Scientific
Research
Publishing

A Natural Language Processing Approach to Promote Gender Equality: Analysing the Progress of Gender-Inclusive Language on the Victorian Government Website

Arushi Raichur, Nicole Lee, Rezza Moieni

Cultural Infusion Pty. Ltd., Melbourne, Australia
Email: aarushi.r@culturalinfusion.org.au

Abstract

This research paper explores the role of language in shaping cultural and social attitudes towards gender and the importance of using gender-inclusive language to promote gender equality and eliminate gender bias. Specifically, we analysed the Victorian government website in Australia https://www.vic.gov.au/ over some years from 1970-2023 to Measure and compare the evolution of its language across this period in terms of gender neutrality. To conduct this research, we created three datasets by scraping data from three different websites. The first two datasets comprise a list of masculine and feminine words used in the English language, while the third dataset comprises a list of gender-neutral words approved by the Victorian government. We compared this list of words with the data provided by the Victorian government website, and based on this analysis, we assessed how gender-neutral the website was. The findings show that using gender-inclusive language is a powerful way to promote gender equality and eliminate gender bias. Moreover, our analysis reveals that the Victorian government website has become more gender-neutral over the years, which signals their commitment to promoting gender equality and inclusivity. Overall, our research underscores the importance of using gender-inclusive language in all communication, including website content, to create a more inclusive and equitable society.

Keywords

1. Introduction

The Victorian government has prioritized gender diversity and inclusive language in recent years. They have developed guidelines that recommend inclusive and respectful language while avoiding gender bias. These guidelines serve as a resource for government departments, agencies, and employees to promote inclusive communication. Efforts have been made to adopt gender-neutral language in government communications, using inclusive terms and avoiding assumptions or exclusions based on gender. The Victorian government website serves as a comprehensive resource for information and services provided by the Victorian government. It covers a range of services, programs, news, and announcements from various departments, including health, education, business, jobs, wellbeing, and transportation services.

The phrase "Ladies and Gentlemen!" is a common way to address a group of people. However, it reinforces a binary view of gender by categorizing individuals into two groups: women (ladies) and men (gentlemen). This language assumes that everyone can be neatly classified into these two categories, disregarding and excluding those who do not identify within the traditional gender binary. Gender bias refers to the unequal treatment or perception of individuals based on their gender, often resulting in favoring one gender over the other. This bias can manifest in various ways, including attitudes, behaviors, social norms, and institutional practices that reinforce stereotypes, preconceived notions, and discriminatory practices.

Consistently using gendered language like "Ladies and Gentlemen" highlights the dominance of gender over other social categories. It emphasizes gender as a primary means of identification and reinforces societal norms and expectations associated with being a man or a woman.

From a young age, children are socialized into these gender norms through their families, peers, media, and education systems. They are exposed to gender stereotypes and often taught that certain behaviours, interests, and roles are appropriate for boys or girls. This early exposure to gender stereotypes can shape their beliefs and attitudes throughout their lives. The reference to Steffens and Viladot (2015) suggests that research supports the theory of early development and acquisition of gender stereotypes. This research likely provides an overview of studies examining the influence of socialization on the formation and perpetuation of gender stereotypes from childhood into adulthood.

Using inclusive language ensures that people are not excluded from conversations or work. It recognizes the diversity of individuals we interact with and serve, both in direct communication and when describing someone who is not present. Inclusive language is a way to acknowledge and respect the range of people in our society.

Inclusive language acknowledges and respects the diversity of bodies, genders, and relationships. By reviewing website content, (including the homepage, About us, and Services pages), we assess the use of gender-neutral terms and alterna-

tives to gender-specific pronouns in order to determine the level of inclusivity on these websites, comparing data with masculine, and feminine, and gender-neutral datasets, and drawing conclusions about their inclusivity rate.

2. Literature Survey

2.1. The Importance of Gender and Age Inclusivity in Digital Spaces

With our global society's digitisation, websites have become integral conduits through which organisations reach out to and interact with their clients, stakeholders, and the general public. It is essential for these digital platforms to reflect and respect the diversity of their visitors, including their varying genders and age groups. However, a growing body of evidence suggests that websites, like other forms of media, may exhibit biases based on gender and age in their design, language, and content (Ward & Grower, 2020). These biases can have negative consequences, such as perpetuating stereotypes, creating an unwelcoming environment for certain user groups, or alienating potential users and customers who feel underrepresented or excluded. Therefore, promoting gender and age inclusivity in digital domains is not only a matter of social justice but also a strategic necessity for organizations (CSW, 2023) that aim to connect with a diverse audience.

2.2. Understanding Gender and Age Bias in Online Content

The content of a website, including its language, visuals, and design elements, plays a pivotal role in creating an environment that is either inclusive or exclusive for its users. Biases can subtly manifest in the choice of words, tone of language, imagery, and even the design of the user interface (Kaye, Wall, & Malone, 2016).

