INTERNATIONAL JOURNAL OF ADVANCED ARTIFICIAL INTELLIGENCE RESEARCH

DEVELOPING AI ASSISTANCE FOR INCLUSIVE COMMUNICATION IN ITALIAN FORMAL WRITING

Dr. Alessia Romano
Department of Linguistics and Artificial Intelligence, University of Bologna, Italy

Prof. Marco Bianchi
Department of Computer Science, Sapienza University of Rome, Italy

Published Date: 13 December 2024 // Page no.:- 15-19

ABSTRACT

The push for inclusive language has gained global momentum, recognizing its crucial role in fostering respectful and equitable communication. While English has seen significant advancements in AI-assisted inclusive writing, highly gendered languages like Italian present unique challenges due to their complex grammatical structures, where gender is intrinsically linked to nouns, articles, adjectives, and verb participles. This article proposes a novel approach for developing an AI-powered system designed to assist in writing inclusive language within Italian formal communications. Our methodology integrates sophisticated Natural Language Processing (NLP) techniques, leveraging advanced Italian Language Models fine-tuned on expert-annotated datasets of formal texts. We outline the process of identifying non-inclusive patterns, generating grammatically correct and contextually appropriate inclusive alternatives, and the critical role of Explainable AI (XAI) to ensure user trust and understanding. Anticipated outcomes suggest that this system will significantly enhance the efficiency and accuracy of inclusive writing in Italian formal settings, promoting broader adoption of gender-neutral and person-first language principles.

Keywords: Inclusive language, Italian language, AI assistance, Formal communication, Natural Language Processing, Gender neutrality, Explainable AI.

INTRODUCTION

The adoption of inclusive language has become a cornerstone of modern communication, reflecting a societal commitment to diversity, equity, and respect across various communities [3, 6, 19, 29]. Inclusive language aims to avoid bias, stereotypes, and discriminatory expressions that could marginalize or exclude individuals or groups based on gender, ethnicity, disability, age, or other characteristics [51]. While guidelines and tools for inclusive writing are increasingly prevalent in languages like English [3, 6, 19, 20, 29], their development for highly inflected and grammatically gendered languages such as Italian presents unique and complex challenges.

Italian, a Romance language, assigns grammatical gender (masculine or feminine) to virtually all nouns, and this gender agreement extends to articles, adjectives, pronouns, and past participles, pervading sentence structure. This intrinsic grammatical gendering complicates the direct application of strategies common in less inflected languages, such as simply substituting gendered pronouns with "they/them" [48]. For instance, a job title like "ministro" (minister) traditionally masculine, has become a point of linguistic and social debate when referring to a woman, leading to discussions on forms like "ministra" [8]. Italian academic institutions and governmental bodies are actively developing guidelines for gender-inclusive communication [10, 11, 12, 13, 15, 16, 17, 31, 32], underscoring the pressing need for effective tools.

The rise of Artificial Intelligence (AI) and Natural Language Processing (NLP) offers a promising avenue to address these complexities [4, 24]. AI models, particularly large language models (LLMs), have demonstrated remarkable capabilities in understanding, generating, and transforming text [7, 36, 37]. However, applying these general capabilities to the nuanced and culturally specific requirements of inclusive language in Italian formal communications demands specialized development. Existing research has explored user-centric gender rewriting [1], de-biasing techniques [2], and gender-neutral translation [33, 44, 52, 54] in various languages, yet a tailored, robust AI-driven solution specifically for Italian formal contexts is still an emerging field.

This article proposes a comprehensive methodology for developing an AI-assisted system aimed at facilitating inclusive language writing in Italian formal communications. We hypothesize that by combining advanced NLP techniques with meticulously curated, expert-annotated Italian textual data and incorporating Explainable AI (XAI) principles, we can create a tool that not only identifies non-inclusive patterns but also suggests accurate, grammatically sound, and contextually appropriate inclusive alternatives. The subsequent

sections detail the proposed methods, outline anticipated results, and discuss the broader implications and future directions of this research.

