CASE STUDY: TRANSLATION FROM ENGLISH INTO EUROPEAN PORTUGUESE USING GENDER-NEUTRAL LANGUAGE
DO AI CHATBOTS PERFORM BETTER THAN MT SYSTEMS?

Lisbeth Ferreira[1]

lferreira@letras.up.pt

FACULDADE DE LETRAS DA UNIVERSIDADE DO PORTO (PORTUGAL)

ABSTRACT. There is a growing demand for gender-neutral writing in translation tasks motivated by the need for inclusion, neutrality or economic factors. Studies suggest that current Machine Translation (MT) and Artificial Intelligence (AI) systems tend to mirror the source text's (and real-world) gender bias or default generic masculine. It is important to understand whether, with documents in a gender-neutral style, the MT and AI tools provide better gender-neutral translations by resorting to standard and established grammatical solutions. Additionally, the effectiveness of AI systems, with customised prompts, in surpassing MT limitations regarding gender-neutral translation is also analysed.

KEY-WORDS. Gender-Neutral Language, Gender-Neutral Translation, Machine Translation, Artificial Intelligence (AI), Prompts.

RESUMO. A procura pela escrita de género neutro na tradução tem aumentado por vários motivos que vão desde a inclusão e a neutralidade até aos fatores económicos. Os estudos sugerem que a Tradução Automática (TA), fornecida por motores de tradução e sistemas de Inteligência Artificial (IA), tende a refletir o género do texto de partida ou a utilizar o masculino genérico. É importante compreender se, com documentos escritos num estilo de género neutro, a TA oferece melhores resultados em termos de escrita de género neutro, recorrendo a soluções gramaticais padrão e estabelecidas. Analisa-se, também, a eficácia de *prompts* personalizados nos sistemas de IA para ultrapassar as limitações da TA neste tipo de escrita.

PALAVRAS-CHAVE. Linguagem Neutra de Género, Tradução Neutra de Género, Tradução Automática, Inteligência Artificial (IA), *Prompts*.

[1] Estudante do 3.º ano do 3.º Ciclo de Estudos em Ciências da Linguagem.

1. Introduction

Over the last few years, there has been an increasing demand for the use of Gender-Neutral Language (GNL) (Cho, Kim, J. Kim, S. & Kim, N. 2019b; Kostikova, Daems & Lazarov 2023; Sun, Webster, Shah, Wang & Johnson 2021). One of the main reasons for this is a broader social awareness in terms of equality, inclusivity and recognition of different gender identities so as to respect the identity of the speaker/author and to avoid misgendering text recipients. Moreover, the use of GNL is also particularly useful when the gender of the target audience is unknown and when there is a need to make communications more universal and suitable for a diverse audience.

In institutional communications such as official documents, business communications, or academic papers, the use of GNL is becoming standard practice (Piergentili, Fucci, Savoldi, Bentivogli & Negri 2023; Sun *et al.* 2021). It helps to maintain a professional and inclusive tone, avoiding assumptions about the gender identity of the recipients while also avoiding the use of the generic masculine. In theory, the use of GNL also simplifies the translation process, especially between languages that have similar approaches to gender conventions. It can allow a more efficient, straightforward and cost-effective translation process across different linguistic and cultural contexts.

It is not the aim of this research to assess the reasons that may lead to this demand for Gender-Neutral Translation (GNT) by customers, but rather to analyse the quality of GNT provided by the currently available Machine Translation (MT) and Artificial Intelligence (AI) systems. In this context, our primary emphasis is on GNL, particularly resorting exclusively to the use of rephrasing and neutralisation strategies, thus Indirect Non-Binary Language (INL)[2].

Even though gender bias affects several monolingual tasks in natural language processing (NLP), it is even more noticeable in bilingual tasks, such as translation. When

[2] It is important to note that the incorporation of neopronouns and neomorphemes (Direct Non-Binary Language – DNL) falls outside the scope of this research.

While languages such as English have strategies that are gaining widespread acceptance, Romance languages like Portuguese, Spanish and Italian cannot resort to the "they" singular solution, for example. These languages are incorporating neopronouns (e.g., EN ze/zir instead of he/she/him/his/her), neomorphemes (e.g., ES -e/-es instead of -o/-os and -a/-as), and other alternatives (e.g., graphemic devices such as IT -@ instead of -a/-o/-e/-i), which allow reference to individuals without resorting to generic words (Piergentili *et al.* 2023).

To bridge this gap, the Portuguese "Elu" system was developed. It consists on the introduction of neopronouns, such as "elu/delu" instead of the traditional "ele/dele" or "ela/dela", as well as neomorphemes, thus changing traditional morphological structures. For instance, the Elu system employs neutral forms like "-e" or "-es", as in "bonite/bonites", for example, instead of "bonito/bonitos" (masculine endings "o" and "os") or bonita/bonitas" (feminine endings "a" and "as").

translating from a notional gender language (a language in which grammatical gender is inferred or implied rather than marked by grammatical forms, as is the case in English) into a grammatical gender language, the challenge becomes even greater. This process requires neutralisation strategies that can sometimes compromise the effectiveness and neutrality of the texts and always involve making choices.

In notional gender languages, gender is primarily conveyed through personal pronouns and possessive adjectives (e.g., he/him/hers; she/her/hers) and lexically gendered words (e.g., man; woman). On the other hand, grammatical gender languages, such as Portuguese (as well as Spanish, French and Italian, among others), exhibit a morphosyntactic agreement system, where various parts of speech, including verbs, determiners, and adjectives, have gender inflexions. There are also genderless languages (e.g. Finnish and Turkish) in which gendered words are limited to certain lexical pairs.

Current MT systems, such as Google Translate[3] and DeepL[4], operate at the sentence level and are based on large corpora that include texts from different subjects and styles. As a result, they tend to follow the morphological number and gender cues used in the source text. In general, this can cause information to be lost, namely the lack of number and gender agreement when translating from a gender-marked to a gender-poorer language, often leading to morphological mistakes that ultimately affect the quality of the translation. When it comes to gender, the issue extends beyond potential grammatical agreement problems; the absence of anaphoric or cataphoric reference can also result in the use of the generic masculine (which may reveal gender bias according to recent studies) instead of a non-marked form (Cho, Kim J., Kim, M. & Kim, N. 2019a; Sun *et al.* 2021). While human translators use contextual information to ensure gender agreement or to avoid bias, most current MT systems do not (Vanmassenhove, Hardmeier & Way 2019), and they depend mainly on large corpora and statistical information.

Although not specifically designed for translation tasks, AI systems leverage extensive contextual understanding and nuanced language generation capabilities, which can potentially enhance gender-neutral translations. This potential improvement forms the basis for investigating whether AI systems can outperform traditional MT systems in producing gender-neutral translations.

[3] https://translate.google.com/?sl=en&tl=pt&op=translate
[4] https://deepl.com

In this case study, we compare the translations provided by MT systems (Google Translate and DeepL) and AI systems (ChatGPT[5] and Google Bard[6]) for sentences extracted from gender-neutral writing guidelines[7] to assess whether the latter provide better Gender-Neutral Translation (GNT) results than the former. We hypothesise that more detailed prompts will yield better results with AI systems, as recent studies (Jiao, Wang, Huang, Wang, Shi & Tu 2023; Yamada 2023) have shown that customising prompts with specific instructions and context improves the quality of AI-generated translations. In the provided hyperlink[8], it is possible to access the 38 selected sentences, the international guidelines used, and the translations obtained for the 38 sentences that constitute the dataset. The AI chatbots were provided with prompts ranging from simple instructions with no information on GNL to richly informed ones meant to function as a sort of 'translation brief' aimed at producing fully gender-neutral translations. It should be considered that the sentences tested in this study, as they are taken from international guidelines already written using GNL, have greater potential for generating good results in terms of gender-neutral writing. A similar case study with a completely different corpus would likely have rather different results.

