A Computational Framework for Measuring and Analyzing Gender Bias in Portuguese-language Literary Texts

Mariana O. Silva[1], **Michele A. Brandão**[1], **Mirella M. Moro**[1]

[1]Department of Computer Science
Universidade Federal de Minas Gerais (UFMG) – Brazil

{mariana.santos,michele.brandao,mirella}@dcc.ufmg.br

Abstract. *Literary texts often reproduce and reshape social biases, including gender bias, through recurring linguistic patterns that influence how characters are represented. While traditional literary criticism offers detailed interpretations, it is limited in scale, and most computational studies of gender bias focus on English, leaving Portuguese-language literature underexplored. This dissertation addresses this gap by proposing* PORTALIA*, a modular computational framework for measuring and analyzing gender bias in Portuguese-language literary texts. The framework integrates Natural Language Processing techniques for character identification, automated gender inference, linguistic analysis, and quantitative bias measurement. It is evaluated on Portuguese literary corpora and supports large-scale, reproducible analyses of gendered patterns in character representation, descriptors, and narrative agency.*

Category	Doutorado
Author	Mariana de Oliveira Santos Silva
Graduate Program	Programa de Pós-Graduação em Ciência da Computação
University	Universidade Federal de Minas Gerais (UFMG)
Advisor	Profa. Dra. Mirella Moura Moro
Co-advisor	Profa. Dra. Michele Amaral Brandão
Defense and Approval Date	December 11, 2025
Grade/Concept	Approved
Honors or Distinctions	Not applicable

Examination Committee

Member	Affiliation
Profa. Dra. Gisele Lobo Pappa	Departamento de Ciência da Computação – UFMG
Profa. Dra. Karin Becker	Instituto de Informática – UFRGS
Profa. Dra. Ticiana L. Coelho da Silva	PPG em Modelagem e Métodos Quantitativos – UFC
Prof. Dr. Evandro L. T. P. Cunha	Faculdade de Letras – UFMG

Highlights

- Reproducible framework for gender bias analysis in Portuguese-language literature.
- Literary NER models adapted to Portuguese-language literary texts.
- Annotated corpus for entity recognition in Portuguese-language literary works.
- Gendered patterns are measured in texts and computational methods.
- Linguistic metrics reveal gender asymmetries in character portrayal.

Code and data: https://github.com/marianaossilva/gender_pipeline

1. Context and Problem

Literary texts encode social structures through linguistic and narrative patterns, including gender bias, which affects how characters are described, act, and are positioned in narratives [Luo et al. 2024, Stuhler 2024]. While traditional literary criticism provides detailed analyses via *close reading*, it is limited in scale and reproducibility. In contrast, computational approaches enable *distant reading*, allowing literary corpora to be treated as structured textual data that can be processed, analyzed, and compared at scale [Schulz and Štěpán Bahník 2019, Hoyle et al. 2019]. This is also relevant for Computer Science, as literary data are part of the corpora used to train language models, which may inherit and amplify such biases [Caton and Haas 2024].

Despite advances in computational bias analysis, most research focuses on English and high-resource languages [Lima and Araujo 2023]. Portuguese remains underrepresented in datasets, tools, and large-scale studies, especially for literary texts [Freitas and Martins 2023, Freitas and Santos 2023]. Additionally, bias is not directly observable, emerging from multiple signals such as entity recognition, gender attribution, and linguistic patterns. Thus, the problem addressed is *how to measure and analyze gender bias in Portuguese-language literary texts in a systematic and scalable way*. To do so, this Ph.D. dissertation proposes *PORTALIA*, a computational framework integrating character identification, gender inference, and bias measurement.

2. Research Goals and Contributions

The main goal of this dissertation is to *develop a computational framework for measuring and analyzing gender bias in Portuguese-language literary texts*. This goal is grounded on a systematic mapping of related work, which reveals four key challenges that guided the proposed solution. These challenges are addressed through four research goals (**RGs**), each corresponding to a core component of the framework.

[RG1] Develop a computational framework for gender bias analysis. This goal focuses on designing a modular framework that integrates linguistic and computational methods, structured around three components: *Character Identification*, *Gender Inference*, and *Gender Bias Measuring*.

[RG2] Develop domain-adapted methods for literary character identification. This goal addresses the recognition of characters in Portuguese literary texts through annotated corpora and domain-adapted Named Entity Recognition (NER) models, improving robustness in the presence of stylistic variation.

[RG3] Develop an automated approach for gender inference. This goal proposes a hybrid method that combines name-based information with contextual linguistic cues, enabling scalable and accurate gender attribution in literary texts.

