind changes made to the text.

5 Removing Linguistic Barriers for Democratic Participation

5.1 Sentence complexity prediction

The English part of the original corpus consists of over 76 parallel texts, primarily sourced from the Scottish care service, political manifestos for the 2024 UK general election, and newsletters from the national charity Disability Equality Scotland. These texts span a diverse range of topics, including health care

In 2018-20	life expectancy at birth in Scotland was	76.8 years for males	and 81.0 years for females.
From 2018 to 2020	babies born in Scotland were expected to live	77 years if they were boys	and 81 years if they were girls.
Modulation	Explanation	Synonymy,Syntax	Synonymy,Syntax

Table 1. Segment alignment for the original (top) and simplified (bottom) sentences. The third row indicate the strategies needed for each segment.

services, environmental policies, the legal system, waste management, disability advocacy, and linguistic accessibility. The French part of the corpus was based on the Réfugiés.info website and covered a range of topics relevant to the refugees.

We developed a range of macro-strategies which can help in explaining why complex sentences in our heterogenous corpus require simplification, see an example in Table 1. The macro-strategies have been thought as points in a continuum between two poles: those resulting in most addition of text (explanation) to those resulting in the most deduction of text (omission), the middle being constituted by transcription, with no addition or deduction of text.

For this experiment, we selected 155 complex sentences from the English source texts and 370 sentences for French. We also annotated them with the macro-strategies required for their simplification. The dataset was divided using an 80/20 split for training and evaluation, respectively. To avoid redundancy and ensure that each sentence pair was annotated with only one simplification category, equivalent complex sentences were manually generated in cases where a specific simplification typology was not present in the original corpus. The eight categories—Omission, Compression, Illocutionary Change, Syntactic Changes, Transcription, Transposition, Synonymy, and Explanation—were selected to create a balanced dataset that captures a diverse range of linguistic transformations and simplification strategies.

The annotation process consisted of a first analysis of the parallel texts, and a review of the existing typologies used to illustrate translation operations, both in the fields of computational linguistics and translation studies.

The training dataset consists of 130 Standard English sentences paired with their simplified counterparts. Each simplified counterpart was designed to include precisely one simplification strategy, where a single complexity was restored to its original form. This design ensures that the relationship between a sentence and its simplified version highlights specific simplification strategies, allowing the model to associate each sentence with different parts of the complexity being resolved. To streamline classification, these fine-grained simplification strategies were mapped to broader macro-categories based on a predefined hierarchical structure, simplifying the labels while preserving their semantic distinctions.

We fine-tuned four different pre-trained transformer models to perform the task of multiclass classification, using BERT and RoBERTa, both mono- and multilingual versions. We employed stratified 5-Fold cross-validation to ensure robust evaluation and generalizability. We used early stopping, where training

Typology Strategy	**Spanish**		**Catalan**	
	# Sent.	%	# Sent.	%
Total Sentences	336	100.00	382	100.00
Complex Sentences	312	92.86	365	95.55
Omission	3	0.89	1	0.26
Compression	5	1.49	3	0.79
Syntactic Changes	61	18.15	104	27.23
Transcript	2	0.60	2	0.52
Transposition	14	4.17	10	2.62
Synonymy	205	61.01	217	56.81
Modulation	14	4.17	16	4.19
Explanation	8	2.38	12	3.14

Table 2. Sentence counts and proportions of simplification strategies in Spanish and Catalan datasets from -4 (Omission) to +4 (Explanation)

was terminated if the validation loss did not improve for the patience period. This ensured efficient use of resources while retaining the best model.

A challenge during training concerned class imbalance in the dataset, where certain strategies were underrepresented. To address this, we replaced the traditional cross-entropy loss with a weighted cross-entropy loss function. Class weights were calculated based on the inverse frequency of each category. This approach ensured that underrepresented classes contributed more significantly to the overall loss, improving the model's ability to predict these minority classes, as otherwise the models tended to predict to the majority classes.