For example, certain words or phrases may be predominantly associated with either masculinity or femininity due to societal norms and stereotypes. Words like "power", "race", "strength", and "competition" are typically linked to masculinity (Gruys & Munsch, 2020), while words such as "love", "sympathy", and "emotions" tend to be associated with femininity. Overreliance on gendered language can unintentionally introduce a gender bias into website content, making it more appealing to one gender while potentially alienating the other.

Likewise, age bias can emerge when website content caters primarily to the interests, linguistic styles, or technological proficiency of a specific age group, leaving others feeling marginalized or unable to fully engage with the site. For instance, the use of trendy slang or references may attract a younger audience but could disenchant older users. On the other hand, complex navigation or small font sizes may pose accessibility challenges for visually impaired users (Hart et al., 2008).

Therefore, ensuring gender inclusivity in website content is crucial to fostering an unbiased and welcoming digital space. This not only promotes safety for

gender-diverse people but also helps organizations expand their user base and enhance user satisfaction.

3. The Impact of Gender Inclusivity on User Engagement and Website Success

Success

The inclusivity of website content plays a crucial role in determining user engagement and overall success. When websites embrace gender inclusivity, they create an environment where gender-diverse users feel a sense of belonging, leading to increased engagement and return visits (Sundar et al., 2014).

Research found that websites using gender-inclusive language and imagery resulted in improved user satisfaction for all genders (Remigio & Talosa, 2021).

Moreover, websites that provide gender-inclusive content attract a more diverse audience, expanding their user base. This is particularly beneficial for businesses as it allows access to a wider market and potential customers, contributing to growth and profitability. For non-profit or informational platforms, it enables the dissemination of messages or services to a larger demographic, maximising their societal impact.

4. Measuring Gender and Age Inclusivity Using Python and Machine Learning

For years, the concept of diversity has been overlooked and poorly defined by social scientists and demands a comprehensive understanding (Moieni et al., 2017) (Moieni & Mousaferiadis, 2022) proposed a fractal analysis with four dimensions—ethnicity, country of birth, languages, and worldviews—to assess diversity. Additionally, they introduced a method in 2021 to measure the level of representation or mutuality between two cohorts (Moieni et al., 2017). Advanced technologies like machine learning offer promising solutions for detecting and addressing gender biases in job advertisements. By programming machine learning algorithms to recognize gender-specific words and phrases, unbiased analyses of potential biases in job descriptions can be conducted (Moieni & Mousaferiadis, 2022). Recently, Moieni et al. (2023) published a machine learning algorithm to predict organizational diversity in 2023.

Machine learning, a subset of artificial intelligence, that enables systems to learn from data and improve through experience, provides an effective approach for evaluating and mitigating gender biases in website content. Natural Language Processing (NLP), a specialized field within machine learning focused on computer-human language interactions (IBM, 2023), is particularly valuable for detecting gender biases in textual information.

In our data analysis, we leveraged the power of both Excel and Python. Python libraries such as BeautifulSoup and Pandas were utilised for basic data cleaning, sorting, and filtering tasks. Python complemented Excel by offering extensive li-

braries like sci-kit-learn and TensorFlow, providing a wide range of machine learning algorithms and advanced analytics techniques. Python's ecosystem facilitated training and deploying machine learning models, enabling predictive analytics and more sophisticated analyses.

By incorporating Python algorithms alongside machine learning, it becomes possible to train them to identify words and expressions commonly associated with specific genders. For instance, an algorithm can detect an abundance of "masculine" words like "strength" and "competition" (Gruys & Munsch, 2020), or "feminine" words like "love" and "emotions", suggesting potential gender bias in the content. By integrating these tools into the design and content creation process, organizations can proactively monitor and rectify biases, ensuring a more inclusive digital environment for all users.

5. Challenges and Opportunities in Leveraging Machine Learning for Inclusivity

Harnessing the power of Python and machine learning to identify and rectify gender biases on websites presents both opportunities and challenges. Acquiring a representative dataset for gender analysis can be difficult, as gender information is often not readily available in website content and may require manual annotation or reliance on external data sources, which is time-consuming and prone to errors.

Additionally, gender analysis involves subjective interpretation, necessitating domain expertise and an understanding of social and cultural contexts. Different interpretations can introduce subjectivity into the analysis.

Another challenge is the lack of standardized methodologies and benchmarks for gender analysis on websites, making it difficult to compare and generalize findings across different studies.

However, Python's versatility, extensive library ecosystem, and strong community support make it a popular choice for data analysis. With practice and familiarity, users can overcome these challenges and leverage Python's strengths to efficiently conduct advanced data analysis tasks.

