

Occupational gender bias in ungendered languages an LLMs: Comparing Hungarian and Chinese

Anonymous ACL submission

Abstract

This paper examines occupational gender bias and stereotypes, in a cross-linguistic setting. We analyze ratings of 50 job titles collected from speakers of two languages without grammatical gender markings: Hungarian and Mandarin Chinese. Participants were instructed to rate how typical it is for a certain job to be done by men or women, according to their own perceptions. Our results show that in both languages the occupational nouns carry societal biases, despite the fact that the job titles themselves have no grammatical gender markings. We analyze the ratings by participant gender and perform intra-linguistic and cross-linguistic comparisons, highlighting key differences and offering insights that range from peculiarities in word formation to broader cross-cultural generalizations. Additionally, we also compared the human raters' responses with that of a few popular generative AI engines. Interestingly, the biases exhibited by Large Language Models (LLMs) were found to be even stronger than those shown by our human participants.

1 Introduction

— Check [Kaukonen et al. \(2025\)](#).

1.1 Background

...

— Hungarian has no grammatical gender, and most occupations are used in gender-neutral terms.

— Chinese only marks gender when writing the 3rd person singular pronoun – 他 *tā* ‘he’ / 她 *tā* ‘she’ – but that too is a relatively recent invention, going back to the May Fourth Movement of 1919 ([Bi, 2013](#)), and similarly to Hungarian, most occupations are unmarked for gender.

— [Gabor](#): Describe word formation for job titles in Hungarian, why most jobs are unmarked for gender, and when they are not. Add cultural notes? See Materials.

— [Wenhui](#): Describe word formation of job titles in Chinese, why job titles are unmarked for gender, what does adding 女 ‘woman’ and 男 ‘man’ “do”, and any related cultural aspects:

In Chinese, occupational titles are primarily formed through compound word formation, with their semantic core focused on describing the work content, function, or domain rather than directly denoting the gender of the practitioner. Regarding the morphological patterns of Chinese occupational compounds, scholars have proposed various classification methods ([Packard, 2000](#); [Fu, 2014](#)). Among these, the ‘head + quasi-suffix’ derivational pattern represents one of the most predominant models ([Fu, 2014](#)). These quasi-suffixes typically denote specific identities or categories of people, functionally analogous to suffixes like ‘-er’, ‘-ian’ or ‘-ist’ in English, but retain a degree of semantic substance without complete grammaticalization ([Fu, 2014](#)). Examples include 师 *shī* in 教师 *jīaoshī* ‘teacher’ and 工程师 *gōngchéngshī* ‘engineer’, 家 *jiā* in 科学家 *kēxuéjiā* ‘scientist’ and 画家 *huàjiā* ‘painter’, and 员 in 服务员 *fúwùyuán* ‘waiter/waitress’. While these morphemes possess independent lexical meanings, such as 师 denoting ‘military division’ and 家 meaning ‘home/residence’, their quasi-suffixal derivation shows in the formation of occupational nouns. Another significant pattern is the ‘verb-object’ (V-O) compound ([Packard, 2000](#)), which directly describes the core action of the work and its object. For instance, in 编剧 *biānjù* ‘playwright/screenwriter’, 编 (*verb*: ‘to organize; create’) governs 剧 (*noun*: ‘drama; script’) as its object. Similar structures include 司机 *sījī* ‘driver’ (司 *v.*: ‘to operate; control’ + 机 *n.*: ‘machine/vehicle’) and 保安 *bǎoān* ‘security guard’ (保 *v.*: ‘to protect’ + 安 *n.*, ‘safety’). Furthermore, Chinese occupational titles incorporate loanwords, such as 模特 *mó tè*, from French *modèle* or English *model*, and 博客 *bō kè*, from English *blogger*. Upon entering Chinese, these terms

typically conform to native morphological habits and usage rules, and their forms likewise carry no grammatical gender information.

Chinese is generally acknowledged as a language that lacks grammatical gender from a structural perspective (Li and Thompson, 1989). In contrast to languages with mandatory noun gender systems such as French, German, or Spanish, Chinese nouns, adjectives, and articles lack distinctions of masculine, feminine, or neuter gender. Thus, certain scholars contend that the Chinese linguistic system is fundamentally gender-neutral and may not inherently reflect gender classification (Li and Thompson, 1989; Packard, 2000).