As this is not a real translation task with a specific purpose, audience and application, but rather a study to assess if GNL is used in two types of automatic translation tools, no post-editing was done on any of the translations provided by the MT or AI systems. The assessment was made: a) taking into account only the GNL request; and b) based on the assumption that translation support tools as the ones used for this study require post-editing; therefore, a final version cannot be used "as-is", i.e., without the intervention of a translator.

[5] https://chatgpt.com

[6] https://gemini.google.com
On 8 February 2024, Google Bard changed its name to Gemini. As the case study was developed before, the former designation was maintained - https://blog.google/products/gemini/bard-gemini-advanced-app/.

[7] The following guidelines were used as references for Table 1 and for the translation data analysis. English: E1 - United Nations, 2018; E2 – General Secretariat, Council of the European Unit, 2018; E3 – European Parliament, 2018; E4 – North Atlantic Treaty Organisation, 2020 / Portuguese: P1 – Commission for Citizenship and Gender Equality (CIG); P2 – European Parliament, 2018; P3 – IGEN – Fórum; P4 – University of Porto.

[8] https://github.com/LF-DCL/Gender_Neutral_Translation_Case_Study - Includes a downloadable Excel file with the complete dataset, as well as the prompts/interactions used in the study.

2. Related work

Research on gender bias in translation technologies has gained significant prominence over the last few years, and the presence of bias in MT has been extensively studied. This bias has a negative impact on NLP (Sun, Gaut, Tang, Huang, ElSherief, Zhao, Mirza, Belding, Chang & Wang 2019), namely MT. According to Friedman & Nissenbaum (1996), bias in computer systems can be divided into three categories: pre-existing bias, technical bias, and emergent bias. These categories can also be applied to gender bias in MT (Savoldi, Gaido, Bentivogli, Negri & Turchi 2021).

Mitigating gender bias in MT requires a comprehensive approach encompassing historical, technical, and emergent dimensions. Refining data curation, making informed architectural choices, and aligning systems with evolving societal values are essential for inclusive and equitable language processing.

Underrepresentation and stereotyping in MT involve biased visibility and negative generalisations. Underrepresentation occurs when MT misrepresents women or overlooks non-binary individuals, which has an impact on communication. Stereotyping perpetuates biased views, affecting self-esteem and reinforcing societal prejudices. The widespread integration of MT in web applications magnifies these challenges, emphasising the interconnected nature of underrepresentation and stereotyping (Savoldi *et al.* 2021). Studies of GNL either approach the incorporation of meta-information in the source text or focus on rewriting techniques (Vanmassenhove, Emmery & Shterionov 2021). Most studies in this field adopt a binary perspective on gender, with a considerably smaller prevalence of studies taking a non-binary approach.

2.1. Binary approach

The binary approach refers to the use of gender-specific terms rather than gender-neutral alternatives. This approach reflects traditional linguistic norms, where masculine forms are often used generically, reinforcing stereotypes and perpetuating inequalities (Piergentili *et al.* 2023). Translation with a generic masculine approach predominates due to volume-related and social bias-related aspects. This is why "cops are usually men and nurses are usually women" (Cho *et al.* 2019b).

Dinu, Mathur, Federico & Al-Onaizan (2019) explore a method for eliminating bias by augmenting the MT input with additional information: a black-box approach where MT learns

how to use external terminology provided. Meanwhile Prates, Avelar & Lamb (2020) analyse the phenomenon of machine bias in the translation of pronouns with a comprehensive list of job positions in 12 different gender-neutral languages that they translate into English using Google Translate. They conclude that Google Translate has a higher probability of translating sentences with masculine pronouns than with feminine or neutral pronouns, especially in STEM (Science, Technology, Engineering and Mathematics) areas. Vanmassenhove *et al.* (2019), for their part, present a compilation of vast datasets with information from speakers of 20 languages and develop a set of experiments that incorporate the gender information through tags into neural machine translation for different language pairs, concluding that this can greatly improve translation quality in some pairs.

Stanovsky, Smith & Zettlemoyer (2019) present a challenge set and assessment protocol for analysing gender bias in MT for eight target languages with grammatical gender, based on a morphological analysis (Spanish, French, Italian, Russian, Ukrainian, Hebrew, Arabic and German). They conclude that all tested MT systems (Google Translate, Microsoft Translator, Amazon Translate and Systran) reveal bias. Aligned with these findings, Kocmi, Limisiewicz & Stanovsky (2020) use the WinoMT challenge set to analyse coreference and bias in translation from English into grammatically gendered languages, also observing that all systems they tested show worse performance when the professional in question is female. Saunders & Byrne (2020) introduce a novel perspective on addressing gender bias through a domain adaptation framework. Their study demonstrates significant enhancements in the WinoMT challenge set when adapting to datasets across three language pairs, advocating small-domain adaptations as a more efficient strategy for mitigating gender bias in MT.

Rescigno, Vanmassenhove, Monti & Way (2020) present the results of an assessment of MT systems, namely Google Translate, DeepL and Bing Microsoft Translation, and the rendering of grammatical gender, providing a qualitative, quantitative and statistical analysis of the "frequency of female, male and neutral forms in the translation of a list of personality adjectives, nouns related to professions and bigender nouns". Pursuing a different direction, Bentivogli, Savoldi, Negri, Di Gangi, Cattoni & Turchi (2020) question what happens with speech translation in particular, aiming to understand if input as audio signals can provide additional information resulting in a reduction of bias.

Research has shown that the presence of gender bias in MT is strong. This bias reinforces inequalities as does the use of the generic masculine, which is a manifestation of linguistic normativity (Piergentili *et al.* 2023).

2.2. Non-binary approach

A non-binary approach involves strategies to avoid gender-specific language, promoting inclusivity by using gender-neutral terms. Non-binary language can be divided into two categories: direct and indirect. Direct non-binary language (DNL) uses neopronouns and neomorphemes explicitly designed to avoid gender-specific language. Indirect non-binary language (INL) resorts to pluralisation or sentence restructuring to avoid gender-specific terms without explicitly changing the original words (Vanmassenhove *et al.* 2021).

Recent investigations focus on non-binary approaches, such as gender-neutral rewriting. These studies analyze the capacity of MT systems to adapt to GNL through rewriting and post-editing steps.

Sun *et al.* (2021) propose a rewriting and post-processing approach for generating sentences that incorporate the singular "they" and an evaluation benchmark of 500 sentences from Twitter, Reddit, jokes and movie quotes. They define a new task where the system is required to change to neutral forms (e.g. changes such as fireman to firefighter) with "no human-labelled training data". The authors emphasise the scarcity of natural corpora featuring gender-neutral sentences, significantly hindering non-binary GNL studies. Vanmassenhove *et al.* (2021) develop a similar study but with larger synthetic and natural benchmarks.

Kostikova *et al.* (2023) study the capabilities of Adaptive Machine Translation (AMT) when using GNL, namely focusing on non-binary language in English-German translation using the Modern MT engine. They investigate two gender-neutral writing strategies in German, both with Direct Non-binary Language (DNL): the DeE-System and the gender star system. Despite signs of masculine bias in Modern MT, their findings suggest AMT systems can adapt to gender-neutral writing, and increased exposure to gender-neutral corpora can facilitate their evolution.

3. Gender-Neutral Language

Languages contemplate three categories of gender: natural gender, grammatical gender, and social gender (Rescigno *et al.* 2020). Natural gender is typically linked to the biological sex of a person or animal. Grammatical gender, on the other hand, may not align with a word's semantic category and varies across languages, being shaped by cultural

contexts. Finally, social gender involves associating the natural gender of a person with a word based on specific properties, often influenced by societal stereotypes.