[RG4] Measure and analyze gender bias in texts and models. This goal investigates gender representation in literary texts and its reflection in computational methods, including embeddings and generative models, through quantitative and linguistic analyses of character portrayal.

3. State-of-the-art Advancement

This work advances the state of the art by addressing three key limitations of prior research. First, existing approaches often analyze bias in isolation (e.g., word associations

or model outputs) without integrating entity extraction, gender attribution, and linguistic analysis. Second, most studies focus on English, limiting applicability to other languages. Third, many analyses rely on black-box models with limited interpretability. *PORTALIA* addresses these issues by providing a unified and modular framework that integrates multiple stages of analysis and enables interpretable, dependency-based bias metrics. It also introduces resources and methods tailored to Portuguese, enabling more accurate and context-aware analysis. Finally, it bridges bias in language and bias in computational methods, supporting more comprehensive and auditable analyses.

4. Summary of the Proposed Solution

PORTALIA is a modular, computational framework for measuring and analyzing gender bias in Portuguese-language literary texts. It integrates insights from literary studies and NLP into a unified, reproducible workflow that transforms unstructured narratives into analyzable data. The framework is organized into three interconnected components (character identification, gender inference, and bias measurement) explicitly modeling their dependencies through well-defined inputs and outputs.

Character Identification. The first component of the framework focuses on identifying literary characters. Literary texts pose specific challenges for NER, including stylistic variation, historical language use, and indirect or descriptive references to characters. To address these challenges, *PORTALIA* adopts domain-adaptive NER strategies tailored to Portuguese-language literary corpora. This module relies on an annotated corpora and fine-tuned transformer models tailored to Portuguese literary data, improving robustness in capturing narrative entities.

Gender Inference. This second component assigns gender labels through a hybrid strategy that combines name-based evidence with contextual linguistic cues, such as pronouns and agreement patterns. This approach improves robustness over single-source methods and supports scalable gender attribution while preserving interpretability.

Gender Bias Measurement. The final component measures bias at two levels. At the linguistic level, it captures asymmetries in descriptors, actions, and narrative roles using dependency-based analysis. At the computational level, it evaluates how these patterns are encoded in models, including embeddings and generated texts. This dual perspective enables a comprehensive and auditable analysis of gender bias as both a textual and computational phenomenon.

5. Evaluation

The framework was evaluated through experiments aligned with its four components. **Character Identification.** Literary NER was assessed using the *SemEval 2013 Task 9.1* scheme, comparing general-domain and domain-adapted models. Fine-tuned models consistently outperform the BERT-CRF baseline, improving recall and reducing missing entities. Under the **Strict** setting, F1 scores reach approximately 77%, exceeding 81% under relaxed criteria. Performance is strongest for the `PERSON` category, which is the main goal of the *Character Identification* component.

Gender Inference. The gender inference component was evaluated on a gold-standard dataset of 1,000 `PERSON` entities labeled as *female*, *male*, or *unknown*, with substantial

inter-annotator agreement ($\kappa \approx 0.65$). Name-based strategy achieves high precision but limited coverage, while context-based cues significantly expand coverage. Combined, the hybrid method achieves 76.2% accuracy and 93.0% coverage, reaching 98.8% accuracy in binary classification (excluding *unknown*).

Gender Bias in Language. Applying the framework to literary corpora in Portuguese (1,297 public-domain works) reveals consistent gender asymmetries. Male entities dominate (often over 60%) and are more frequently associated with action verbs and agentive roles, while female entities appear more often in descriptive or relational contexts. Lexical analysis shows systematic differences: female entities are linked to appearance and emotional attributes, whereas male entities are associated with authority, profession, and action.

Gender Bias in Computational Models. Analyses of embeddings and generated texts show that these patterns propagate to computational models. Generated narratives reproduce male over-representation (60–68%) and exhibit gender confusion, particularly misclassifying female entities. WEAT [Caliskan et al. 2017] analyses indicate stronger associations between male-related terms and domains such as *Work* and *Risk*, while female-related terms align with *Home* and *Feel*. In several cases, models amplify biases present in the source corpus.

6. Scientific Production

This Ph.D. dissertation resulted in a consistent body of scientific, technical, and open research contributions, including peer-reviewed publications, datasets, and code resources.