2. Methods

The development of an AI-assisted inclusive language writing tool for Italian formal communications involves several interconnected stages: corpus creation and annotation, guideline implementation, model architecture design, and evaluation.

2.1 Corpus Creation and Expert Annotation

A foundational step is the creation of a specialized corpus of Italian formal communications (e.g., administrative documents, official reports, academic papers, policy texts). This corpus must then undergo meticulous expert annotation to identify instances of non-inclusive language and provide corresponding inclusive alternatives [35]. Given the linguistic complexities and the evolving nature of inclusive language norms in Italian, human expertise is indispensable for generating high-quality labeled data. Annotation platforms like Label Studio [50] can facilitate this process, ensuring consistency and manageability of the dataset. The annotation process will strictly adhere to established guidelines from authoritative Italian academic and European institutions [10, 11, 12, 13, 15, 16, 17, 31, 32].

2.2 Inclusive Language Guidelines and Pattern Identification

The system's core logic will be built upon a detailed set of Italian inclusive language guidelines. These guidelines dictate transformations for various linguistic phenomena, including:

1. Gendered Nouns and Adjectives: Identifying common collective nouns (e.g., "i professori" - male professors or mixed group of professors) and suggesting gender-neutral alternatives (e.g., "i docenti" - the faculty) or dual forms (e.g., "studenti e studentesse" - male and female students).
2. Pronouns and Articles: Addressing the challenges of generic masculine pronouns and articles and suggesting appropriate rewrites [48].
3. Participles and Agreement: Handling the agreement of past participles with subjects of mixed gender or indeterminate gender.
4. Addressing and Salutations: Providing alternatives for gender-specific forms of address in formal letters.
5. Person-First Language: While a broader aspect, the system will also incorporate elements of person-first language, particularly relevant for discussions around disability [51].

The system will leverage NLP techniques to identify patterns indicative of non-inclusive language. This involves syntactic parsing to understand grammatical roles and semantic analysis to interpret context [49].

2.3 AI Model Architecture for Rewriting

The core of the AI assistance will be a sophisticated language model capable of sequence-to-sequence rewriting. Given the success of transformer-based architectures in various NLP tasks, we propose exploring fine-tuning large pre-trained Italian language models.

1. Pre-trained Italian Language Models: Models such as BERT-based (e.g., GePpeTto [28], ALBERTO [34], Italian BERT/ELECTRA models [45]) or T5-based (e.g., IT5 [40], BART-IT [23]) are suitable candidates. These models have been pre-trained on vast Italian text corpora, giving them a strong foundation in Italian grammar and semantics [14, 53].

2. Fine-tuning for Rewriting: The selected pre-trained model will be fine-tuned on our expert-annotated dataset for the specific task of inclusive language rewriting [7, 36]. This involves mapping non-inclusive input sentences to their corresponding inclusive output sentences. The model will learn to identify context, grammatical dependencies, and apply the rules derived from the inclusive language guidelines.

3. Contextual Understanding: Crucially, the model must be able to understand the context of the sentence to provide appropriate suggestions. For example, rewriting "the manager" (if gendered in Italian) requires knowing if the manager's gender is known or unknown, or if "manager" refers to an individual or a collective. Approaches to user-centric gender rewriting [1] and de-biasing models [2] can inform this contextual sensitivity.

4. Handling Ambiguity: Italian gendered terms can be ambiguous (e.g., "i parenti" can mean male relatives or relatives of mixed gender). The model should be designed to offer multiple inclusive options or prompt the user for clarification in such cases.

2.4 Explainable AI (XAI) Integration

To build user trust and enable effective human-AI collaboration, the system will incorporate Explainable AI (XAI) components [9, 21, 27]. For each suggested rewrite, the system should explain why a particular change was recommended and how the inclusive alternative was derived. This could involve highlighting the specific non-inclusive words or phrases identified and referencing the relevant inclusive language guideline that prompted the change [39, 55, 56]. XAI is particularly important in a domain as sensitive and nuanced as language transformation.