The use of GNL proves to be a valuable solution with a wide range of benefits. Besides fostering inclusivity and respecting different identities, it enables the elimination of misgendering, especially when gender identification is a challenge due to insufficient data.

Piergentili *et al.* (2023) highlight three main desiderata for determining if a gender-neutral output is required: a) refraining from expressing gender when it cannot be reliably assumed from the source, which is particularly relevant when translating from a notional gender language into a grammatical gender one; b) using appropriate expressions of gender if they are indirectly conveyed in the source – gender cues (such as third-person pronouns, terms of address, and gender-specific nouns) play a crucial role in determining whether or not a gender-neutral translation is needed; c) refraining from propagating generic masculine in the source language – even though recognizing gender cues may not be straightforward, especially with generic masculine.

3.1. Grammatical gender

All linguistic systems encode or represent gender in some form, whether through grammatical structures, lexical choices, or social conventions. In English (as in Danish, for instance), a notional gender language, gender is primarily conveyed through personal pronouns and possessive adjectives (e.g., he/him/his; she/her/hers) and lexically gendered words (e.g., man; woman). On the other hand, grammatical gender languages like Portuguese (as well as Spanish, French and Italian, among others) exhibit a morphosyntactic agreement system, where various parts of speech, including verbs, determiners, and adjectives, have gender inflexions. There are also "genderless" languages (e.g., Finnish and Turkish) that lack grammatical gender but still convey gender through lexical or cultural mechanisms. In these languages, gender distinction is often based on "animacy" in some lexical pairs.

These differences are particularly noteworthy in translation, especially when the source language lacks gender information about a referent and the target language is a grammatical gender language, as is the case in translation from English into Portuguese. In terms of automatic translation, this language pair is prone to mistakes related to gender bias.

3.2. Guidelines for neutral writing

Several international institutions provide guidelines on gender-neutral writing. Despite the fact that these guidelines address monolingual contexts, they can provide useful insights for translation tasks that require GNL.

The United Nations (2018), the European Union (2018), the European Parliament (2018) and NATO (2020) (along with others) provide guidelines for writing in GNL in English. The guidelines share a common emphasis on the importance of GNL in challenging stereotypes and promoting inclusivity. As for Portuguese, the guidelines made available by the European Parliament (2018), the Portuguese Economic and Social Council (2021) and the IGEN Forum (2018) similarly encourage the use of gender-inclusive language, aligning with international efforts.

3.3. Neutralisation strategies

Efforts to promote linguistic fairness and inclusivity for all gender identities can be categorised into two primary approaches: i) INL, the adoption of gender-neutral formulations, and ii) DNL, the introduction of novel linguistic elements (Piergentili *et al.* 2023).

DNL solutions, such as the use of neopronouns and neomorphemes, cannot be neglected and are indeed necessary in translation in specific contexts, especially when the speaker deliberately uses a neutral pronoun. Despite the importance of these solutions, their usage is not universally deemed grammatically correct or socially acceptable in many linguistic environments and, as mentioned earlier, this linguistic solution falls outside the scope of this text. To employ gender-inclusive language in more formal settings, the use of INL with established gender-neutral conventions within the standard language is expected. Neutralisation strategies can be as simple as choosing gender-neutral words (*firefighter* instead of *fireman*, for example) or may require a total sentence rephrasing without neoelements.

Table 1 presents a summary of neutral solutions and strategies that comply with institutional guidelines as well as with gender neutralisation techniques proposed by Vanmassenhove *et al.* (2021) and Piergentili *et al.* (2023), and follows the presentation proposed by the latter.

A. Epicene synonyms		
EN	E4	*Chairman* → Chair(person)
PT	P4	*Os enfermeiros/As enfermeiras* → O pessoal de enfermagem
B. Pluralisation/Singularisation (towards generic or epicene forms)		
EN	E1	A judge must certify that *he* has familiarised *himself* with... → All judges must certify that they have familiarised themselves with...
PT	P2	*Os diretores*… / → A direção…
C. Relative and indefinite pronouns		
EN	E4	If a staff member is not satisfied..., *he* can ask for a rehearing.
		→ Any staff member who is not satisfied... can ask for a rehearing.
PT	P2	*Os requerentes* devem... / → Quem proceder à apresentação de um requerimento deve…
D. Collective and role nouns		
EN	§	Please contact one of the *waiters*. / → Please contact our staff.
PT	P2	Os *supervisores*… / → As pessoas responsáveis pela supervisão…
E. Omission		
EN	§	A person must reside... before *he* may apply for permanent residence. → ...before _ _ applying for permanent residence.
PT	P2	*O juiz* proferirá a sentença. / → A sentença será proferida.
F. Repetition		
EN	E2	A manager may apply... if permission has been granted by *his* institution. → ...if permission has been granted by that manager's institution.
PT	§	A direção pode recorrer da decisão caso *esta* assim o pretenda. → A direção pode recorrer da decisão caso a direção assim o pretenda.
G. Passive voice		
EN	E4	Each action officer must send *his* document. / → Documents must be sent.
PT	P1	*O candidato* deve enviar o formulário de candidatura até dia x. → O formulário de candidatura deve ser enviado até dia x.
H. Imperative forms		
EN	E4	Each staff member is requested to submit *his* information. → Please submit all information.
PT	§	*O candidato* deve enviar o seu CV. / → Envie o seu CV.
I. Impersonal forms		
PT	§	*O cidadão* perde o direito… / → Perde-se o direito…

Table 1 - Examples of neutralisation strategies[9].

[9] Table format follows Piergentili *et al.* (2023) adapted for EN-PT-PT. In red, *italic,* the generic masculine formulations; in green, underlined, the gender-neutralisation. Column 2 identifies source guidelines; self-created examples (§) or omissions, when not applicable.

The neutralisation strategies presented in the guidelines are targeted towards monolingual tasks. However, they can also be applied in bilingual tasks, especially when gender-neutral writing is necessary in the target text, regardless of whether the source text adheres to this type of writing or not.

4. Proposed method

To develop this comparative case study, 38 sentences in English were taken from the international guidelines on gender-neutral writing, as outlined by the strategies in Table 1. The sentences that make up the analysed dataset[10] are written according to these neutralisation strategies. The aim of this study is to compare the performance of MT systems and AI platforms in the translation of these gender-neutral texts, particularly when AI models are provided with customised prompts. For that, all the sentences were translated with two MT systems and two AI platforms to determine which technology offers better results in maintaining gender-neutral language.

The selected MT systems were DeepL and Google Translate, which are MT tools, while the other two systems used, ChatGPT and Google Bard[11], are Artificial Intelligence (AI) platforms that, although capable of generating translations, were not specifically designed for this purpose. In the case of the AI platforms, three versions of the translations were collected: the first version was generated from a simple and direct prompt; the second from a slightly more GNL-specific prompt; and the third prompt was based on the concept of the "translation-brief" (Nord 1997).

4.1. Translation prompts

One of the features of AI chatbots is the ability to provide detailed instructions, that is to say, fine-tuning prompts so as to render the desired output. The use of AI platforms is still recent, and users are exploring ways to make the most of their features. It is here questioned whether it is possible to achieve a better GNL translation when the AI features are combined with customised prompts that include the translation "skopos" (i.e., the purpose or intended function of the translation) and further details that work as a "translation brief" (understood as

[10] The full dataset is available at: https://github.com/LF-DCL/Gender_Neutral_Translation_Case_Study.
[11] ChatGPT 3.5 and Google Bard Pro 1.0 – both versions available for free online at the time of the study.

a set of instructions outlining the purpose, target audience, and other specific requirements of the translation) (Nord 1997, 2018; Pym 1993; Reiss & Vermeer 2014).

Other recent studies have been exploring the possibility of enhancing translations generated by AI systems through the use of detailed and customised prompts.