Machine learning models can be constantly updated and honed as they learn from more data, enabling them to adapt to evolving language use and cultural norms. The use of machine learning in tandem with human supervision and interpretation can thus provide a potent tool for overseeing and improving the inclusivity of digital spaces (Ribeiro et al., 2016).

Moving forward, research and development in machine learning and AI ethics can contribute to more refined and effective tools for bias detection and mitigation. With the rising recognition of the significance of digital inclusivity, the integration of machine learning in web design and content creation offers immense potential for a more inclusive and equitable digital world.

6. Key Findings

Based on the bar graph (see Figure 1 and Table 1), we can observe the following

insights regarding the overall masculinity among the four departments: Communities, Public Sector, Health & Social Support and Victorian Health:

1) Department Comparison: Among the four departments, the Victorian Health Department (Category D) stands out with the highest number of masculine words, comprising 91 out of 222 words. This indicates a relatively strong presence of masculine language within the content scraped from websites associated with this department.

2) Lowest Masculinity: On the other hand, the communities department (Category A) exhibits the lowest number of masculine words, with only 36 out of 222 words. This shows that the websites within this department tend to have a comparatively lower usage of masculine language. The significant difference in masculinity levels between the departments implies that the choice of language and the degree of masculinity in website content varies.

Upon further analysis of the bar graph (see Figure 2 and Table 2), it becomes evident that the usage of feminine words across the departments follows a similar pattern to that of masculine words. Department D, representing the Victorian Health Department, again stands out with the highest count of feminine words, totaling 88 out of 210 words. This suggests a notable presence of feminine language within the content scraped from websites associated with this department.
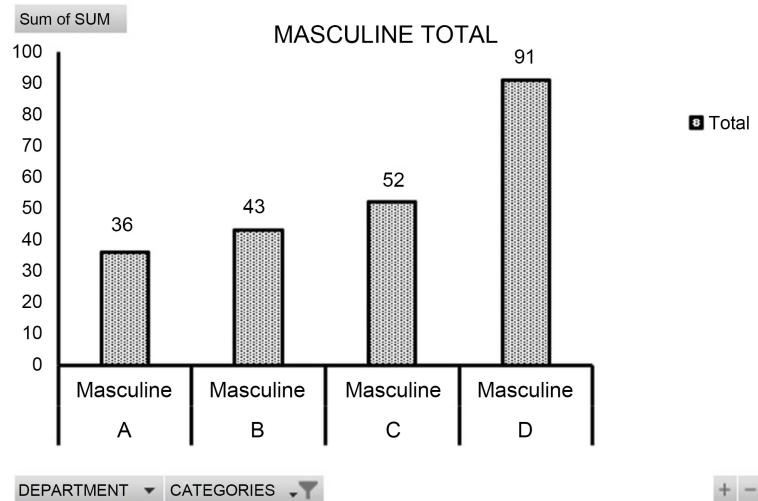
Figure 1. Department analysis with most masculine words.

Table 1. Masculine words.

DEPARTMENT	SUM
A	36
B	43
C	52
D	91
SUM	222
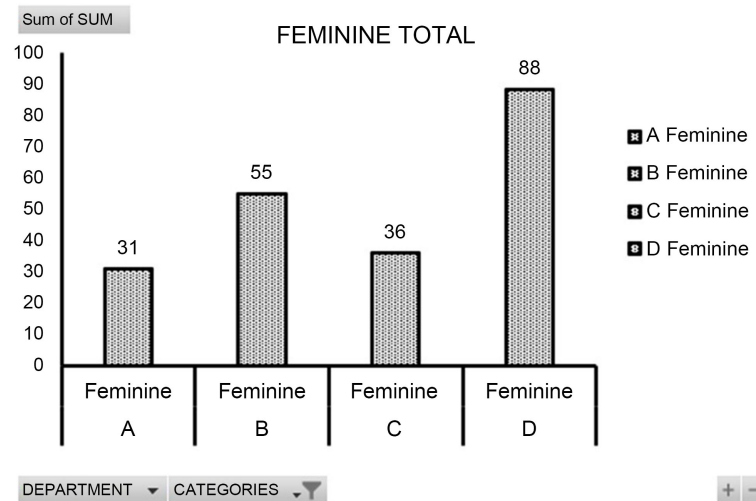
Figure 2. Department analysis with most feminine words.

Table 2. Feminine words.

DEPARTMENT	SUM
A	31
B	55
C	36
D	81
SUM	210

Conversely, the Communities Department (Category A) exhibits the lowest count of feminine words, with only 31 out of 210 words. Department C falls in the middle range, indicating a moderate usage of feminine language on the websites associated with health and social support.