2.5 Evaluation Metrics

The performance of the AI-assisted system will be evaluated using both automated and human-centric metrics:

1. Automated Metrics: Standard NLP metrics like BLEU [30] and ROUGE [26] can assess the fluency and grammatical correctness of the generated inclusive text against reference translations. However, these metrics might not fully capture inclusivity.

2. Inclusivity Metrics: Custom metrics or adaptations of existing bias detection metrics [41, 42, 46, 47] will be developed to quantify the degree of inclusivity achieved in the rewritten text.

3. Human Evaluation: Expert human evaluators will assess the semantic accuracy, grammatical correctness, contextual appropriateness, and overall naturalness of the suggested inclusive rewrites. Inter-annotator agreement (e.g., using Cohen's Kappa [25]) will be crucial for validating human judgments. User studies will also gauge user satisfaction, perceived helpfulness, and ease of use.

3. Results (Anticipated)

The development and deployment of an AI-assisted inclusive language writing tool for Italian formal communications, following the proposed methodology, is anticipated to yield significant positive results across several dimensions:

- Improved Accuracy in Inclusivity: The system is expected to demonstrate high accuracy in identifying non-inclusive language patterns specific to Italian grammar and context. Leveraging large, fine-tuned Italian language models will enable precise recognition of subtle gendered forms, even in complex sentence structures.

- Enhanced Grammatical Correctness and Fluency: By training on high-quality, expert-annotated formal Italian texts, the AI model will be capable of generating grammatically correct and fluent inclusive alternatives that seamlessly integrate into the original text. This will address a major challenge in manual rewriting, which often leads to awkward phrasing or grammatical errors.

- Significant Reduction in Manual Effort: For authors, editors, and communicators working with Italian formal documents, the tool will drastically reduce the time and effort currently spent on manually reviewing and rewriting texts for inclusivity. This automation will streamline communication workflows within institutions and organizations.

- Greater Consistency in Inclusive Language Adoption: The system will promote a more consistent application of inclusive language guidelines across different authors and departments within an organization. This consistency is crucial for establishing and maintaining a unified inclusive communication standard.

- Increased User Trust and Acceptance: The integration of Explainable AI (XAI) will be a critical factor in user acceptance. By providing clear justifications for suggested changes, the system will empower users to understand the rationale behind the recommendations, thereby fostering trust and encouraging broader adoption of inclusive practices.

- Positive Social and Cultural Impact: By making inclusive writing more accessible and manageable, the tool will contribute to the wider adoption of gender-neutral and person-first language in Italian formal discourse. This will promote a more equitable and respectful communication environment, aligning with the values of diversity and inclusion.

- Foundation for Future Linguistic Research: The creation of a large, expertly annotated Italian inclusive language corpus will serve as a valuable resource for future research in NLP, computational linguistics, and sociolinguistics, particularly concerning gender bias in language and its mitigation.

These anticipated outcomes collectively underscore the potential of this AI-assisted solution to bridge the gap between linguistic complexity and the growing demand for inclusive communication in Italian formal settings.

4. Discussion

The proposed AI-assisted system for inclusive language writing in Italian formal communications represents a crucial step towards addressing a significant linguistic and societal need. The inherent gendered nature of Italian grammar poses unique challenges that simple find-and-replace mechanisms cannot adequately resolve. Our approach, by focusing on a comprehensive methodology that encompasses expert-annotated data, sophisticated language models, and Explainable AI, aims to provide a robust and practical solution.

The emphasis on expert-annotated data [35] is paramount. General-purpose language models, despite their vast training corpora, often reflect biases present in the data they were trained on [41, 42, 46, 47]. For a sensitive domain like inclusive language, particularly in formal contexts where precision and nuance are critical, a dedicated, high-quality dataset is essential to teach the AI the subtle rules and appropriate contextual transformations. This aligns with the understanding that trustworthy AI systems require careful attention to data quality and bias mitigation [27].

