Inclusiveness Check Feature in Grammar-checking AI Systems: Does AI Force Us to Be Politically Correct?

Dewi Yuri Cahyani
Universitas Udayana

Abstract

This paper aims to conduct a theoretical review of the inclusiveness check feature in grammar-checking AI systems, such as those found in Grammarly and Microsoft Word. The examination of this inclusiveness check feature will involve the consideration of Politically Correct Language (PCL) and the theoretical debate surrounding it, and the process of semantic change. Drawing upon the author's experience of using Grammarly as a writing assistant for her dissertation, this paper argues that such a feature should be critically addressed. It is suggested that these features effectively call out users for being politically incorrect and function as moral police for their way of expressing themselves.

Keywords: artificial intelligence, inclusiveness check, political correctness, politically correct language, semantic change

Introduction

In Indonesia, scientific publications have become one of the important indicators of academic performance. Writing such publications is part of the three obligations of academic community, which include teaching, research, and community service. Academics are also encouraged to publish their research in reputable international journals, many of which require English manuscripts. As non-native English speakers, many academics then turn to artificial intelligence (AI) to assist them in preparing manuscripts – one common tool is grammar-checking AI such as Grammarly (see, among others, Fadhilah (2018), Nova (2018), Novianti (2020), Fitria (2021), and Marliyanda et al (2022)).

The use of grammar-checking AI has proven to be a highly effective tool, especially in scientific writing, which demands clarity and precision. For example, software like Microsoft Word will flag any instances of incorrect tenses or misspelled words. Grammarly, a service dedicated to grammar-checking, allows users to assess correctness, clarity, engagement, delivery, and language style in their writing. However, grammar-checking AI systems like Microsoft Word and Grammarly do more than just check for grammar. For example, Microsoft Word offers optional inclusive style checks that can be turned off in the settings. Even with inclusiveness checks enabled, Microsoft Word does not make changes to documents without user consent, and users can choose to accept or ignore suggestions.

A different case applies to Grammarly. Inclusiveness checks are enabled by default, but, like with Microsoft Word, they do not require us to accept the suggestions. However, Grammarly's rating system is based on the number of corrections we make based on its suggestions, which can negatively impact our overall writing performance if we do not follow the suggestions. Such an arrangement is quite frustrating since most of us dislike low performance, and thus, it makes us vulnerable to falling into their suggestion trap. It also seems unfair since users pay for the service, in the first place, to enhance their writing – not to correct their way of saying or expressing things.

Language plays a significant role in shaping culture, and in turn, culture influences politics. Therefore, using inclusive language and avoiding potentially offensive or demeaning words is crucial. Doing so can help eliminate incorrect stereotypes associated with certain words or

terms, particularly those referring to minority and marginalized groups. However, constantly being corrected by AI to replace certain words for being deemed non-inclusive, offensive, or not "woke" enough can be frustrating mainly because many of us use the software to improve our grammar, not to alter our use of words (which also means, to change our perceptions of reality) or to modify our language style to match the software's preferences (e.g., Grammarly's preference for active over passive voice).

While we can argue that bias and offensiveness should be valid considerations in proper grammar, the primary argument against inclusiveness checks focuses on the software's prescriptive nature. That is, the software dictates what we should and should not say – a characteristic often associated with defenders of Politically Correct Language (PCL). In the past, humans made decisions about what was considered politically correct. Now, in our creation of AI, we are incorporating the notion of Political Correctness (PC) into the programs we create. In this regard, software such as Microsoft Word and Grammarly use AI not only to ensure grammatical correctness but also to promote PCL, yet they do not explicitly label this feature as a "political correctness check," even though that is essentially what it is.

Based on such a standing, this paper aims to theoretically review the inclusiveness check feature in grammar-checking AI systems (e.g., Grammarly and Microsoft Word). Previous studies on grammar-checking AI in Indonesia have focused on how such software can help write scientific papers. For example, Fadhilah (2018) studied the use of Grammarly among lecturers; Ambarwati (2021) explored Grammarly's performance in providing formative feedback; and Nova (2018), Novianti (2020), Fitria (2021), and Marliyanda et al. (2022) researched the use of Grammarly for English writing among university students. Therefore, by focusing on the inclusiveness check feature of grammar-checking AI systems, this paper seeks to address the gap in the studies; furthermore, rather than focusing on just specific software like Grammarly, this paper will also include other software with an inclusiveness check feature as its subject of observation.