Yamada (2023) explores how to optimise MT in ChatGPT by customising prompts, namely with the translation's purpose and target audience, and concludes that this approach can, in fact, greatly increase the output quality. His assessment, developed from a professional translator's perspective, finds that this method improves translation quality based on industry standards, especially for marketing documents and idioms tied to specific cultures.

Jiao *et al.* (2023) also researched ChatGPT's capabilities in terms of MT and concluded that, when strong prompts are used, the quality of the output is equivalent to that of commercial MT systems, but it shows greater limitations with low-resource language pairs. Gao, Wang & Hou (2023) similarly propose several translation prompts, including details on the translation task, context domain data, and part-of-speech (POS) tags, to further unlock ChatGPT's translation potential. Like Yamada (2023) and Jiao *et al.* (2023), the authors conclude that ChatGPT performs better when more detailed prompts are employed.

5. Analysis and results

Each of the 38 sentences that constitute the analysed dataset yielded eight different translations: one by each MT system (Google Translate and DeepL) and three by each AI system (ChatGPT and Google Bard). Three different prompts were used (ranging from simple ones with no information on GNL to customised and detailed prompts), each prompt generating three different outcomes.

The first translations obtained from the AI systems (identified as GPT1 and BD1 in the dataset) were comparable to those generated by DeepL and Google Translate, as no information about GNL was provided. The chatbots were only required to translate the sentences from English into European Portuguese. Providing such limited information in the first prompt was crucial to determine if the customised prompts would, as hypothesised, generate better results.

The results of the second translation produced by the AI systems (GPT2 and BD2) were obtained from an improved prompt deemed fitting, for instance, for human translators who were to produce a gender-neutral translation of the sentences.

The first attempt at generating a gender-neutral translation using a customised prompt was the following: "*Translate the sentences below from English into European Portuguese considering the following instructions. Please translate the following sentences from English into European Portuguese using a gender-neutral language. You can use rewriting and reformulation strategies to avoid masculine and feminine nouns, pronouns, adjectives, and verbs, for instance, passive voice, imperative forms, and impersonal forms among others. You can also replace gendered nouns for epicene synonyms or collective nouns. You cannot use neopronouns or neomorphemes, such as "Elu", "Delu", "amigue", "bonite", nor solutions like "Car@", "Carxs" or solutions that include the male/female like "Caro/as", "bonito/a".*

Before devising the final customised prompt, which was expected to produce the best results for the purposes of this study, several attempts were made with two of the sentences[12]. These tests to fine-tune the customised prompt were done on ChatGPT[13], and the same final prompt was then used on Google Bard. It can be argued that the same process of prompts testing in order to find the most effective one might have produced a different result if run on Google Bard. However, for the purpose of this study, it was necessary to use the same prompt on both systems to achieve the third and final version of the translations.

As shown in the previously provided hyperlink, the level of detail ChatGPT was provided with was different, and additional interactions with the chatbot were necessary so as to provide the machine information regarding why some of the translations presented did not comply with the instructions given. The tests clearly demonstrated that it was quite a challenge to establish the appropriate amount of information the prompts should contain; providing either too much or too little information can create "noise" and affect the performance of the AI system. The third set of AI-produced translations of the sentences (see GPT3 and BD3 in the dataset) were obtained from the prompt that seemed to have rendered better results than the earlier tests.

The final customised prompt used was: *Translate the following sentence from English into European Portuguese. The target sentence must be totally gender-neutral. The nouns, pronouns, adjectives and verbs used in the translation CANNOT be gender-specific, so the sentences MUST be 100% neutral as the client forbids the use of gender marks for professions, job titles, groups of people, etc. Example: instead of "fotógrafo" / "fotógrafa" you*

[12] The interactions with the AI chatbots, as well as the three prompts used, can be accessed via the previously provided hyperlink.
[13] The tests were done on ChatGPT because it was launched before Google Bard.

could use "profissional de fotografia". You can use several rephrasing, paraphrasing strategies such as passive voice, imperative, collective nouns and other strategies. Example: instead of "Se estiver interessado..." / "Se estiver interessada..." you could use "Em caso de interesse…" You cannot use dual gender approaches such as "O/A candidato/a" in any circumstance. You also cannot use neo systems such as neopronouns "elu/delu".

This prompt was considered the most balanced version of all three since it provides some grammatical information but avoids complex metalanguage. It also provides some examples, suggests neutralisation strategies, and defines which expressions and solutions align with the pseudo-client's instructions.

5.1. Assessment criteria

In order to analyse the translations produced by Google Translate, DeepL, ChatGPT and Google Bard, we set up a group of assessors composed of three professional translators: T1 has 1 year of experience, T2 has 4 years of experience, and T3 has over 10 years of experience.

Each of the assessors was given the 38 original sentences in English and the eight different automatic translations for each sentence, totalling 304 translations per assessor. They were instructed that the main criterion of analysis was not the overall quality of the translation. As professional translators accustomed to post-editing, they should understand that the output of MT is not intended to serve as a final translation. They were instructed to assess each of the translation versions with the understanding that they could classify sentences based on more than one criterion, when necessary, and according to the following criteria: *1) Adequate gender-neutral translation; 2) Uses generic masculine; 3) Uses a dual gender approach (e.g Ele/Ela; bonito/bonita); 4) Uses a non-grammatical approach (e.g. todxs/tod@s/todes); 5) Uses a biased-gender approach (e.g. translate* nurse *as feminine); 6) Brazilian Portuguese; 7) Mixed approaches*. The three elements of the assessment group worked separately on the task and did not discuss results with each other.

5.2. Quantitative analysis

Gender-neutral translations

	GT	DL	GPT 1	GPT 2	GPT 3	BD 1	BD 2	BD 3
T1	12	15	11	13	15	10	9	9
T2	10	12	9	11	11	8	7	9
T3	11	13	11	12	13	10	9	9

Table 2 - Number of translations considered gender-neutral.

Non gender-neutral translations

	GT	DL	GPT 1	GPT 2	GPT 3	BD 1	BD 2	BD 3
T1	26	23	27	25	23	28	29	29
T2	28	26	29	27	27	30	31	29
T3	27	25	27	26	25	28	29	29

Table 3 - Number of translations considered Non gender-neutral.

As for the quantitative results, there was very little difference in the total number of sentences that each of the assessors considered to be written in accordance with the required criteria of gender neutrality. Above, Table 2 shows the total number of sentences considered neutral by the three assessors, while Table 3 shows the total number of sentences considered non-neutral by the three assessors. The non-neutral translation results correspond only to criteria "Uses generic masculine" (2) and "Uses a gender-biased approach" (5), as none of the systems produced translations that met criteria (3), (4) and (7). The data in Tables 2 and 3 above encompass the European and the Brazilian varieties of Portuguese. Nevertheless, variety-specific outputs were not identified since this variable did not affect gender neutrality. Yet, the number of sentences translated into Brazilian Portuguese (PT-BR), but considered correct by T1, T2 and T3 regarding the gender neutrality criterion were as follows: two results for GT; two for GPT3; six for BD1; six for BD2 and one for BD3. It should be noted that all three prompts explicitly stated that the translations must be in PT-PT.

5.3. Qualitative analysis

For a more detailed analysis of the qualitative data, only the cases in which the translations were unanimously rated as gender-neutral by the three assessors are analysed. To present the relevant information concisely, the following tables display only the translations,

the ratings from each assessor and the reference to the original sentence. The most relevant cases are also commented on; however, it is possible to analyse and explore all the various responses by accessing the dataset made available in the provided hyperlink, along with the 38 original source sentences identified by their corresponding reference numbers.