A noteworthy observation to consider is the historical establishment of Department D's websites. As these websites have been in existence for a longer period, dating back to the early stages of Victorian website development, it is understandable why they exhibit higher dominance of both masculine and feminine words. The linguistic patterns found in Department D's websites may have been influenced by societal norms, communication trends, or the overall evolution of language over time.

Overall, this analysis not only highlights the presence of feminine words across the departments but also draws attention to the broader context of Department D's websites being established for a significant duration.

Upon a detailed examination of the gender-inclusive websites, it is evident that the public sector department comprises the highest number of gender-inclusive websites (see **Figure 3** and **Table 3**). Out of the total 478 websites analyzed, 154 websites in Department B prioritize gender inclusivity in their content. This indicates a significant effort by the public sector to ensure gender diversity and inclusiveness in their online presence.
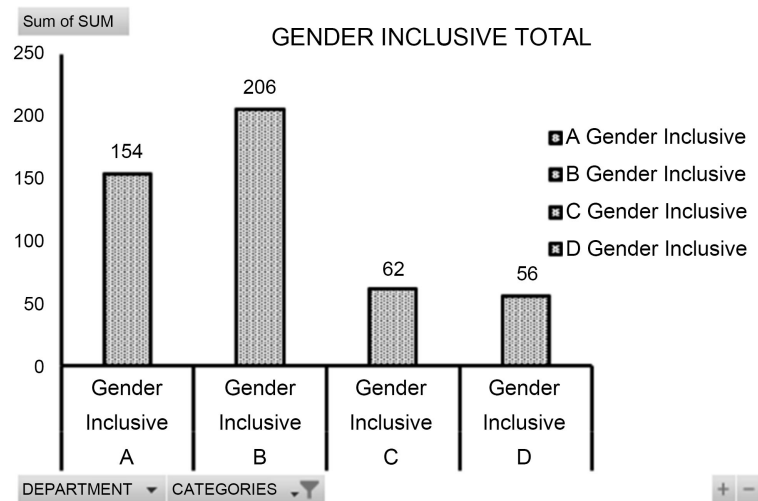
Figure 3. Department analysis with most gender-neutral words.

Table 3. Gender neutral words.

DEPARTMENT	SUM
A	154
B	206
C	62
D	56
SUM	478

In contrast, Department D, which represents the Victorian Health Department, has the fewest number of gender-diverse websites, with only 56 out of 478. It is worth noting that the websites in this department might have been established prior to the emphasis on gender inclusivity in web content. However, it is important to consider that after the launch of the public sector websites in 2014, all subsequent Victorian government websites have prioritized gender diversity.

Category A (Communities) ranks next in terms of gender diversity, with 152 gender-inclusive websites out of the total 478. This suggests that websites associated with community-focused departments also place a significant emphasis on gender inclusivity in their content.

Following Category A, Category C, which represents better health and social support, demonstrates a commendable effort with a substantial number of gender-inclusive websites.

These findings highlight the varying degrees of gender inclusivity among the different departments. Department B (Public Sector) takes the lead in fostering gender diversity, while Department D (Victorian Health) lags, potentially due to the historical establishment of its websites. The launch of the public sector websites by the Victorian government in 2014 marks a turning point for gender inclusivity across Victorian government websites, with subsequent websites prioritizing gender diversity.

The bar graph analysis reveals intriguing insights into the gender inclusivity and linguistic characteristics of various government department websites in the Victorian region. Department A (Communities) exhibits a remarkable dedication to gender inclusivity, with 154 gender-inclusive words, alongside conscious efforts to minimize masculine (36) and feminine (31) language (see Figure 4 and Table 4). This focus on gender inclusivity can be attributed to the department's primary emphasis on supporting LGBTQ+ and gender-inclusive workplaces through its four dedicated websites, which were launched in 2016. Overall, the graph demonstrates noteworthy progress in fostering gender diversity and inclusivity within the Victorian government's online presence, with particular attention to the evolving landscape of gender-inclusive language and representation.

The analysis of Department B (Public Sector) websites reveals their efforts towards gender inclusivity and equality (see Figure 5 and Table 5). Notably, the department demonstrates a substantial lead in gender-inclusive language, with 206 gender-inclusive words used across its websites. This emphasis on gender inclusivity is further exemplified by the existence of an all-inclusive language guide and a workplace gender equality action plan among the four websites associated with the department.