To examine the inclusiveness check feature in grammar-checking AI systems, this paper will start by discussing Politically Correct Language (PCL) and the theoretical debate surrounding it, followed by a discussion of changes in language semantics. Finally, this paper will make an argument against the inclusiveness check feature, followed by a conclusion on the matter.

Politically Correct Language (PCL)

The term "political correctness" has been given a variety of definitions in recent years. According to the Oxford Dictionary, it means "*the avoidance of terms and behavior considered to be discriminatory or offensive to certain groups of people*." Meanwhile, the online Lexico Dictionary, powered by Oxford, defines it as "*the avoidance of forms of expression or action that are perceived to exclude, marginalize, or insult groups of people who are socially disadvantaged or discriminated against.*" However, the use of passive forms (e.g., considered, perceived) and the absence of a subject in such definitions raise a question: Who decides what is offensive?

Another definition of political correctness, according to the online Cambridge Dictionary, is "*the act of avoiding language and actions that could be offensive to others, especially those relating to sex and race.*" In addition, the online Merriam-Webster Dictionary considers being politically correct as "*conforming to a belief that language and practices which could offend political sensibilities (as in matters of sex or race) should be eliminated.*" Here, the use of the modal verbs "can" or "could" raises another question: Are the language or actions in question truly offensive as alleged?

In short, Politically Correct Language (PCL) is a language intended to minimize offense, especially when describing groups identified by external markers such as race, gender, culture, or sexual orientation. (Roper, 2024) Despite "political correctness" being a loose term, its advocates argue that it positively impacts society. They claim that it prevents the use of words with negative or offensive connotations, thus showing respect to people who are unjustly stereotyped. In this view, political correctness aims to prevent bullying and offensive behavior. (O'Neill, 2011: 279) PC's advocates also aim to replace terms with offensive connotations with more politically correct ones. For example, instead of using terms like "black" or "yellow" to describe people of African or Asian descent, the term "people or person of color (POC)" is introduced, which is perceived to be more neutral. This strategy is said to have two recognized advantages: first, it reduces the social acceptability of using offensive terms, and second, it discourages the reflexive use of words that import a negative stereotype, thus promoting fair and conscious thinking in describing others based on their merits. (O'Neill, 2011: 280)

Despite critiquing the use of words with negative connotations or offensive language as a matter of civility, proponents of political correctness (PC) claim that the very notion of PC is actually a myth. It is an invention of critics who allege it to be a progressive initiative, but in reality, it is designed to undermine their opponents without substantial argument. Hutton (2001) asserts that PC is a tool used by the American Right in the mid-1980s to undermine American liberalism. Toynbee (2009) further argues that PC is merely a baseless right-wing accusation meant to empower its users to attack those on the left. (O'Neill, 2011: 280)

In reality, even though the term "political correctness" is often denied by those it is attached to, it has become a significant and much-debated concept over the last few decades. It has been the subject of discussion, dispute, criticism, and satire by commentators from across the political spectrum. (Roper, 2024) Historically, the term first appeared in Marxist-Leninist vocabulary following the Russian Revolution of 1917 to describe adherence to the policies and principles of the Communist Party of the Soviet Union — which is, the party line. During the late 1970s and early 1980s, the term began to be used by liberal politicians to refer to the extremism of some left-wing issues, particularly regarding what was perceived as an emphasis on rhetoric over content. In the early 1990s, the term was used by conservatives to question and oppose what they perceived as the rise of liberal left-wing curricula and teaching methods on university and college campuses in the United States. By the late 1990s, the usage of the term had again decreased, and it was most frequently employed by comedians and others to lampoon political language. At times, the left also used it to scoff at conservative political themes. In today's society, political correctness has found its new battle cry — that is, Cancel Culture. (Thiele, 2021) Cancel Culture itself is not new, since time and again, we have witnessed a phenomenon where various parties have made demands that something not be shown, said, or exhibited publicly.