5.3.1. DeepL and Google Translate

Ref.	Translation	T1	T2	T3
1	Cada Representante Permanente deve apresentar as suas credenciais ao Protocolo	**1**	**1**	**1**
4	Os planos de externalização de alguns serviços não avançaram ao ritmo previsto, uma vez que se regista uma escassez significativa de mão de obra.	1	1	1
5	Os planos de externalização de alguns serviços não avançaram ao ritmo previsto, uma vez que existe uma escassez significativa de pessoal	1	1	1
15	Serão responsáveis pela gestão dos procedimentos de contratação pública.	1	1	1
16	Uma pessoa pode não saber o seu número de contribuinte.	1	**1**	**1**
19	Cada visitante deve apresentar o seu bilhete de identidade para obter um temporário de visitantes.	1	1	1
21	Cada participante deve certificar-se de que não deixa para trás os seus documentos classificados quando sai da sala.	1	1	1
22	Cada participante deve certificar-se de que não deixa os seus documentos classificados ao sair da sala.	1	1	1
25	Para requerer a residência permanente, é necessário ter residido continuamente no país durante 5 anos.	1	1	1
27	Cada agente de ação deve enviar o seu documento à secretaria para distribuição.	1	1	1
28	Os documentos devem ser enviados ao registo para distribuição.	1	1	1

Table 4 - DeepL translations assessed as GNT by T1, T2 and T3.

Contrary to our hypothesis, that is, that the GNT results generated by AI chatbots would be more accurate due to the use of customised prompts, DeepL was the system that presented the highest number of sentences translated into a completely gender-neutral style (11 sentences). The results may be attributed to the fact that the source sentences were written in a neutral and "institutional" style. It would be worthwhile to conduct a study to verify if different types of texts, not originally written with GNL, would generate different outcomes, potentially with a higher use of the generic masculine. By looking at the sentences generated in a gender-neutral style, it may be safe to assume that the strategies used by DeepL essentially involve using impersonal or collective nouns. The phrases “Cada representante/visitante/participante” and “Uma pessoa”, for example, avoid gender markings and make sentences impersonal.

Ref.	Translation	T1	T2	T3
4	Os planos para terceirizar cerca de 19 serviços não prosseguiram no ritmo previsto, pois há escassez significativa de mão de obra.	1	1	1
5	Os planos para terceirizar cerca de 19 serviços não prosseguiram no ritmo previsto, pois há uma escassez significativa de pessoal.	1	1	1
15	Serão responsáveis pela gestão dos procedimentos de contratação pública.	1	1	1
16	Alguém pode não saber seu número fiscal.	1; 6	1; 6	1; 6
19	Cada visitante deverá apresentar o seu documento de identificação para obter um passe de visitante temporário.	1	1	1
21	Cada participante deverá certificar-se de não deixar os seus documentos confidenciais ao sair da sala.	1	**1**	**1**
22	Cada participante deverá certificar-se de não deixar os seus documentos confidenciais ao sair da sala.	1	1	1
25	Uma pessoa deve ter residido continuamente no país durante 5 anos antes de solicitar residência permanente.	1	1	1
27	Cada oficial de ação deve enviar seu documento ao cartório para distribuição.	1; 6	1; 6	**1; 6**
28	Os documentos devem ser enviados ao cartório para distribuição.	1	1	1

Table 5 - Google Translate translations assessed as GNT by T1, T2 and T3.

Similar to the results obtained with DeepL, the number of sentences with GNT did not vary significantly. GT yielded 10 outcomes in GNL. As for the strategies used, no significant differences were observed, and it often resorted to impersonal and collective nouns (e.g. 'Uma pessoa'), indefinite pronouns (e.g. 'Alguém'), and a few occurrences of the passive voice.

5.3.2. ChatGPT and Google Bard

Ref.	Source	Translation	T1	T2	T3
1	GPT 1	Cada Representante Permanente deve apresentar as suas credenciais ao Protocolo.	1	1	1
1	GPT 2	Cada Representante Permanente deve submeter os seus credenciais ao Protocolo.	1	1	1
1	GPT 3	Cada Representante Permanente deve submeter suas credenciais ao Protocolo.	1; 6	1; 6	1; 6
4	GPT 1	Os planos para terceirizar cerca de 19 serviços não avançaram no ritmo previsto, devido a significativas escassezes de mão de obra.	1	1	1
4	GPT 2	Os planos de externalizar cerca de 19 serviços não avançaram no ritmo previsto, devido a significativas carências de recursos humanos.	1	1	1
4	GPT 3	Os planos de externalizar cerca de 19 serviços não avançaram no ritmo esperado, uma vez que existem importantes carências de recursos humanos.	1	**1**	**1**
5	GPT 1	Os planos para terceirizar cerca de 19 serviços não avançaram no ritmo previsto, devido a significativas carências de pessoal.	1	1	1
5	GPT 2	Os planos de externalizar cerca de 19 serviços não avançaram no ritmo previsto, devido a significativas carências de pessoal.	1	**1**	**1**
5	GPT 3	Os planos de externalizar cerca de 19 serviços não avançaram no ritmo esperado, uma vez que existem importantes carências de pessoal.	1	1	1
14	GPT 3	Em caso de interesse, por favor, envie um e-mail.	1	1	1
16	GPT 1	Alguém pode não saber o seu número de contribuinte.	1	1	1
16	GPT 2	Alguém pode não saber o seu número de contribuinte.	1	1	1
16	GPT 3	Alguém pode não saber o seu número de contribuinte.	1	1	1
19	GPT 2	Cada visitante deve apresentar o seu documento de identificação para obter um passe temporário de visitante.	1	1	1
21	GPT 1	Cada participante deve garantir que não deixe os seus documentos classificados ao sair da sala.	1	1	1
21	GPT 2	Cada participante deve garantir que não deixa para trás os seus documentos classificados ao sair da sala.	1	1	1
21	GPT 3	Cada participante deve garantir que não deixa os seus documentos classificados para trás ao sair da sala.	1	1	1
22	GPT 1	Cada participante deve garantir que não deixe os seus documentos classificados ao sair da sala.	1	1	1
22	GPT 2	Cada participante deve garantir que não deixa para trás os seus documentos classificados ao sair da sala.	1	1	1
22	GPT 3	Cada participante deve garantir que não deixa os seus documentos classificados para trás ao sair da sala.	1	1	1
25	GPT 1	Uma pessoa deve ter residido continuamente no país por 5 anos antes de solicitar residência permanente.	1	**1**	**1**
25	GPT 2	Uma pessoa deve ter residido continuamente no país durante 5 anos antes de solicitar residência permanente.	1	1	1
25	GPT 3	Uma pessoa deve ter residido continuamente no país por 5 anos antes de solicitar a residência permanente.	1	1	1; 6
27	GPT 1	Cada responsável pela ação deve enviar o seu documento ao arquivo para distribuição.	1	1	1
27	GPT 2	Cada responsável pela ação deve enviar o seu documento para o registo para distribuição.	1	**1**	**1**
27	GPT 3	Cada responsável pela ação deve enviar o seu documento para o arquivo para distribuição.	1	1	1
28	GPT 1	Os documentos devem ser enviados ao arquivo para distribuição.	**1**	**1**	**1**
28	GPT 2	Os documentos devem ser enviados para o registo para distribuição.	1	1	1
28	GPT 3	Os documentos devem ser enviados para o arquivo para distribuição.	1	1	1

Table 6 - ChatGPT translations assessed as GNT by T1, T2 and T3.

Between the ChatGPT translation version GPT1 and versions GPT2 & 3, only one more sentence was translated following the gender neutrality criterion: GPT1 produced 9 results, while GPT2 & 3 yielded 10 GNL outcomes. However, the sentence that showed improvement between versions GPT2 & 3 varied because gender neutrality was inconsistent across versions. Thus, in GPT1 & 2, the translation of sentence 19 changed from "Todo visitante" (the generic masculine) in GPT1 to "Cada visitante" (neutral) in GPT2; one would expect the customised prompt in GPT3 to replicate or produce a corresponding result. Yet, GPT3 mistakenly retrieved the generic masculine form "Todo o visitante". Sentence 14 shows a significant improvement in GTP3 compared to the earlier versions. While GPT1 & 2 resorted to the generic masculine ("Se estiver interessado"), GPT3 yielded an effectively gender-neutral translation by using the impersonal expression "Em caso de interesse".