Language	**Precision**	**Recall**	**F1-score**	**Accuracy**	**Predicted**	**Original**
Spanish	1.00	0.93	0.96	0.93	312	336
Catalan	1.00	0.96	0.98	0.96	365	382

Table 3. Performance of the complexity classification model for Spanish and Catalan

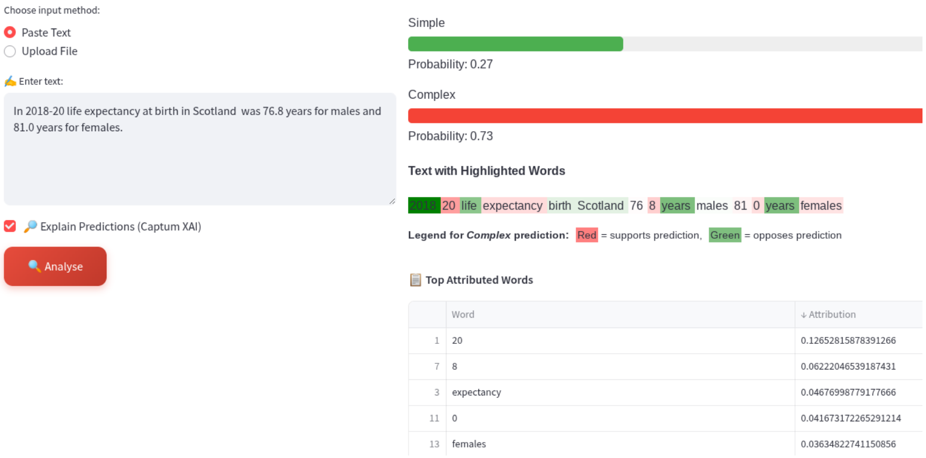
Given the class imbalance, as the overall measure of performance we chose the weighted macro F1-score [29], which better reflects the classifier's ability to handle both frequent and rare simplification strategies. The fine-tuned multilingual classifier model achieved a weighted macro F1-score of 0.8089, demonstrating its ability to generalize across majority and minority classes.

Our pipeline follows a two-step classification approach. First, we apply a complexity classifier that determines whether a sentence requires simplification. Then, for sentences predicted as complex, we apply a second-stage typology classifier that assigns one of eight simplification strategies.

We tested this approach on our developed institutional corpora in Spanish and Catalan. Table 4 summarizes the content distribution of the tested corpora. The Spanish dataset primarily consists of informative texts (59%), along with political commentary and news articles. The Catalan dataset also includes a

Genre	Spanish (%)	Catalan (%)
Informative Texts	59	57
Political & Ideological Articles	18	–
News Articles	18	–
Policy & Legislative Documents	5	21
Social Justice & Public Policy Analysis	–	14

Table 4. Distribution of document genres in the developed Spanish and Catalan corpora.



Choose input method:

● Paste Text
○ Upload File

✍ Enter text:

In 2018-20 life expectancy at birth in Scotland was 76.8 years for males and 81.0 years for females.

☑ 🔎 Explain Predictions (Captum XAI)

Simple

Probability: 0.27

Complex

Probability: 0.73

Text with Highlighted Words

2018 20 life expectancy birth Scotland 76 8 years males 81 0 years females

Legend for *Complex* prediction: Red = supports prediction, Green = opposes prediction

📋 **Top Attributed Words**

	Word	↓ Attribution
1	20	0.12652815878391266
7	8	0.06222046539187431
3	expectancy	0.04676998779177666
11	0	0.041673172265291214
13	females	0.03634822741150856

Fig. 1. Interpretation of predictions via Integrated Gradients. Orange tokens contribute most to the decision of treating this sentence as complex.

strong informative component (57%) but consists of a greater proportion of policy-related documents and public policy analysis.

The performance of the complexity classifiers is shown in Table 3. Results indicate high accuracy and robustness, with F1-scores of 0.96 for Spanish and 0.98 for Catalan, and complex sentences identified in 93% and 96% of the texts, respectively.