However, apart from the theoretical debate regarding political correctness, which is still ongoing to this day, and anyone related to the PC is generally refusing to be identified with the movement, the interest of this paper is to test whether the alleged purpose of the PCL – which is, to prevent bullying and offensive behavior, is true. For this particular intention, we need to examine the process of semantic change and the reason that terms become offensive or inoffensive.

Semantic Change and the Alleged Purpose of PCL

To understand the motivation behind using Politically Correct Language (PCL), it is essential to comprehend its intended purpose. Advocates of PCL argue that its use can discourage the use of offensive or negatively connotated words, thereby preventing the unfair stereotype depiction of individuals or groups. To test the veracity of this claim, it is crucial to examine the etymology of words – the history of words and how their meaning has changed over time. Why are words such as "nigger," "retarded," "moron," "dwarf," or "queer" considered derogatory? Are they inherently offensive, or is it the context in which they are used that makes them offensive?

Ben O'Neill (2011) argues that the word's lexicology does not indicate a hostile meaning. To support his argument, he uses the term "mentally retarded" as an example. He notes that to "retard" something means to hinder or impede it, to make it slower or diminish its development or progress in some way. Thus, describing someone as "mentally retarded" means that their mental processes are somehow impeded, hindered, diminished, or slowed down. This meaning is certainly accurate; that means the term "mentally retarded" is actually also an accurate descriptor of such a condition. It is also neutral because the term itself does not imply a value judgment about such diminished mental function.

Perhaps, though, the term contains an implied negative meaning because the brain condition that is generally expected to have is one that functions without any obstacles, so being "mentally retarded" is considered something bad and undesirable. This is where perhaps the negative connotation of this term comes from. However, according to O'Neill, this is insufficient to justify the claim that the term is offensive.

O'Neill argues that the perceived offensiveness of terms is not necessarily inherent in their literal meaning but rather in the way they are delivered and the context in which they are used. (2011: 281) For example, suppose the term "mentally retarded" is used by schoolyard bullies to insult or humiliate someone with diminished mental function. In that case, the term becomes offensive due to the spiteful tone and context in which it is used. If such use becomes widespread, the term may develop an additional meaning as an insult, carrying with it the implication that mental retardation is shameful and deserving of ridicule. Thus, in this case, a term (retarded) that is intrinsically a neutral predictor of certain mental conditions may become a negative connotation because of its use as an insult.

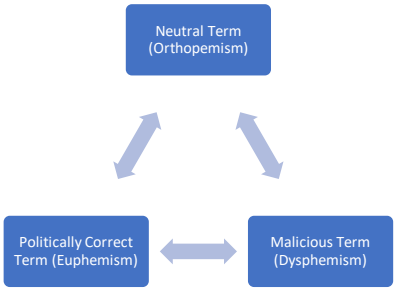
In this regard, Politically Correct Language (PCL) is claimed to be designed to solve this bullying problem and its etymological by-product; where its proponents adopt the strategy of periodically replacing the words used as insults with new terms to avoid negative connotations imbued in existing terms.

Unfortunately, bullies will not stop their actions just because new words are used to refer to the characteristics they want to use for insulting people. When a word like "mentally retarded" is replaced by a euphemism like "differently abled," those who want to use it as a neutral descriptor of mental condition will do so. However, those who want to use it as an insulting term will do so as well. Following this logic, the term "differently abled" will gradually lose its neutral meaning and become another offensive or malicious term. In this context, feminist Germaine Greer notes that euphemisms tend to lose their function quickly through their association with the reality they represent. As a result, they need to be regularly replaced by new euphemisms. (Greer (1971) on O'Neill, 2011: 282)

Therefore, this word-replacement strategy of political correctness creates a cycle that leads to what Pinker (1994) called "the euphemism treadmill." In this process, a term that starts as neutral (an orthophemism) can become negative over time as it is used as an insult, turning into

a malicious term (a dysphemism). This term is then replaced with a more politically correct term (a euphemism), which becomes widely accepted and seen as the appropriate neutral expression, even if its actual meaning may not be entirely neutral. This cycle repeats itself.