From Table 6, we can observe that, in the three sets of translations provided by ChatGPT, the system tended to apply impersonal and collective nouns and phrases (e.g.

"pessoa", "cada visitante") as well as to the use of impersonal and/or passive structures in order to produce gender-neutral outcomes[14].

Ref.	Source	Translation	T1	T2	T3
4	BD 1	Os planos de terceirização de alguns 19 serviços não avançaram no ritmo previsto, pois há deficiências significativas de recursos humanos.	1	1	1
4	BD 2	Os planos de terceirização de alguns 19 serviços não seguiram o ritmo previsto, pois há uma escassez significativa de pessoal.	1	1	1
4	BD 3	Os planos de terceirização de alguns 19 serviços não têm avançado no ritmo esperado, devido a uma escassez significativa de mão de obra.	1	**1**	**1**
5	BD 1	Os planos de terceirização de alguns 19 serviços não avançaram no ritmo previsto, pois há deficiências significativas de pessoal.	1	1	1
5	BD 3	Os planos de terceirização de alguns 19 serviços não avançaram como esperado devido à escassez de pessoal.	1	1	1
6	BD 3	Antes de submeter o seu documento, envie-o ao ponto focal para revisão. O ponto focal irá devolver-lhe o documento com comentários.	1	1	1
9	BD 3	Uma pessoa que apresentar uma queixa e não estiver satisfeita com a decisão do conselho pode solicitar uma nova audiência.	1	1	1
15	BD 3	Serão responsáveis pela gestão de procedimentos de contratação pública.	1	1	1
16	BD 1	Alguém pode não saber seu número de contribuinte.	1; 6	1; 6	1; 6
16	BD 2	Alguém pode não saber seu número de contribuinte.	1; 6	1; 6	1; 6
16	BD 3	Poderá haver pessoas que não conheçam o seu número de identificação fiscal.	1	1	1
21	BD 1	Cada participante deve se certificar de não deixar seus documentos classificados para trás ao sair da sala.	1; 6	1; 6	1; 6
21	BD 2	Cada participante deve garantir que não deixe seus documentos classificados para trás ao sair da sala.	1	**1; 6**	1; 6
22	BD 1	Cada participante deve se certificar de não deixar seus documentos classificados para trás ao sair da sala.	1; 6	1; 6	1; 6
22	BD 2	Cada participante deve garantir que não deixe seus documentos confidenciais para trás ao sair da sala.	1; 6	1; 6	1; 6
25	BD 1	Uma pessoa deve ter residido continuamente no país por 5 anos antes de solicitar a residência permanente.	1	1	1; 6
25	BD 2	Uma pessoa deve residir continuamente no país por 5 anos antes de solicitar a residência permanente.	1	1	1; 6
25	BD 3	Para se candidatar à residência permanente, é necessário residir continuamente no país por 5 anos.	1	1	1; 6
27	BD 1	Cada oficial de ação deve enviar seu documento ao registro para distribuição.	1; 6	1; 6	1; 6
27	BD 2	Cada oficial de ação deve enviar seu documento ao registro para distribuição.	1; 6	1; 6	**1; 6**
28	BD 1	Documentos devem ser enviados ao registro para distribuição.	1; 6	1	1; 6
28	BD 2	Os documentos devem ser enviados ao registro para distribuição.	1; 6	1	1; 6
28	BD 3	Os documentos devem ser enviados ao registo para distribuição.	1	1	1

Table 7 - Google Bard translations assessed as GNT by T1, T2 and T3.

Google Bard usually presents a range of alternative responses to prompts, including comments and explanations. For the purposes of this study, we always considered the first proposal. We observed a very small difference in the total number of neutrally translated results among the translation outcomes resulting from the three Google Bard prompt versions: BD1 produced eight neutral results; BD2, only seven; and BD3, also eight. However, although they differ by only one sentence overall, the use of GNL from one version to the other was not consistent: GNL outcomes produced by one version were no longer gender-neutral in the next. For example, in BD1 & 3, the word "pessoal" is used as a neutral solution for the English word "staffing", while BD2 translated the same word with the generic masculine "funcionários".

[14] Although the 3 model prompts clearly indicated that the sentences were to be translated into PT-PT, this criterion was followed in GPT1 & 2, whereas, in GPT3, 3 of the gender-neutral translations were rendered in PT-BR.

Sentences 21, 22 and 23 are noteworthy because they complied with the gender neutrality criterion in BD1 & 2, but, in BD3, which was expected to produce better results for being the most detailed prompt, the three sentences were no longer gender-neutral. The English phrases “Each participant” (21 and 22) and “Each action officer” (23) were, at first, correctly neutralised in Portuguese as “Cada participante” and “Cada oficial de ação”, respectively. However, BD3 generated the male-biased translations “todos os participantes” and “todos os agentes de ação”. Sentences 6, 9 and 15, in fact, showed improvement in BD3.

The prompt version in BD3 seems to have efficiently improved the sentences above. Be that as it may, as in the case of ChatGPT, we could not identify a pattern or provide any justification to explain why BD1 & 2 prompt versions generated gender-neutral outcomes but not BD3, which produced generic masculine structures despite being the most customised prompt, thus expected to yield better results.

It should also be noted that, out of the eight translation proposals, BD3 was the only version that provided a gender-neutral solution for sentence 9 (“A complainant who is not satisfied with the board's decision can ask for a rehearing”). This sentence proved to be a challenge for all the other versions, which used the generic masculine, whereas BD3 achieved the desired gender neutrality with the expression "uma pessoa que apresentar uma queixa".

In fact, a closer look at all the gender-neutral solutions yielded by Google Bard reveals that, in the vast majority of cases, the system resorted to words in Portuguese which carry no gender marking, such as “pessoal”, “mão-de-obra”, “alguém” or “cada pessoa/oficial/participante”. One aspect of the results produced by Google Bard merits particular attention: the number of results in PT-BR dropped substantially from six (in BD1 & 2) to just one (in BD3).

5.4. Summary

A close analysis of the sets of sentences considered to have been translated in a gender-neutral style by the assessment team revealed that, out of the 38 sentences, only four met the GNT criteria in all eight translation versions (see Table 8 below).