For typology classification, the distribution of simplification strategies is reported in Table 2. In both languages, *Synonymy* was by far the most common transformation strategy, followed by *Syntactic Changes*. Interestingly, the least frequent strategies were *Omission* and *Compression*, suggesting that simplification in these corpora tends to rely more on rewording and structural changes than on deletion or pragmatic reformulation.

5.2 Interpretability

Our classifier for predicting the difficulty of sentences offers good performance. However, we were also interested in the reasons why the classifier would predict sentences as being complex. So we carried out experiments of AI explainability with Integrated Gradients (IG) method [15], so that we can detect which words or syntactic constructions commonly affect readability, as well as which of them aligns with human annotation. IG achieves this by calculating the gradients of the model's output with respect to its input, thereby highlighting the importance of individual features. For the sentence in Figure 1, IG offers actionable insights by attributing importance scores to specific words, revealing their influence on the predictions, including the lexical complexity (*expectancy*, which needs to be explained) and complex information packaging (the *2018-20*).

By applying IG across our annotated dataset, we identified a total of 1303 complex words from the original sentences. These words were then compared against their corresponding simplified, easy to read versions to determine which complex words were removed during simplification. This comparison yielded 877 removed words, representing 67.31% of the total complex words identified. The removed words are indicative of tokens that were deemed complex by both the model and human editors, as their removal from the easy to read versions suggests that they were perceived as difficult or unnecessary for simplified comprehension.

5.3 Sentence simplification

Automatic text simplification is a technology to adapt the content of a text by removing the linguistic barriers which are an obstacle to comprehension [24], therefore it can be seen as a way to automatize the translation or transformation of original texts into easy-to-read versions [18]. State-of-the-art simplification systems [27] predominantly use decoder-only auto-regressive LLMs (e.g., GPT-4), which generally outperform other architectures due to their strong few-shot capabilities. However, the use of commercial, closed-source LLMs like GPT-4 poses challenges for the project due to privacy concerns concerns, costs, and the inability to fine-tune them. In our project, we opted to use Salamandra family LLMs [11] which perform exceptionally well on European languages, particularly Romance languages which are the languages of our use cases. Given that Salamandra models are decoder-only and offer instruction-tuned versions, the project's initial strategy revolves around simple few-shot prediction system. Our choice is motivated by recent success in few-shot text simplification by leveraging the model's existing "knowledge" to perform eary-to-read adaptation.

Given that synonymy is the most prominent complexity phenomena in our corpora (see Table 2), we limit our description of simplification to lexical simplification [22], the process of replacing complex words in a given sentence with simpler alternatives of equivalent meaning (i.e., Synonymy). We take advantage of a manually curated corpus of lexical simplification examples in Catalan and Spanish [25] to test our prompting methodology to obtain viable, simpler substitutes for complex words. The dataset is composed of sentences with marked target words and lists of easier synonyms provided by human informants. We test zero and few-shot methods and compare the performance to the baseline used in the recent MLPS 2024 evaluation challenge [27] on simplification. Our prompting strategy conditions or instruct the Salamandra model to provide 10 simpler replacements given a sentence and a target word to simplify. Additionally, for few-shot prompting, the model is conditioned with real examples from the trial part of the dataset (not used in the evaluation). Quantitative assessment is performed using accuracy at 1, which is defined as the percentage of instances where the first top-ranked substitute for the target word matches the most frequently suggested synonym in the gold data. The results so far are promising, with our approach obtaining – few-shot – accuracies of 0.35 for Spanish and 0.22 for Catalan compared to a very strong baseline (0.32 for Spanish and 0.20 for Catalan) based on a 10 times bigger model than ours. Still there is considerable work to be done to improve the final accuracy.

6 Conclusions

This research demonstrates how linguistic complexity in institutional texts significantly limits inclusive participation in democratic deliberation, particularly for marginalized groups such as individuals with ICD, older citizens, and migrants. To overcome these barriers, we propose a solution based on developing a typology of simplification strategies and a classifier capable of detecting sentence complexity across multiple languages, generating a wide range of registers, and mediating across communicative divides. Text simplification contributes to mitigate exclusion overcoming barriers for participation and deliberation as well as situations of structural inequality. Clearer and unbiased information improves deliberation quality and accommodates different deliberators to express their views in an egalitarian footing.