Image 1. The Euphemism Treadmill

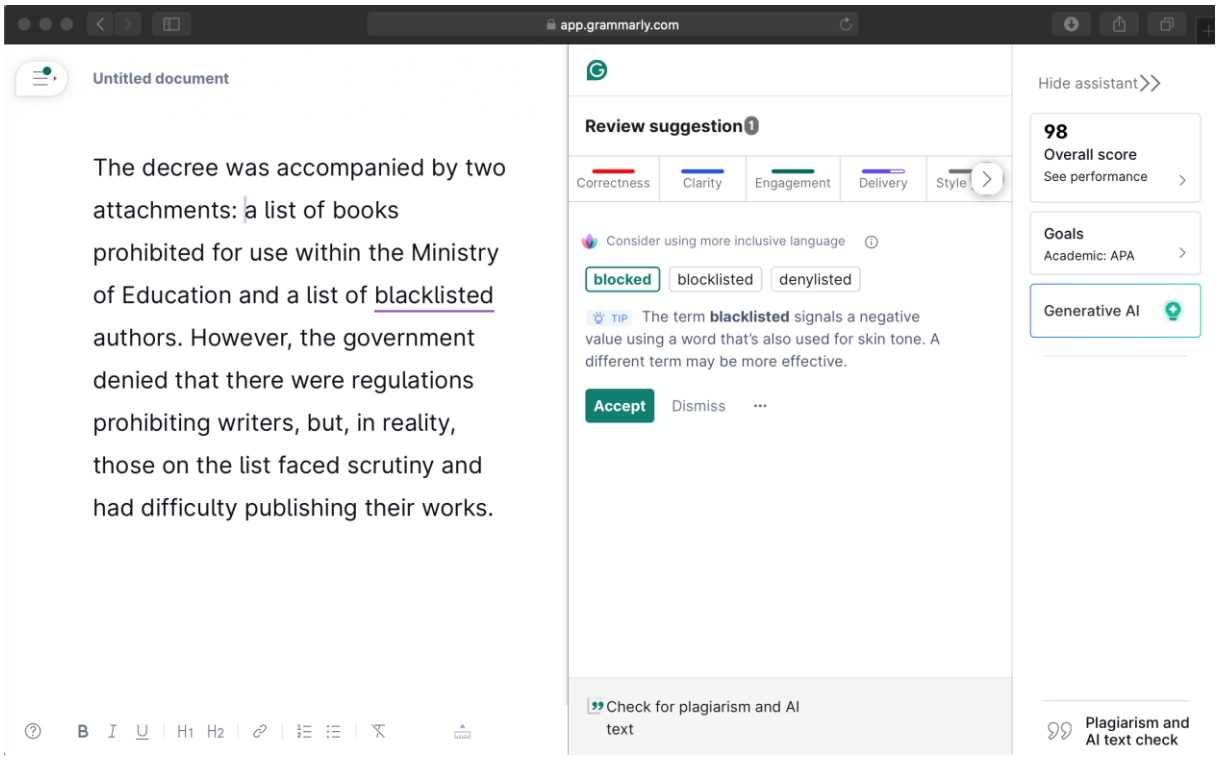


The euphemism treadmill is a slow process, but nonetheless is a cyclical one. Even when a term is found to resolve the problem of negative semantic change, it is short-lived. The new, neutral word eventually becomes an insult as bullies use it. As long as the social dynamics remain the same, the cycle repeats indefinitely, leading to a growing list of discarded dysphemisms. This explains why political correctness can never be effective over a long period of time. (Burkhardt, 2010 in O'Neill, 2011: 282-283)

Discussion: The Impact of PCL on Discourse and The Case Against Inclusiveness Check Feature

I will use my experience in writing my dissertation to present an argument against the inclusiveness check feature found in software like Microsoft Word and Grammarly. For example, when I used the word "blacklist" in a piece discussing book banning, Grammarly suggested alternative words, explaining that "blacklist" could be offensive to certain groups. I understand that in the United States, the term "black" has historically been associated with negative connotations when referring to African-descent communities. Thus, using "black" to describe negative phenomena, such as the banning of books, is seen as reinforcing its negative connotations and being hurtful towards people with a black complexion.