Scientific Publications. The work generated 8 direct publications closely aligned with the thesis contributions, including: (i) 3 journal papers (*Information Processing & Management* [Silva et al. 2024b], *LINGUAMÁTICA* [O. Silva and Moro 2025], and *Journal of the Brazilian Computer Society* [Silva et al. 2025d]), (ii) 2 international conferences (PROPOR [Silva and Moro 2024b], LREC-COLING [Silva and Moro 2024c]), and (iii) 3 national venues (KDMiLe [Silva et al. 2023a], SEMISH [Silva and Moro 2024a], STIL [Silva et al. 2025a]). Among these, the work received a **Second Best Paper Award** [Silva et al. 2023a]. A systematic literature review derived from this research is currently under review at *ACM Computing Surveys* [Silva et al. 2024a].

Datasets and Resources. The dissertation produced publicly available datasets to support research on Portuguese-language literary text mining, including: (i) *PPORTAL_ner*, an annotated corpus for literary Named Entity Recognition [Silva and Moro 2024d], and (ii) datasets associated with bias analysis in masked language models and generative models [Silva et al. 2025b, Silva et al. 2025c]. All resources are available with persistent identifiers (Zenodo).

Code and Models. The proposed framework and models are publicly available, supporting reproducibility and reuse: (i) the *PORTALIA* framework: `https://github.com/marianaossilva/gender_pipeline`, and (ii) domain-adapted NER models: `https://huggingface.co/marianaossilva`.

Further Publications. Beyond the core thesis contributions, the research resulted in 8 complementary publications, including: (i) 2 journal papers (*Journal of Information and Data Management* [Silva et al. 2022b] and *Journal of In-*

formation Systems [Silva et al. 2022a]), (ii) an international conference (PROPOR [Scofield et al. 2022b]), and (iii) 5 national venues (BraSNAM [Silva et al. 2023b, Silva et al. 2021b], DSW SBBD [Silva et al. 2021a], WebMedia [Scofield et al. 2022a], DS4SG SBBD [Silva et al. 2024c]). These works include two awards: **Best Paper** [Silva et al. 2021a] and **Second Best Paper** [Silva et al. 2021b].

Scientific Initiation Co-advisorships. The research also contributed to undergraduate training through co-advising of scientific initiation projects focused on gender and narrative analysis.

Byproducts. Besides the products directly and indirectly related to the Ph.D., the knowledge acquired during the process has also contributed to other publications and datasets (as first author) on the contexts of social and music computing, as well as digital government: WebMedia 2022 (**Best Paper Runner-up)**, Journal on Interactive Systems 2023, Vórtex Music Journal 2023, DSW 2022, WCGE 2023, SBBD 2024 (**Best Paper Runner-up**), Journal on Interactive Systems 2024, Journal of Information and Data Management 2024, Bracis 2024.

References

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics Derived Automatically from Language Corpora Contain Human-like Biases. *Science*, 356(6334):183–186.

Caton, S. and Haas, C. (2024). Fairness in Machine Learning: A Survey. *ACM Comput. Surv.*, 56(7).

Freitas, C. and Martins, F. (2023). Bela, Recatada e do Lar: O Que a Mineração de Textos Literários Nos Diz Sobre a Caracterização de Personagens Femininas e Masculinas. *Fórum Linguístico*, 20(3):9118–9138.

Freitas, C. and Santos, D. (2023). Gender Depiction in Portuguese. In *Conference Reader: 2nd Annual Conference of Computational Literary Studies (CCLS2023)*, pages 4–30.

Hoyle, A. M., Wolf-Sonkin, L., Wallach, H. M., Augenstein, I., et al. (2019). Unsupervised Discovery of Gendered Language through Latent-Variable Modeling. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL*, pages 1706–1716. Association for Computational Linguistics.

Lima, L. F. F. P. d. and Araujo, R. M. d. (2023). A Call for a Research Agenda on Fair NLP for Portuguese. In *STIL*, pages 187–192. SBC.

Luo, K., Mao, Y., Zhang, B., and Hao, S. (2024). Reflecting the Male Gaze: Quantifying Female Objectification in 19th and 20th Century Novels. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 13803–13812, Torino, Italia. ELRA and ICCL.

O. Silva, M. and Moro, M. (2025). Aprimorando o Reconhecimento de Entidades Nomeadas em Textos Literários em Português com Modelos Adaptativos. *Linguamática*, 17(1):77–94.

Schulz, D. and Štěpán Bahník (2019). Gender Associations in the Twentieth-century English-language Literature. *Journal of Research in Personality*, 81:88–97.

Scofield, C., Silva, M., and Moro, M. (2022a). What makes a book successful? a study on portuguese-language literature. In *Anais Estendidos do XXVIII Simpósio Brasileiro de Sistemas Multimídia e Web*, pages 69–72, Porto Alegre, RS, Brasil. SBC.