In contrast to gender-inclusive language, the department utilizes a relatively smaller number of masculine (43) and feminine (55) words. This discrepancy in the usage of gender-specific language reveals the efforts made to promote gender-neutral language.
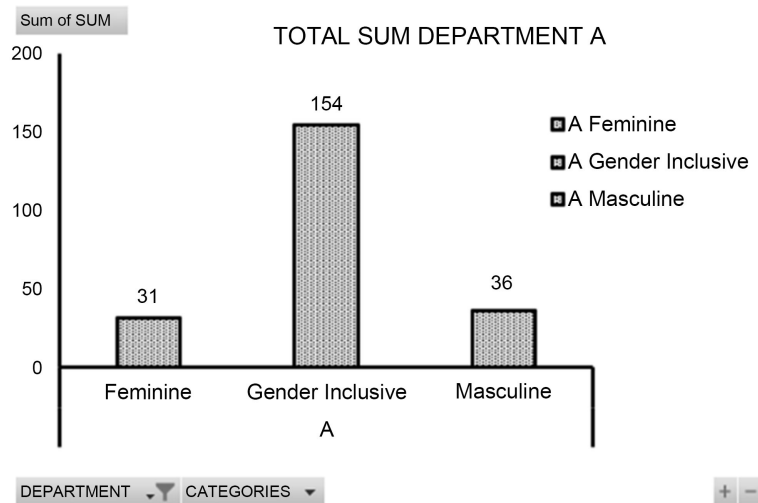
Figure 4. Department A analysis.

Table 4. Department A word distribution.

DEPARTMENT A	SUM
FEMININE	31
MASCULINE	36
GENDER INCLUSIVE	154
SUM	221
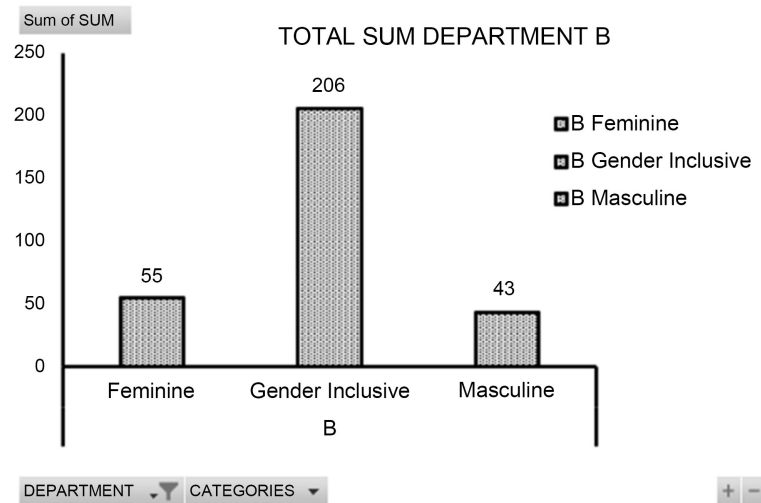
Figure 5. Department B analysis.

Table 5. Department B word distribution.

DEPARTMENT B	SUM
FEMININE	55
MASCULINE	43
GENDER INCLUSIVE	206
SUM	304

The significant gap between gender-inclusive words and masculine–feminine words is a testament to the public sector's commitment to fostering an inclusive and diverse online environment. The introduction of gender-inclusive language guides and gender equality action plans shows the public sector department's proactive approach to ensuring equitable representation and communication in their web content.

The analysis of Department C (Health and Social Support) websites reveals an intriguing dynamic in their language usage (see Figure 6 and Table 6). While the department exhibits a positive effort towards gender inclusivity, with 62 gender-inclusive words, it is noteworthy that masculine words (52) are not far behind, and feminine words (36) are also present.

Department C's primary focus on family well-being, conflicts, and family planning may explain the relatively higher usage of masculine and feminine words. Traditionally, topics related to family and conflicts might have been associated with gender-specific language. However, the emergence of 62 gender-inclusive words shows a conscious effort by the department to adapt its language practices and promote a more inclusive approach to addressing these subjects.

It is important to recognize that the Health Department has a long history, which might have influenced its initial language usage. As societal attitudes shift towards greater recognition of gender issues, people are becoming more empowered to stand up for themselves and demand inclusivity in all aspects, including language.
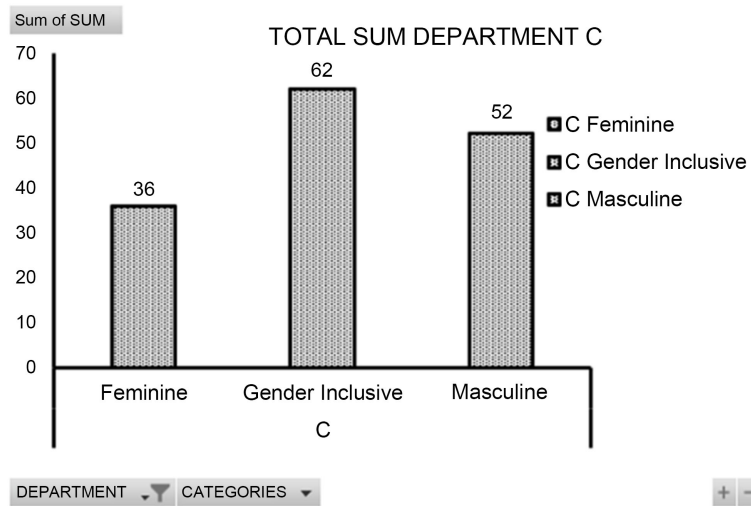
Figure 6. Department C analysis.