Image 2. Grammarly Screenshot



However, I rejected the suggestion. One of the reasons is that the other suggested terms such as "blocked," "blocklisted," "denylisted," and sometimes "banned" or "prohibited" do not accurately describe the phenomenon that I want to explain. The situation I want to convey involves authors being placed on a list for further scrutiny, rather than being immediately banned. Therefore, I declined the suggestion out of concern that it might change the way I express myself, which I believe should be a matter of personal preference, and even worse, it could alter the meaning of the expression.

Secondly, I did not use the word "black" to insult or make an offense against anyone or any group. In this regard, where does the alleged rudeness or offensiveness of the term come from? As we can see in Image 2, Grammarly gives perfect scores for the aspects of correctness, clarity, engagement, and language style but not for the delivery aspect, where they emphasize that "*the term blacklisted signals a negative value using a word that's also used for skin tone. A different term may be more effective.*" The term "black" itself intrinsically does not contain value judgment. Oxford Dictionary, for instance, defines black as "*the very darkest color owing to the absence of or complete absorption of light; belonging to any human group having dark-colored skin; and characterized by tragic or disastrous events.*" As O'Neill (2011: 281) argues, the offensiveness of a term relies not only on its literal meaning but also on its delivery, including the tone and context in which it is delivered. In this case, the term "blacklist" itself, as we can see, is not inherently offensive and I do not use the word "black" to insult people or in an offensive tone of voice. Thus, Grammarly warns that the term I use may be offensive, where actually there is no offense intended.

The inclusiveness check feature as such, therefore, actually bull-eyes the wrong targets. Although the reason for avoiding terms that are considered politically incorrect seems noble, in reality, bullies remain bullies. In practice, even if such a term as "blacklist" is defamed and replaced by new terms, bullies will seek other "degrading" words to insult, intimidate or offend black communities if that is what they want. They will not curb their actions merely because a

new word is now used to refer to the characteristics that they wish to use as a basis for insulting people. For example, the more accepted term for people with a darker complexion today is a person of color. Bullies will easily use this new term also to mock or insult dark-colored people or communities.

The act of bullying can be displayed through people's intention to offend, which can be shown in three observable factors: the language used, the context of the remarks, and the tone in which they are delivered. (O'Neill, 2011: 284) Someone who intends to insult others primarily does so through their tone of voice and may not necessarily use explicitly insulting language, although they may still use insulting words. Advocates for politically correct language often focus solely on the choice of words. This means that proponents of Politically Correct Language may encourage people to take offense at remarks where no offense is actually intended. For instance, a doctor uses terms such as mentally retarded or dwarfism to describe a patient's condition, or older people use outdated terms out of ignorance of the new, more acceptable term. Consequently, periodically replacing words considered offensive or having negative connotations with new terms may not effectively address the goal of solving bullying problems, as defenders of Politically Correct Language claim.

Moreover, the problem with this drive for Politically Correct Language is that it attempts to deal with the problems of negative semantic change by outlawing accurate descriptors rather than by trying to rehabilitate them or use them with proper context and tone (e.g., replacing black with person of color; dwarf with vertically-challenged person; mental disability with special needs; wheelchair-bound with movement problem; pedophile with minor-attracted people; etc.). While avoiding "derogatory" terms is the ideal, the absence of accurate descriptors for certain conditions can cause other problems, such as wrong treatments and more discrimination against the groups despised by the use of such terms. Thus, in my opinion, the proper approach to dealing with the problems of negative semantic change is to continue to use the existing words in circumstances that make it clear that no negative intention is intended, and the process of semantic change is acceptable as long as it does not alter our thoughts via speech. Because in most cases, the enemy of political correctness is actually not bullies, but the studious, literate persons who understand the proper meaning of words and want to use them correctly. In this regard, the movement of PCL can have a long-term impact on discourse since it can change the way we express and see the world. As the Sapir-Whorf hypothesis illustrates, our perception of reality is determined by our thought processes, which are influenced by the language we use. In this way, language shapes our reality and tells us how to think about and respond to that reality.