Scofield, C., Silva, M. O., de Melo-Gomes, L., and Moro, M. M. (2022b). Book Genre Classification Based on Reviews of Portuguese-Language Literature. In *15th International Conference Computational Processing of the Portuguese Language - PROPOR*, pages 188–197. Springer.

Silva, M., ao, M. B., and Moro, M. (2025a). Gender Bias in Portuguese Literary Texts: A Masked Language Model Approach. In *Anais do XVI Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 407–419, Porto Alegre, RS, Brasil. SBC.

Silva, M., Brandão, M., and M. Moro, M. (2025b). Gender Bias in Portuguese Literary Texts: A Masked Language Model Approach.

Silva, M., Brandão, M., and M. Moro, M. (2025c). Rewriting Stories with LLMs: Gender Bias in Generated Portuguese-language Narratives.

Silva, M. and Moro, M. (2024a). NLP Pipeline for Gender Bias Detection in Portuguese Literature. In *Anais do LI Seminário Integrado de Software e Hardware (SEMISH)*, pages 169–180. SBC.

Silva, M. O., ao, M. A. B., and Moro, M. M. (2024a). Gender Bias Detection in Literary Texts: A Systematic Literature Review. *ACM Computing Surveys*. Under submission.

Silva, M. O., Brandão, M. A., and Moro, M. M. (2025d). Rewriting Stories with LLMs: Gender Bias in Generated Portuguese-language Narratives. *Journal of the Brazilian Computer Society*, 31(1):1120–1136.

Silva, M. O., de Melo-Gomes, L., and Moro, M. M. (2024b). From Words to Gender: Quantitative Analysis of Body Part Descriptions within Literature in Portuguese. *Information Processing & Management*, 61(3):103647.

Silva, M. O., Melo-Gomes, L., and Moro, M. (2023a). Gender Representation in Literature: Analysis of Characters' Physical Descriptions. In *Anais do XI Symposium on Knowledge Discovery, Mining and Learning*, pages 17–24, Porto Alegre, RS, Brasil. SBC.

Silva, M. O. and Moro, M. M. (2024b). Evaluating Pre-training Strategies for Literary Named Entity Recognition in Portuguese. In *Proceedings of the 16th International Conference on Computational Processing of Portuguese, PROPOR 2024, Santiago de Compostela, Galicia/Spain, 12-15 March, 2024*, pages 384–393. Association for Computational Lingustics.

Silva, M. O. and Moro, M. M. (2024c). PPORTAL_ner: An Annotated Corpus of Portuguese Literary Entities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2024*, pages 12927–12937. European Language Resources Association (ELRA).

Silva, M. O. and Moro, M. M. (2024d). PPORTAL_ner: An Annotated Corpus of Portuguese Literary Entities.

Silva, M. O., Oliveira, G., and Moro, M. (2023b). Analyzing Character Networks in Portuguese-language Literary Works. In *Anais do XII Brazilian Workshop on Social Network Analysis and Mining*, pages 115–126, Porto Alegre, RS, Brasil. SBC.

Silva, M. O., Oliveira, G., and Moro, M. (2024c). Data Insights on Gender Representation: Analyzing the Book and Music Industries. In *Anais Estendidos do XXXIX Simpósio Brasileiro de Bancos de Dados*, pages 338–347. SBC.

Silva, M. O., Scofield, C., de Melo-Gomes, L., Botelho, J. E., et al. (2022a). Brazilian Reading Preferences in Goodreads: Cross-state and Cross-region Analyses. *iSys - Brazilian Journal of Information Systems*, 15(1):25:1–25:20.

Silva, M. O., Scofield, C., de Melo-Gomes, L., and Moro, M. M. (2022b). Cross-collection Dataset of Public Domain Portuguese-language Works. *Journal of Information and Data Management*, 13(1).

Silva, M. O., Scofield, C., and Moro, M. M. (2021a). PPORTAL: Public Domain Portuguese-language Literature Dataset. In *Anais do III Dataset Showcase Workshop*, pages 77–88, Rio de Janeiro, Brazil. SBC.

Silva, M. O., Scofield, C., Oliveira, G., Seufitelli, D., et al. (2021b). Exploring Brazilian Cultural Identity Through Reading Preferences. In *Anais do X Brazilian Workshop on Social Network Analysis and Mining*, pages 115–126, Porto Alegre, RS, Brasil. SBC.

Stuhler, O. (2024). The Gender Agency Gap in Fiction Writing (1850 to 2010). *Proceedings of the National Academy of Sciences*, 121(29):e2319514121.