Table 6. Department C word distribution.

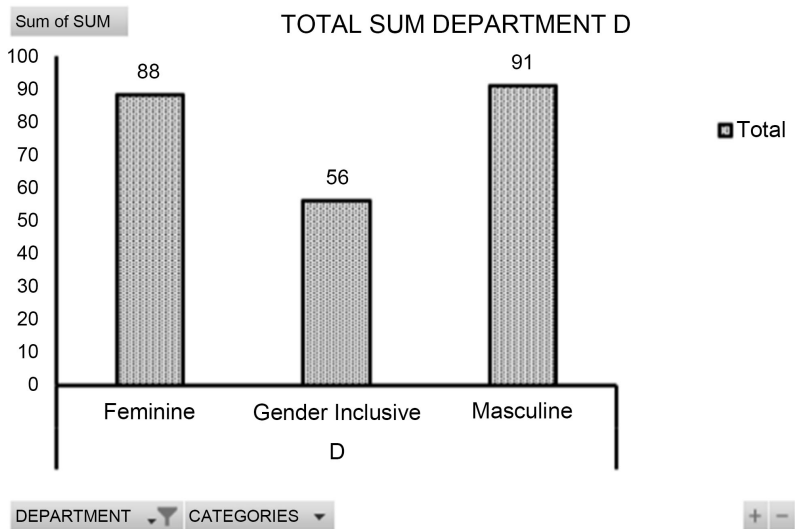| DEPARTMENT C	SUM
FEMININE	36
MASCULINE	52
GENDER INCLUSIVE	62
SUM	150

Societal changes are likely influencing the Health Department's evolving approach to gender-inclusive language. As it adapts to the changing landscape, the Health Department appears to show a growing awareness and commitment to gender-neutral language.

The analysis of Department D (Better Health) websites provides valuable insights into its communication style and focus on promoting healthy living for Victorians (see **Figure 7** and **Table 7**). The department's emphasis on physical activities, outdoor sports, and active lifestyles is reflected in its language usage.

The presence of male-dominated words (91) indicates a historical pattern that might have been influenced by the traditional association of physical activities and sports with masculinity. Similarly, the occurrence of feminine words (88) shows some recognition of gender diversity but still lags behind male-dominated words.

The department's use of gender-inclusive words (56) demonstrates a growing awareness of the importance of inclusive language. As society evolves, it becomes crucial for organizations like the Better Health Department to adapt their communication practices to reflect a more inclusive approach.

Given that the Better Health department was one of the initial departments launched under the Victorian government websites, its language usage might have initially followed traditional gender norms. However, the recent launch of a gender-inclusive guide suggests a positive step towards adopting more inclusive language in their content.

Figure 7. Department D analysis.

Table 7. Department D word distribution.

DEPARTMENT D	SUM
FEMININE	88
MASCULINE	91
GENDER INCLUSIVE	56
SUM	253

The analysis of the top 10 gender-inclusive, feminine, and masculine words used on the Victorian government website provides compelling evidence of its commitment to gender diversity and inclusivity (see **Figure 8** and **Table 8**). Among these words, "GENDER" stands out as the most frequently used, appearing 81 times. This reflects a conscious effort by the government to acknowledge and address matters related to gender identity and representation.

In terms of feminine words, "WOMEN" appears prominently with a count of 38 (see **Figure 9** and **Table 9**). This highlights the government's dedication to recognizing and promoting issues that specifically concern women, their rights, and their well-being.

Likewise, the use of "MEN" 22 times showcases the government's attention to male-related matters (see **Figure 10** and **Table 10**), ensuring that their experiences and concerns are adequately represented and addressed.

The presence of gender-inclusive, feminine, and masculine words demonstrates the Victorian government's commitment to embracing a comprehensive and balanced approach to communication. By catering to the diverse needs and experiences of all genders, the government aims to foster an inclusive and supportive environment for every human being.

Furthermore, the establishment of different departments, such as Communi-

ties, Public Sector, Health and Well-being, and Better Health, reflects the government's proactive strategy to address a wide range of social, health, and public service needs. Each department's focus on gender diversity and inclusivity contributes to an overarching effort to improve the overall lifestyle and well-being of all citizens, regardless of gender or identity.