Leveraging pre-trained Italian language models (e.g., BERT, T5 families) and fine-tuning them for the specific rewriting task is a computationally efficient and effective strategy. These models possess a deep understanding of Italian syntax, semantics, and morphology [23, 28, 34, 40, 45, 53], which is indispensable for generating grammatically correct and natural-sounding inclusive alternatives. The challenge lies in guiding these models to consistently produce outputs that adhere to established inclusive language guidelines without sacrificing fluency

or meaning. This is where approaches focused on user-centric rewriting [1], de-biasing [2], and gender-neutral translation [33, 44, 52, 54] become highly relevant, providing frameworks for guiding the model's output towards desired inclusive patterns.

The integration of Explainable AI (XAI) is not merely a technical add-on but a fundamental component for usability and trust [9, 21]. In the context of language, where personal preferences and cultural nuances play a significant role, blindly accepting AI suggestions can be problematic. By providing explanations for its recommendations [39, 55, 56], the system empowers users to understand the reasoning, learn the guidelines, and make informed decisions, fostering a collaborative rather than dictatorial human-AI interaction [22]. This transparency is especially vital in formal communications where accuracy and adherence to specific institutional guidelines are critical.

While the anticipated results are promising, several challenges and future directions exist. The evolution of inclusive language itself is ongoing, requiring continuous updates to guidelines and, consequently, to the AI model and its training data. Cultural acceptance and the pace of linguistic change will also influence adoption. Furthermore, extending inclusivity beyond gender to encompass other dimensions such as disability [51], ethnicity, and socioeconomic status would add further complexity but enhance the system's overall impact.

Future work could include:

- Real-time Integration: Developing plug-ins or APIs to seamlessly integrate the AI assistant into popular word processors and communication platforms, offering suggestions as users type.
- User Customization: Allowing users or institutions to customize the inclusive language guidelines, accommodating specific organizational policies or regional linguistic preferences.
- Multimodality: Exploring how AI can assist in inclusive communication across different modalities, such as in presentation slides or visual content, where language choices still play a role.
- Continuous Learning: Implementing mechanisms for the model to continuously learn from user feedback and newly published inclusive language guidelines, ensuring its relevance and accuracy over time.

CONCLUSION

In conclusion, by addressing the unique linguistic challenges of Italian through a robust AI-driven approach, this research aims to significantly advance the practice of inclusive communication in formal settings. Such a tool not only offers practical benefits in efficiency and accuracy but also plays a vital role in promoting a more equitable, respectful, and representative linguistic landscape.

REFERENCES

[1] Alhafni, B., Habash, N., & Bouamor, H. (2022). User-centric gender rewriting. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics:[1] Human Language Technologies (pp. 618–631). Association for Computational Linguistics.

[2] Amrhein, C., Schottmann, F., Sennrich, R., & Läubli, S. (2023). Exploiting biased models to de-bias text: A gender-fair rewriting model. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics[2] (Vol. 1: Long Papers) (pp. 4486–4506). Association for Computational Linguistics.

[3] American Psychological Association. (2023). Inclusive Language Guidelines. Retrieved August 2023 from https://www.apa.org/about/apa/equity-diversity-inclusion/language-guidelines

[4] Attanasio, G., Greco, S., La Quatra, M., Cagliero, L., Tonti, M., Cerquitelli, T., & Raus, R. (2021). E-MIMIC: Empowering multilingual inclusive communication. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data) (pp. 4227–4234). IEEE.

[5] Attanasio, G., Pastor, E., Di Bonaventura, C., & Nozza, D. (2023). Ferret: A framework for benchmarking explainers on transformers. In Proceedings of the 17th Conference of the European Chapter of the Association for Computational[3] Linguistics: System Demonstrations. Association for Computational Linguistics.[4]

[6] Babinszki, T., Cavender, A., Gower, M., Hoehl, J., Lima, D., Manser, E., & Trewin, S. (2019). Inclusive writing. In Y. Yesilada & S. Harper (Eds.), Web Accessibility—A Foundation for Research (2nd ed., pp. 135–152). Springer.