Nonetheless, sociolinguistic research argues that language use, particularly at the pragmatic level, is significantly shaped by socio-cultural attitudes (Labov, 1972). Many occupations are associated with relatively strong gender stereotypes (Sun, 1997; Su et al., 2021). For instance, 护士 *hùshì* translates to ‘nurses’ and 保姆 *bǎomǔ* to ‘babysitter’, both of which are predominantly female occupations, whereas 警察 *jīngchá* means ‘police officers’ and 高管 *gāoguǎn* refers to ‘executive’, which are predominantly male roles. This societal perception results in pragmatic asymmetry: although occupational terms are grammatically gender-neutral, speakers or language users frequently form default gender assumptions in communication contexts based on societal stereotypes. Consequently, when a practitioner’s gender diverges from the conventional stereotype associated with their profession, or when particular circumstances require explicit gender identification, speakers often precede occupational terms with gender markers 男 *nán* for ‘male’ or 女 *nǚ* for ‘female’. This results in Chinese occupational phrases such as 男护士 *nánhùshì* for ‘male nurse’, 女警察 *nǚjīngchá* for ‘female police officer’, 女高管 *nǚgāoguǎn* for ‘female executive’, and 男保姆 *nánbǎomǔ* for ‘male babysitter’. This practice of incorporating gender markers demonstrates a dual effect (Sun, 1997; Gustafsson Sendén et al., 2015; Su et al., 2021). On the one hand, it emphasises the transcendence of gender constraints by individuals, thereby increasing the visibility of specific groups, particularly those in non-traditional gender roles. Conversely, scholars suggest that this gender-focused terminology may, particularly in contradictory contexts, reinforce prevailing gender stereotypes (Sun, 1997). Conversely, scholars suggest that this gender-focused terminology may, particularly in contradictory contexts, rein-

force prevailing gender stereotypes (Sun, 1997; Gustafsson Sendén et al., 2015). For instance, under societal cognition specific to China, the aforementioned occupation title 高管 *gāoguǎn* ‘executive’ subconsciously defaults to a male referent. Highlighting 女高管 *nǚgāoguǎn* or ‘female executive’ designates the individual as an anomaly and may inadvertently maintain the gender stereotype in China that associates the term ‘executive’ predominantly with males, showing an intrinsic uniqueness or necessity for explicit distinction when women assume this position.

In summary, Chinese occupational titles are inherently gender-neutral at the grammatical level due to the linguistic features, specifically the lack of grammatical gender and their primary function of solely describing the work itself. The addition of gender modifiers such as 男 *nán* ‘male’ or 女 *nǚ* ‘female’ operates as a pragmatic device rather than a grammatical requirement. This strategy primarily signals socio-cognitive markedness by highlighting deviations from occupation-gender stereotypes (i.e., denoting non-prototypical associations), as empirically demonstrated in Chinese corpora where terms like 女总统 *nǚzǒngtǒng*, ‘female president’ occur more frequently than their male-marked counterparts in leadership roles (Su et al., 2021; Farris, 1988). Concurrently, it fulfills referential specificity in contexts demanding explicit gender identification, such as legal depositions or medical documentation where deixis resolution is pragmatically necessary (Hellinger and Bußmann, 2003; Stahlberg et al., 2011).

...

We are interested in these unmarked words, as they do not inherently possess gender bias – certainly not grammatically – but according to our expectations they will be rated according to the prevailing societal stereotypes nonetheless.

...

2 Experimental setup

For both languages we designed a simple survey-based experiment in which participants were asked to rate job titles on a 7-point Likert scale. Participants were instructed to make decisions on how likely is an occupation to be pursued by men or by women, according to their own perception.¹ First,

¹While this study focuses on people who identify or are identified as either male or female, we acknowledge the presence of non-binary people in the workforce.

we will introduce the Hungarian experiment, then the Chinese one, and then move on to the results.

2.1 Hungarian

2.1.1 Participants

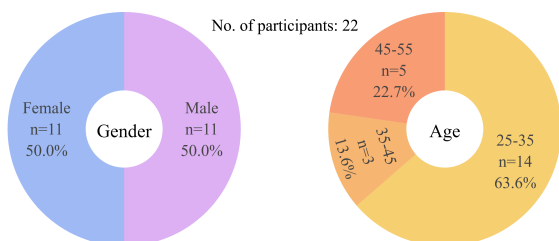


Figure 1: Demographics of the Hungarian participants.

A total of 24 native Hungarian speakers filled our questionnaire, and after validating the responses (reviewing attention checks and manually checking for anomalies) 2 were rejected. Participants were recruited online using the participant recruitment platform Prolific, with screeners set for location (Hungary) and first language (Hungarian); raters were compensated for their time with a small monetary reward. In the end, the Hungarian ratings dataset had 22 participants (n=11 female, n=11 male), with ages ranges of 25-35 (n=14), 35-45 (n=3), and 45-55 (n=5). See Figure 1 for the distribution.

2.1.2 Materials

The Hungarian survey contained 44 items and 6 attention-check items, each a commonly occurring job title in Hungary, such as: *modell* ‘model’ or *katona* ‘soldier’, in no particular order. The attention-check items were removed from the final analysis, these were *pincérnő* ‘waitress’, *titkárnő* ‘secretary (female)’, *tanárnő* ‘teacher (female)’, *takarítónő* ‘cleaning lady’, *ápolónő* ‘nurse (female)’, and *házvezetőnő* ‘housekeeper (female)’. These words are compounds, they explicitly determine the gender of the worker by appending *-nő* ‘woman’ to the base noun. If participants paid attention, all these items should be rated ‘completely female’ (3). Participants who rated any of these lower than 2, or rated them lower than 3 more than once were rejected.