Overall, the analysis underscores the Victorian government's dedication to promoting a gender-diverse, inclusive, and supportive environment for the entire population. By using gender-inclusive language and recognizing the unique needs of various genders, the government demonstrates its commitment to creating positive social change and enhancing the quality of life for every individual in the region.
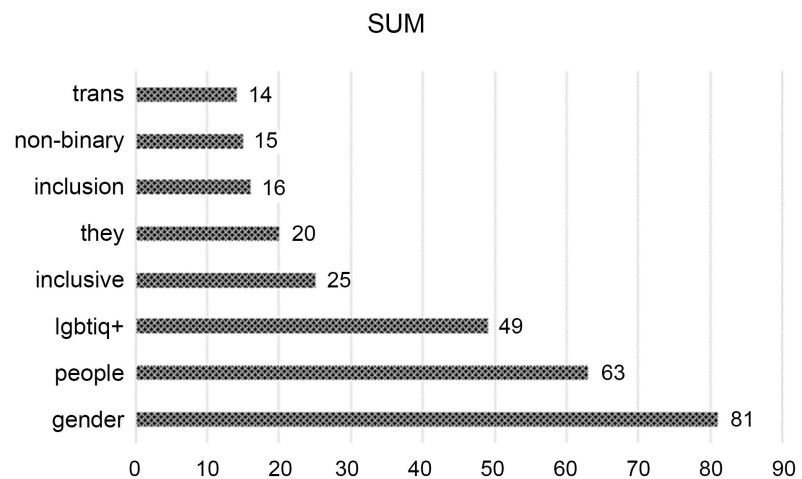
Figure 8. Top 10 gender inclusive words.

Table 8. Top 10 gender inclusive words.

GENDER INCLUSIVE WORDS	COUNT
TRANS	14
THEY	20
PERSON	14
PEOPLE	63
NON-BINARY	15
LGBTIQ+	49
INCLUSIVE	25
INCLUSION	16
GENDER DIVERSE	14
GENDER	81
SUM	311

TOP FEMININE WORD COUNT
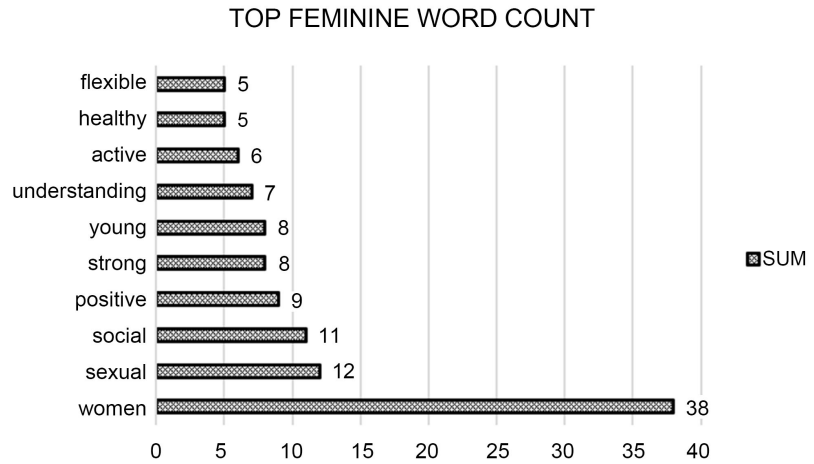
Figure 9. Top 10 feminine words.

Table 9. Top 10 feminine words.

FEMININE WORDS	COUNT
FLEXIBLE	5
HEALTHY	5
ACTIVE	6
UNDERSTANDING	7
YOUNG	8
STRONG	8
POSITIVE	9
SOCIAL	11
SEXUAL	12
WOMEN	38
SUM	109

TOTAL MASCULINE WORD COUNT
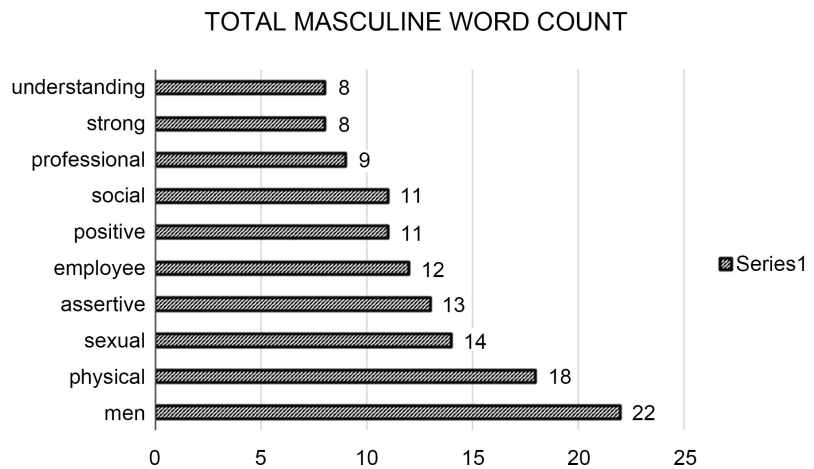
Figure 10. Top 10 masculine words.