[7] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language[5] models are few-shot learners. In Proceedings of the Advances in Neural Information Processing[6] Systems, Vol. 33 (pp. 1877–1901).

[8] Cignarella, A. T., Lai, M., Andrea, M., & Sanguinetti, M. (2021). La ministro è incinta: A twitter account of women's job titles in Italian. In Proceedings of the 8th Italian Conference on Computational Linguistics (CLiC-it '21), Vol. 3033. CEUR-WS.org (pp. 1–7).

[9] Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B., & Sen, P. (2020). A survey of the state of explainable AI for natural language[7] processing. In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing (pp. 447–459). Association for Computational Linguistics. Retrieved from https://aclanthology.org/2020.aacl-main.46

[10] Università degli studi dell'Aquila. (2023). Per un uso della lingua Italiana rispettoso dei generi. Retrieved August 2023 from https://www.univaq.it/include/utilities/blob.php?item=file&table=allegato&id=4925

[11] Università degli studi di Bologna. (2023). Linee guida per la visibilità del genere nella comunicazione istituzionale dell'università di Bologna. Retrieved August 2023 from https://www.unibo.it/it/allegati/linee-guida-per-la-visibilita-

del-genere-nella-comunicazione-istituzionale-dell2019universita-di-bologna/@@download/file/Linee-Guida-Genere-2020.pdf
[12] Università degli studi di Milano. (2023). Vademecum sul linguaggio di genere. Retrieved August 2023 from https://www.unimi.it/sites/default/files/2021-12/Vademecumlinguaggio%20di%20genere_Universit%C3%A0%20degli%20Studi%20di%20Milano.pdf
[13] Università degli studi di Torino. (2023). Un approccio di genere al linguaggio amministrativo. Retrieved August 2023 from https://www.unito.it/sites/default/files/linee_guida_approccio_genere.pdf
[14] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1: Long and Short Papers) (pp. 4171–4186). Association[8] for Computational Linguistics.
[15] Università Mediterranea di Reggio Calabria. (2023). Indicazioni per un uso del linguaggio rispettoso delle differenze. Retrieved August 2023 from https://www.unirc.it/documentazione/media/files/ateneo/pari_opportunita/File_allegato_2.pdf
[16] Università di Siena. (2023). Generi e linguaggi: Linee Guida per un Linguaggio administrative e istituzionale inclusivo. Retrieved August 2023 from https://www.unisi.it/comunicazione/linee-guida-un-linguaggio-amministrativo-e-istituzionale
[17] Politecnico di Torino. (2023). Guida pratica per una comunicazione inclusiva. Retrieved August 2023 from https://www.polito.it/sites/default/files/2023-07/vademecum%20esteso_170723.pdf
[18] Diesner-Mayer, T., & Seidel, N. (2022). Supporting gender-neutral writing in German. In Proceedings of the Mensch Und Computer 2022. ACM.
[19] UK Government. (2014). Inclusive Communication. Retrieved August 2023 from https://www.gov.uk/government/publications/inclusive-communication
[20] UK Government. (2020). Breaking Down Gender Stereotypes in Legal Writing. Retrieved August 2023 from https://civilservice.blog.gov.uk/2020/01/10/breaking-down-gender-stereotypes-in-legal-writing/
[21] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box[9] models. ACM Computing Surveys, 51(5), Article 93, 42 pages.
[22] Kolin, P. C. (2007). Successful Writing at Work. Houghton Mifflin.
[23] La Quatra, M., & Cagliero, L. (2023). BART-IT: An efficient sequence-to-sequence model for Italian text summarization. Future Internet, 15(1), 15.
[24] La Quatra, M., Greco, S., Cagliero, L., & Cerquitelli, T. (2023). Inclusively: An AI-based assistant for inclusive writing. In G. De Francisci Morales, C. Perlich, N. Ruchansky, N. Kourtellis, E. Baralis, & F. Bonchi (Eds.), Machine Learning and Knowledge Discovery in Databases: Applied Data Science and Demo Track (pp. 361–365). Springer Nature, Switzerland, Cham.
[25] Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.