Acknowledgments

This document is part of a project that has received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement No. 101132431 (iDEM Project). The views and opinions expressed in this document are solely those of the authors and do not necessarily reflect the views of the European Union. Neither the European Union nor the granting authority can be held responsible for them. The University of Leeds (UOL) was funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (Grant Agreement No. 10103529). Cristina Astier was partly funded by JDC2023-052963-I, funded by MICIU/AEI/10.13039/501100011033 and by ESF+.

References

1. Afsahi, A.: Disabled lives in deliberative systems. Political Theory **48**(6), 751–776 (2020)
2. Bächtiger, A., Dryzek, J.S., Mansbridge, J., Warren, M.E.: The Oxford handbook of deliberative democracy. Oxford University Press (2018)

3. Behrendt, M., Wagner, S.S., Ziegele, M., Wilms, L., Stoll, A., Heinbach, D., Harmeling, S.: Aqua â€“ combining experts' and non-experts' views to assess deliberation quality in online discussions using llms. In: Proceedings of the First Workshop on Language-driven Deliberation Technology (DELITE) @ LREC-COLING 2024. pp. 1–12. ELRA and ICCL, Torino, Italy (May 2024), https://aclanthology.org/2024.delite-1.1
4. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots: Can language models be too big? In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. pp. 610–623 (2021)
5. Christiano, T.: The constitution of equality: Democratic authority and its limits. Oxford University Press (2008)
6. Christiano, T.: Algorithms, manipulation, and democracy. Canadian Journal of Philosophy **52**(1), 109–124 (2022)
7. Clifford Simplican, S.: Disabling democracy: How disability reconfigures deliberative democratic norms. In: APSA 2009 Toronto Meeting Paper (2009)
8. Cohen, J.: Deliberation and democratic legitimacy. In: Debates in contemporary political philosophy, pp. 352–370. Routledge (2005)
9. Dworkin, R.: What is equality-part 4: Political equality. USFL Rev. **22**, 1 (1987)
10. Elster, J.: Deliberative democracy, vol. 1. Cambridge University Press (1998)
11. Gonzalez-Agirre, A., Pàmies, M., Llop, J., Baucells, I., Dalt, S.D., Tamayo, D., Saiz, J.J., Espuña, F., Prats, J., Aula-Blasco, J., Mina, M., Rubio, A., Shvets, A., Sallés, A., Lacunza, I., Pikabea, I., Palomar, J., Falcão, J., Tormo, L., Vasquez-Reina, L., Marimon, M., Ruíz-Fernández, V., Villegas, M.: Salamandra technical report (2025), https://arxiv.org/abs/2502.08489
12. Guridi, J.A., Cheyre, C., Yang, Q.: Thoughtful adoption of nlp for civic participation: Understanding differences among policymakers. Proc. ACM Hum.-Comput. Interact. **9**(2) (May 2025). https://doi.org/10.1145/3711091, https://doi.org/10.1145/3711091
13. Halliday, M.A.K.: Language as Social Semiotic: The social interpretation of language and meaning. Blackwells, Oxford (1978)
14. Halliday, M.A.K., Matthiessen, C.M.I.M.: Construing experience through meaning: a language-based approach to cognition. Cassell, London (1999)
15. Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., et al.: Captum: A unified and generic model interpretability library for pytorch. arXiv preprint arXiv:2009.07896 (2020)
16. Kreps, S., Kriner, D.: How AI threatens democracy. Journal of Democracy **34**(4), 122–131 (2023)
17. Landemore, H.: Can ai bring deliberative democracy to the masses. In: Human Centered Artificial Intelligence Seminar, Stanford University (2022)
18. Maaß, C.: Easy Language - Plain Language - Easy Language Plus. Balancing Comprehensibility and Acceptability. Frank & Timme (01 2020). https://doi.org/10.25528/042
19. Mansbridge, J.J.: Beyond adversary democracy. University of Chicago Press (1983)
20. McKay, S., Tenove, C.: Disinformation as a threat to deliberative democracy. Political research quarterly **74**(3), 703–717 (2021)
21. Novelli, C., Sánchez-Vaquerizo, J.A., Helbing, D., Rotolo, A., Floridi, L.: A replica for our democracies? on using digital twins to enhance deliberative democracy. arXiv preprint arXiv:2504.07138 (2025)
22. Paetzold, G.H., Specia, L.: A survey on lexical simplification. J. Artif. Int. Res. **60**(1), 549–593 (Sep 2017)