Ref.	Source	Translation
4	E1	*Plans to outsource some 19 services have not proceeded at the anticipated pace, as there are significant manpower shortages.*
4	GT	Os planos para terceirizar cerca de 19 serviços não prosseguiram no ritmo previsto, pois há escassez significativa de mão de obra.
4	DL	Os planos de externalização de alguns serviços não avançaram ao ritmo previsto, uma vez que se regista uma escassez significativa de mão de obra.
4	GPT 1	Os planos para terceirizar cerca de 19 serviços não avançaram no ritmo previsto, devido a significativas escassezes de mão de obra.
4	GPT 2	Os planos de externalizar cerca de 19 serviços não avançaram no ritmo previsto, devido a significativas carências de recursos humanos.
4	GPT 3	Os planos de externalizar cerca de 19 serviços não avançaram no ritmo esperado, uma vez que existem importantes carências de recursos humanos.
4	BD 1	Os planos de terceirização de alguns 19 serviços não avançaram no ritmo previsto, pois há deficiências significativas de recursos humanos.
4	BD 2	Os planos de terceirização de alguns 19 serviços não seguiram o ritmo previsto, pois há uma escassez significativa de pessoal.
4	BD 3	Os planos de terceirização de alguns 19 serviços não têm avançado no ritmo esperado, devido a uma escassez significativa de mão de obra.
16	E3	*Someone may not know their tax number.*
16	GT	Alguém pode não saber seu número fiscal.
16	DL	Uma pessoa pode não saber o seu número de contribuinte.
16	GPT 1	Alguém pode não saber o seu número de contribuinte.
16	GPT 2	Alguém pode não saber o seu número de contribuinte.
16	GPT 3	Alguém pode não saber o seu número de contribuinte.
16	BD 1	Alguém pode não saber seu número de contribuinte.
16	BD 2	Alguém pode não saber seu número de contribuinte.
16	BD 3	Poderá haver pessoas que não conheçam o seu número de identificação fiscal.
25	E4	*A person must have resided continuously in the country for 5 years before applying for permanent residence.*
25	GT	Uma pessoa deve ter residido continuamente no país durante 5 anos antes de solicitar residência permanente.
25	DL	Para requerer a residência permanente, é necessário ter residido continuamente no país durante 5 anos.
25	GPT 1	Uma pessoa deve ter residido continuamente no país por 5 anos antes de solicitar residência permanente.
25	GPT 2	Uma pessoa deve ter residido continuamente no país durante 5 anos antes de solicitar residência permanente.
25	GPT 3	Uma pessoa deve ter residido continuamente no país por 5 anos antes de solicitar a residência permanente.
25	BD 1	Uma pessoa deve ter residido continuamente no país por 5 anos antes de solicitar a residência permanente.
25	BD 2	Uma pessoa deve residir continuamente no país por 5 anos antes de solicitar a residência permanente.
25	BD 3	Para se candidatar à residência permanente, é necessário residir continuamente no país por 5 anos.
28	E4	*Documents must be sent to the registry for distribution.*
28	GT	Os documentos devem ser enviados ao cartório para distribuição.
28	DL	Os documentos devem ser enviados ao registo para distribuição.
28	GPT 1	Os documentos devem ser enviados ao arquivo para distribuição.
28	GPT 2	Os documentos devem ser enviados para o registo para distribuição.
28	GPT 3	Os documentos devem ser enviados para o arquivo para distribuição.
28	BD 1	Documentos devem ser enviados ao registro para distribuição.
28	BD 2	Os documentos devem ser enviados ao registro para distribuição.
28	BD 3	Os documentos devem ser enviados ao registo para distribuição.

Table 8 - Sentences in GNT in the eight versions.

By closely examining the strategies (see Table 1[15]) used by the MT and AI systems to produce GNT (Table 9 below), we can see that a rather limited number of strategies were employed. As a matter of fact, except for one instance where the strategy used was repetition (F), and another in which the imperative form (H) was the structure of choice, for the most part, both MT systems and AI platforms resorted to strategies C, D and E (see Table 1). Strategy C clearly emerges as the most common one.

[15] Strategies identified in Table 1: A - Epicene synonyms; B - Pluralisation/Singularisation (towards generic or epicene forms); C - Relative and indefinite pronouns; D - Collective and role nouns; E - Omission; F - Repetition; G - Passive voice; H - Imperative forms; I - Impersonal forms.

	STRATEGY								
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
DeepL	--	--	6	2	2	--	1	--	--
GT	--	--	6	2	1	--	1	--	--
GPT	--	--	19	6	--	--	3	1	--
BARD	--	--	12	6	1	1	3	--	--

Table 9 - Strategies used by MT and AI systems according to neutralisation strategies defined in Table 1.

Sentence 4 stands out as it is not gender-neutral in English due to the choice of the word "manpower", yet each of the eight results produced a GNT. On the other hand, for sentence 5, which is practically identical to 4, except for being indeed gender-neutral, the BD2 prompt version generated a generic masculine word ("funcionário"). This result is curious, considering the prompt used was exactly the same as for sentence 4.

Understanding the sentences that all systems failed to translate in a gender-neutral style is just as crucial as examining the successful cases. The list below shows the 16 sentences that none of the eight translation outcomes managed to translate in GNL. Since gender markers are widely present in various grammatical categories in PT-PT, translating these sentences with GNL would require substantial changes and more “creative” solutions, which the MT and AI systems seemingly could not carry out.

Ref.	Source	Sentence
2	E1	Professor Smith and Professor Jones will attend the luncheon.
3	E1	Guests are cordially invited to attend with their partners
10	E1	Substitute judges must certify that they have familiarized themselves with the record of the proceedings.
11	E1	The author of a communication must have direct and reliable evidence of the situation being described.
12	E2	An applicant may submit their complaint.
13	E2	A manager may apply for a transfer, provided that permission has been granted by that manager's institution
17	E3	Officials shall carry out their duties.
18	E3	An official's salary is dependent on length of service.
20	E4	All visitors must present their IDs/passes to obtain a temporary visitor's pass.
29	E4	The arrival sheet for newcomers must be completed by every new staff member by the end of the month.
33	E4	I will ask my assistant to make the necessary arrangements.
34	E4	The family allowance shall be granted to staff members whose spouse, within the meaning of the Personnel Regulations, has an overall income (gross income less compulsory social and/or pension contributions) lower than 50% of the basic monthly salary of the scale of the duty station of a Z1/1 grade staff member, plus the basic amount of the allowance.
35	E4	Eligibility shall commence when the staff member and their spouse have established a family unit at the duty station. It shall end when the family unit is dissolved or when the spouse ceases to actually and habitually live with the staff member at the duty station.
36	E4	The COMMITTEE considered that 1 September was not the best time to hold elections because a) it had been noted over the years that the Committee Secretary always had a very heavy workload at that time of year and would be unable to prepare the documents in time and b) many parents had to take leave that day to take their children to school on the first day of the school year;
37	E4	The current Committee Secretary confirmed that she would prefer to hold the elections at a later date that month;
38	E4	All judges must certify that they have familiarized themselves with the process.

Table 10 - Sentences translated with non-gender-neutral language in all translation versions.

6. Conclusion, limitations and further research

First of all, it should be noted that the limited number of sample sentences used in this case study reveals inconsistencies in the use of GNL across the outputs generated by the AI systems despite the use of customised prompts. What seems to stand out from the data is that the strategies used by the systems varied minimally and that they cannot handle structures which may require more elaborate reformulations to produce texts with GNL. Besides, we ought to consider that the outcomes emerged exclusively from source sentences already with GNL.

Even though some recent studies have shown that strong prompts can unlock greater potential from ChatGPT, this did not seem to apply when using gender-neutral language in translations from English into Portuguese.

Without customised prompts, ChatGPT and Google Bard demonstrated a level of quality similar to that of DeepL and Google Translate regarding the use of GNL; likewise, the results yielded by the customised prompts were not substantially different. The attempt at fine-tuning the prompts proved to be a challenge, and further investigation is required into how to best balance the amount of information, its specificity, and the examples provided. However, crafting entirely gender-neutral sentences in Portuguese often requires resorting to rather unnatural structures, which would not be a native speaker's first choice if they were not

asked to use a totally gender-neutral language. Ultimately, it is up to content producers, copywriters, translation clients and other stakeholders to decide when complete neutrality is necessary or appropriate or whether it is more adequate to resort to structures with the traditional use of the generic masculine. This case study does not advocate for either approach but rather aims at contributing to analysing the performance of MT and AI systems when, for whatever reason, a totally GNT is a mandatory requirement.

The findings from this case study have led to two additional research projects, which will be reported on in the near future. The first of these upcoming studies aims to understand how human translators approach gender-neutral language in the translations of the sentences which the MT and AI systems in this study failed to translate using GNL. Furthermore, the change from Google Bard to Gemini makes it relevant to research whether the results achieved by Gemini would differ.