Another problem with the push for Politically Correct Language is that it does not only seek to denounce and replace perceived offensive terms or to create specifically politically correct terms embodying a respectful stance towards minorities or marginal groups, but it also calls for monitoring of what can or cannot be said, especially regarding sex, race, and ethnicity, and ridicule those who continue using outdated, offensive or not inclusive terms. In this regard, political correctness can be a threat to language and mind since such bullying can be used to discredit opponents without proper argumentation. In a democratic society, different opinions and ideologies should ideally find a peaceful way of coexisting. By censoring speech on account of it being potentially offensive – and thus, politically incorrect, or imposing social repression following political incorrectness, political correctness actually silences the very diversity it is supposed to promote. On the contrary, such a drive for political correctness hinders societal progress since progress cannot take place in the absence of dissent, confrontation, conflict, and debate. In this regard, what was stated by the philosopher Karl R. Popper (1984: 146) founds its relevance: persecutions carried out in the name of moral causes (here, advocates of political correctness argue their cause is a moral one to correct the social inequality in our society) are just as bad as those carried out in the name of terrible ones.

Therefore, as Politically Correct Language becomes more prevalent in Artificial Intelligence (AI) and AI plays an increasingly significant role in our lives, it is important to establish the ethical and moral boundaries for these programs. While using inclusive and non-discriminatory language is crucial, the responsibility for upholding moral principles should rest with individuals rather than with AI or its creators. Developing AI is an ongoing process, so the ethical programming of AI will continue to be a major issue in the coming years and requires immediate attention.

Conclusion

The inclusiveness check feature in grammar-checking systems or software, such as those in Microsoft Word and Grammarly, should not be taken for granted. It needs to be critically addressed because this feature calls out users for being politically incorrect, acting as moral police for their way of expressing themselves.

Bibliography

Ambarwati, E. K. (2021). Indonesian university students' appropriating Grammarly for formative feedback. *ELT in Focus, 4*(1), 1-11.

Bihan-Colleran, C. L. (2020). Feminist Linguistic Theories and "Political Correctness": Modifying the Discourse on Women? *The ESSE Messenger, 29*(1), pp. 120-192.

Fadhilah, U. (2018). Penggunaan grammarly untuk penulisan artikel bahasa Inggris dosen Stikes Hangtuah Tanjungpinang. *Jurnal Keperawatan, 8*(2), 884-895.

Fairclough, N. (2003, January). 'Political correctness': the politics of culture and language. *Discourse & Society, 14*(1), 17-28.

Fitria, T. N. (2021). Grammarly as AI-powered English writing assistant: Students' alternative for writing English. . *Metathesis: Journal of English Language, Literature, and Teaching, 5*(1), 65-78.

Fox-Genovese, E. (1995, May-June). A Kafkaesque Trap. *81*(3), 8-14.

Marliyanda, A., Wachyudi, K., & Kartini, D. (2022). Analisis Survei Terhadap Pengguna Grammarly. *Jurnal Educatio, 8*(3), 1147-1152.

McCahill, L. (2018, April 3). *edgy.app*. Retrieved from Edgy: https://edgy.app/micro-how-ai-got-pc-wrong

Nova, M. (2018). Utilizing Grammarly in evaluating academic writing: A narrative research on EFL students' experience. *Premise: Journal of English Education and Applied Linguistics, 7*(1), 80-96.

O'Neill, B. (2011). A Critique of Politically Correct Language. *The Independent Review, 16*(2), 279-291.

Roper, C. (2024, August 6). *Britannica*. Retrieved from britannica.com: https://www.britannica.com/topic/political-correctness#:~:text=political%20correctness%20(PC)%2C%20term,%2C%20culture%2C%20or%20sexual%20orientation.

Scalcau, A. (2020, November). The Paradoxes of Political Correctness. *Theoretical and Empirical Researches in Urban Management, 15*(4), 53-59.

Scanlon, L. (1995, May-June). A Victimless Crime. *Academe, 81*(3), 9-15.

Thiele, M. (2021). Political correctness and Cancel Culture – a question of power!. The case for a new perspective. *Journalism Research, 4*(1), 50-57. Retrieved from journalistik.online: https://journalistik.online/en/debate-en/political-correctness-and-cancel-culture%E2%80%84-%E2%80%84a-question-of-power/