23. Parkinson, J., Mansbridge, J.: Deliberative systems: Deliberative democracy at the large scale. Cambridge University Press (2012)
24. Saggion, H.: Automatic Text Simplification, Synthesis Lectures on Human Language Technologies, vol. 10. Morgan & Claypool Publishers (2017)
25. Saggion, H., Bott, S., Szasz, S., Pérez, N., Calderón, S., Solís, M.: Lexical complexity prediction and lexical simplification for Catalan and Spanish: Resource creation, quality assessment, and ethical considerations. In: Shardlow, M., Saggion, H., Alva-Manchego, F., Zampieri, M., North, K., Štajner, S., Stodden, R. (eds.) Proceedings of the Third Workshop on Text Simplification, Accessibility and Readability (TSAR 2024). pp. 82–94. Association for Computational Linguistics, Miami, Florida, USA (Nov 2024). https://doi.org/10.18653/v1/2024.tsar-1.9, https://aclanthology.org/2024.tsar-1.9/
26. Schramme, T.: Capable deliberators: towards inclusion of minority minds in discourse practices. Critical Review of International Social and Political Philosophy **27**(5), 835–858 (2024)
27. Shardlow, M., Alva-Manchego, F., Batista-Navarro, R., Bott, S., Calderon Ramirez, S., Cardon, R., François, T., Hayakawa, A., Horbach, A., Hülsing, A., Ide, Y., Imperial, J.M., Nohejl, A., North, K., Occhipinti, L., Rojas, N.P., Raihan, N., Ranasinghe, T., Salazar, M.S., Štajner, S., Zampieri, M., Saggion, H.: The BEA 2024 shared task on the multilingual lexical simplification pipeline. In: Kochmar, E., Bexte, M., Burstein, J., Horbach, A., Laarmann-Quante, R., Tack, A., Yaneva, V., Yuan, Z. (eds.) Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024). pp. 571–589. Association for Computational Linguistics, Mexico City, Mexico (Jun 2024), https://aclanthology.org/2024.bea-1.51/
28. Sharoff, S.: Form and function: automatic methods for prediction of functions. In: Wegener, R., McCabe, A., Sellami-Baklouti, A., Fontaine, L. (eds.) Transdisciplinary Systemic Functional Linguistics. Routledge (2025)
29. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Information Processing & Management **45**(4), 427–437 (2009). https://doi.org/10.1016/j.ipm.2009.03.002
30. Stodden, R., Nguyen, P.: Can text simplification help to increase the acceptance of e-participation? In: Proceedings of the First Workshop on Language-driven Deliberation Technology (DELITE) @ LREC-COLING 2024. pp. 20–32. ELRA and ICCL, Torino, Italy (May 2024), https://aclanthology.org/2024.delite-1.3
31. Tessler, M.H., Bakker, M.A., Jarrett, D., Sheahan, H., Chadwick, M.J., Koster, R., Evans, G., Campbell-Gillingham, L., Collins, T., Parkes, D.C., et al.: AI can help humans find common ground in democratic deliberation. Science **386**(6719) (2024)
32. Velikanov, C., Prosser, A.: Mass online deliberation in participatory policy-making. Beyond Bureaucracy. Towards Sustainable Government Informatisation pp. 209–258 (2017)
33. Young, I.M., Nussbaum, M.: Responsibility for Justice. Oxford University Press (2011), https://doi.org/10.1093/acprof:oso/9780195392388.001.0001