To delve deeper into the subject of GNT, further investigations should focus on understanding whether post-editing of MT or whether 'translating from scratch' would produce better outcomes. In the future, it would be extremely useful to conduct similar studies testing the use of AI chatbots with customised prompts for translating different text types and genres. Additionally, it would be worthwhile to verify if different types of texts not originally written with GNL, such as those using the generic masculine, would result in different outcomes.

References

Bentivogli, L.; Savoldi, B.; Negri, M.; Di Gangi, M. A.; Cattoni, R.; Turchi, M. 2020. Gender in Danger? Evaluating Speech Translation Technology on the MuST-SHE Corpus. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 6923–6933. https://doi.org/10.18653/v1/2020.acl-main.619.

Cho, W. I.; Kim, J. W.; Kim, M.; Kim, N. S. 2019a. On Measuring Gender Bias in Translation of Gender-neutral Pronouns. https://github.com/nolongerprejudice/tgbi.

Cho, W. I.; Kim, J. W.; Kim, S. M.; Kim, N. S. 2019b. On Measuring Gender Bias in Translation of Gender-neutral Pronouns. Proceedings of the First Workshop on Gender Bias in Natural Language Processing. Florence: Association for Computational Linguistics, 173–181. https://doi.org/10.18653/v1/W19-3824.

Comunicação e Linguagem Inclusiva. 2018. IGEN - Fórum Organizações Para a Igualdade. https://forumigen.cite.gov.pt/documentos/oferta-formativa-igen/t4-comunicacao-e-linguagem/t4_sara-falcao-casaca.pdf.

Dinu, G.; Mathur, P.; Federico, M.; Al-Onaizan, Y. 2019. Training Neural Machine Translation to Apply Terminology Constraints. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 3063–3068. https://doi.org/10.18653/v1/P19-1294.

Friedman, B.; Nissenbaum, H. 1996. Bias in computer systems. ACM Transactions on Information Systems, 14(3): 330–347. https://doi.org/10.1145/230538.230561.

Gao, Y.; Wang, R.; Hou, F. 2023. How to Design Translation Prompts for ChatGPT: An Empirical Study. http://arxiv.org/abs/2304.02182.

Guidelines on gender-inclusive language. 2020. North Atlantic Treaty Organisation. https://www.nato.int/nato_static_fl2014/assets/pictures/images_mfu/2021/5/pdf/210514-GIL-Manual_en.pdf.

Guidelines on gender-inclusive language in English. 2018. United Nations. https://www.un.org/en/gender-inclusive-language/guidelines.shtml.

Inclusive communication in the GSC. 2018. General Secretariat, Council of the European Union. https://www.consilium.europa.eu/media/35446/en_brochure-inclusive-communication-in-the-gsc.pdf.

Jiao, W.; Wang, W.; Huang, J.; Wang, X.; Shi, S.; Tu, Z. 2023. Is ChatGPT a good translator? Yes, with GPT-4 as the engine. arXiv. https://doi.org/10.48550/arXiv.2301.08745

Kocmi, T.; Limisiewicz, T.; Stanovsky, G. 2020. Gender Coreference and Bias Evaluation at WMT 2020. Proceedings of the Fifth Conference on Machine Translation. Online: Association for Computational Linguistics, 357–364.

Kostikova, A.; Daems, J.; Lazarov, T. 2023. How adaptive is adaptive machine translation, really? A gender-neutral language use case. Proceedings of the First Workshop on Gender-Inclusive Translation Technologies. Tilburg: Open Press TiU, 95-97.

Manual de linguagem inclusiva. 2021. Comissão para a Cidadania e a Igualdade de Género (CIG). Portugal. Conselho Económico e Social. https://www.cig.gov.pt/wp-content/uploads/2021/08/12-Manual-de-Linguagem-Inclusiva-CES.pdf.

Nord, C. 1997. Defining translation functions. The translation brief as a guideline for the trainee translation. Ilha Do Desterro 33: 41–55.

Nord, C. 2018. Translating as a Purposeful Activity (2nd Edition). Routledge. https://doi.org/10.4324/9781351189354.

Parlamento Europeu. 2018. Linguagem neutra do ponto de vista do género no Parlamento Europeu. Retrieved from https://www.europarl.europa.eu/cmsdata/288151/GNL_Guidelines_PT-original.pdf.

Piergentili, A.; Fucci, D.; Savoldi, B.; Bentivogli, L.; Negri, M. 2023. Gender Neutralization for an Inclusive Machine Translation: from Theoretical Foundations to Open Challenges. Proceedings of the First Workshop on Gender-Inclusive Translation Technologies. Tampere: European Association for Machine Translation, 71–83. http://arxiv.org/abs/2301.10075.

Prates, M. O. R.; Avelar, P. H.; Lamb, L. C. 2020. Assessing gender bias in machine translation: a case study with Google Translate. Neural Computing and Applications 32(10): 6363–6381. https://doi.org/10.1007/s00521-019-04144-6.

Pym, A. 1993. Christiane Nord. Text Analysis in Translation. Theory, Method, and Didactic Application of a Model for Translation-Oriented Text Analysis. Translated from the German by Christiane Nord and Penelope Sparrow. Amsterdam/Atlanta GA, Rodopi, 1991, 250 p. ISBN : 90-5183-311-3. TTR 6(2): 184-190. https://doi.org/10.7202/037160ar.

Reiss, K.; Vermeer, H. J. 2014. Towards a General Theory of Translational Action. Routledge. https://doi.org/10.4324/9781315759715.

Rescigno, A.; Vanmassenhove, E.; Monti, J.; Way, A. 2020. A Case Study of Natural Gender Phenomena in Translation. A Comparison of Google Translate, Bing Microsoft Translator and DeepL for English to Italian, French and Spanish. In Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020 (359–364). Accademia University Press. https://doi.org/10.4000/books.aaccademia.8844.

Saunders, D.; Byrne, B. 2020. Reducing Gender Bias in Neural Machine Translation as a Domain Adaptation Problem. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Florence: Association for Computational Linguistics, 7724–7736. https://doi.org/10.18653/v1/2020.acl-main.690.

Savoldi, B.; Gaido, M.; Bentivogli, L.; Negri, M.; Turchi, M. 2021. Gender Bias in Machine Translation. Transactions of the Association for Computational Linguistics 9: 845–874. https://doi.org/10.1162/tacl_a_00401.

Stanovsky, G.; Smith, N. A.; Zettlemoyer, L. 2019. Evaluating Gender Bias in Machine Translation. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 1679–1684. https://doi.org/10.18653/v1/P19-1164.

Sun, T.; Gaut, A.; Tang, S.; Huang, Y.; ElSherief, M.; Zhao, J.; Mirza, D.; Belding, E.; Chang, K.-W.; Wang, W. Y. 2019. Mitigating Gender Bias in Natural Language Processing: Literature Review. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 1630–1640. https://doi.org/10.18653/v1/P19-1159.

Sun, T.; Webster, K.; Shah, A.; Wang, W. Y.; Johnson, M. 2021. They, Them, Theirs: Rewriting with Gender-Neutral English. http://arxiv.org/abs/2102.06788.

Vanmassenhove, E.; Hardmeier, C.; Way, A. 2019. Getting Gender Right in Neural Machine Translation. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 3003–3008. https://doi.org/10.18653/v1/D18-1334.

Vanmassenhove, E.; Emmery, C.; Shterionov, D. 2021. NeuTral Rewriter: A Rule-Based and Neural Approach to Automatic Rewriting into Gender Neutral Alternatives. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana: Association for Computational Linguistics, 8940–8948. https://doi.org/10.18653/v1/2021.emnlp-main.704.

Yamada, M. 2023. Optimizing Machine Translation through Prompt Engineering: An Investigation into ChatGPT's Customizability. Proceedings of Machine Translation Summit XIX, Vol. 2: Users Track. Macau: Asia-Pacific Association for Machine Translation, 195–204. http://arxiv.org/abs/2308.01391.