Table 10. Top 10 masculine words.

MASCULINE WORDS	COUNT
ASSERTIVE	12
EMPLOYEE	12
MEN	22
PHYSICAL	18
POSITIVE	11
PROFESSIONAL	9
SEXUAL	14
SOCIAL	11
STRONG	8
UNDERSTANDING	8
SUM	126

7. Conclusion

The necessity for gender inclusivity on websites is backed by extensive literature underscoring its benefits to user engagement, website success, and ethical digital design practices. As digital domains continue to assume an increasingly crucial role in society, the demand for inclusivity intensifies, and its absence becomes more detrimental.

Python, machine learning, particularly through Natural Language Processing, offers a promising method to gauge and enhance inclusivity. It presents scalability, efficiency, and the capability to analyse copious amounts of data in a manner that human analysis cannot. Nonetheless, challenges persist, especially in terms of the risk of perpetuating biases present in training data and the nuanced understanding required for identifying subtle biases.

Python is a preferred language for machine learning and artificial intelligence (AI) applications. With libraries such as sci-kit-learn, TensorFlow, and PyTorch, Python offers robust tools for building and deploying machine learning models. As the demand for AI-driven data analysis continues to grow, Python's role in this domain will become even more prominent.

In summary, the intersection of machine learning and web design offers significant promise for fostering inclusivity. As research progresses and tools become more sophisticated, the vision of a truly inclusive digital world appears an achievable goal.

Acknowledgements

Quincy Hall, and Catherine McCredie for their insights and constant support of this work. We also would like to thank the University of Melbourne for providing an opportunity to Aarushi Raichur by helping her gain better understanding of real-world practical applications.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

CSW (2023). *Innovation and Technological Change, and Education in the Digital Age for Achieving Gender Equality and the Empowerment of All Women and Girls.* https://www.unwomen.org/sites/default/files/2023-02/CSW67-Expert-Group-Meeting-report-en.pdf

Gruys, K., & Munsch, C. L. (2020). "Not Your Average Nerd": Masculinities, Privilege, and Academic Effort at an Elite University. *Sociological Forum, 35,* 346-369.

Hart, T. A., Chaparro, B. S., & Halcomb, C. G. (2008). Evaluating Websites for Older Adults: Adherence to "Senior-Friendly" Guidelines and End-User Performance. *Behaviour & Information Technology, 27,* 191-199. https://doi.org/10.1080/01449290600802031

IBM (2023). *What Is Natural Language Processing?* https://www.ibm.com/topics/natural-language-processing#:~:text=Natural%20language%20processing%20(NLP)%20refers,same%20way%20human%20beings%20can

Kaye, L. K., Wall, H. J., & Malone, S. A. (2016). Turn That Frown Upside-Down: A Contextual Account of Emoticon Usage on Different Virtual Platforms. *Computers in Human Behavior, 60,* 463-467.

Moieni, R., & Mousaferiadis, P. (2022). Analysis of Cultural Diversity Concept in Different Countries Using Fractal Analysis. *International Journal of Organizational Diversity, 22,* 43-62. https://doi.org/10.18848/2328-6261/CGP/v22i01/43-62

Moieni, R., Mousaferiadis, P., & Roohi, L. (2023). A Study on Diversity Prediction with Machine Learning and Small Data. *Open Journal of Social Sciences, 11,* 18-31. https://doi.org/10.4236/jss.2023.112002

Moieni, R., Mousaferiadis, P., & Sorezano, C. O. (2017). A Practical Approach to Measuring Cultural Diversity on Australian Organizations and Schools. *International Journal of Social Science and Humanity, 7,* 735-739. https://doi.org/10.18178/ijssh.2017.V7.917

Remigio, M. T. R., & Talosa, A. D. (2021). Student's General Attitude in Gender-Inclusive Language. *International Journal of Evaluation and Research in Education, 10,* 864-870. https://doi.org/10.11591/ijere.v10i3.21573

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778

Steffens, M. C., & Viladot, M. A. (2015). *Gender at Work: A Social Psychological Perspective.* Peter Lang. https://doi.org/10.3726/978-1-4539-1534-9

Sundar, S. S., Bellur, S., Oh, J., Xu, Q., & Jia, H. (2014). User Experience of On-Screen Interaction Techniques: An Experimental Investigation of Clicking, Sliding, Zooming,

Hovering, Dragging, and Flipping. *Human-Computer Interaction, 29,* 109-152. https://doi.org/10.1080/07370024.2013.789347

Ward, L. M., & Grower, P. (2020). Media and the Development of Gender Role Stereotypes. *Annual Review of Developmental Psychology, 2,* 177-199. https://doi.org/10.1146/annurev-devpsych-051120-010630