The 6 words above have counterparts without the *nő* ‘woman’ element, i.e. *pincér* ‘waiter’, *titkár* ‘secretary’, *tanár* ‘teacher’, *takarító* ‘cleaner’, *ápoló* ‘nurse’, and *házvezető* ‘housekeeper’, all included in our survey. These words are unmarked for gender, but they do not explicitly refer to men.

The perceived gender bias of these occupations – if any – is due to social factors, not grammatical ones.

Maybe some of this below should go to introduction...

On the surface level, the unmarked form is the base noun, and the feminine-marked form is created by appending *-nő* ‘woman’ to it, which is a regular way of creating female occupational titles in Hungarian. However, in reality this does not equal to a male-female pair, as the unmarked form is not necessarily ‘masculine’. We can observe 3 types in the pragmatic usage of occupational nouns when it comes to the unmarked–marked pairing and its implications for the gendering of the unmarked nouns.

1) **Both forms are common.** Frequently occurring word-pairs in Hungarian would be for example *énekes* ‘singer’ – *énekesnő* ‘female singer’ (not in our dataset). In cases where both versions are well established – i.e., both occur with a relatively high frequency in a balanced corpus – the unmarked word seems to carry some male bias, as the frequent use of a feminine form indicates a need and/or custom for differentiation. We wanted to test if raters perceived this bias or not. In this example, the absolute and relative frequencies (occurrence per a million words) of the two lemmatized nouns in the Hungarian National Corpus (HNC) are 1441/9,4001 for *énekes* and 748/4,8795 for *énekesnő* (Váradi, 2002; Oravecz et al., 2014); the frequency difference here is roughly half (51.9%).

The stark deviations in frequencies for marked–unmarked word pairs such as the above are not an indicator for a strong gender bias – we can assume that both men and women singers would be equally represented in the Hungarian corpus – but reflect that in general, the unmarked forms are used for either males or females when talking about one’s occupation. The female-marked forms are used when there is an explicit intention to specify the gender of the individual, when it is otherwise not known from context (or proper names).²

2) **Only unmarked form is common.** However, there are many cases where the unmarked form is the only one generally used for both genders. Take for example *ügyész* ‘prosecutor’ (8451/55,1287) vs. *ügyésznő* ‘female prosecutor’ (56/0,3653), or

²For example, the sentence *Anyukám tanár.* ‘My mom is a teacher.’ uses the unmarked form, firstly because we want to communicate her job, and secondly because it is obvious from the subject (mother) that she is a woman.

fodrász ‘hairstylist’ (944/6,1580) vs. *fodrásznő* ‘female hairstylist’ (35/0,2283); the deviations in frequency here are over multiple orders of magnitude. In these cases, the unmarked form is the default word to describe anyone practicing the occupation, and appending ‘-nő’ ‘woman’ to it – although possible – would render it unusual, and a bit awkward; but still not as uncanny/forced as Modern English *singress* would be.³

3) **Marked form is common.** Furthermore, there are cases, where the female-marked version ending in *-nő* is so ubiquitous, that it is the unmarked version that will sound a bit unusual, such as *háztartó* ‘housekeeper’ (10/0.0652) vs. *háztartónő* ‘female housekeeper’ (92/0,6001), or, to a small extent *takarító* ‘cleaner’ (169/1.1024) vs. *takarítónő* ‘female cleaner’ (392/2,5571). In short, we are interested in these unmarked words and how people perceive them.

The full list of 44 items is as follows: *modell*, *katona* ‘soldier’, *kórboncnok* ‘pathologist’, *vezérigazgató* ‘CEO’, *menedzser*, ‘manager’ *nővér*, ‘nurse’ *szakács* ‘chef’, *felhasználó* ‘server’, *könyvelő* ‘accountant’, *professzor* ‘professor’, *építész* ‘architect’, *tudós* ‘scientist’, *ápoló* ‘nurse’, *pénztáros* ‘cashier’, *bíró* ‘judge’, *munkás* ‘worker’, *vízimentő* ‘lifeguard’, *jegyárús* ‘ticket seller’, *tűzoltó* ‘firefighter’, *mérnök* ‘engineer’, *rendező* ‘director’, *takarító* ‘cleaner’, *HR-es* ‘HR specialist’, *háztartó* ‘housekeeper’, *légiutas-kísérő* ‘flight attendant’, *pincér* ‘waiter’, *orvos* ‘doctor’, *fodrász* ‘hairstylist’, *földműves* ‘farmer’, *gondozó* ‘caregiver’, *bolti eladó* ‘shop assistant’, *kertész* ‘gardener’, *titkár* ‘secretary’, *PR munkatárs* ‘PR officer’, *dietetikus* ‘dietitian’, *tanár* ‘teacher’, *rendőr* ‘police officer’, *pilóta* ‘pilot’, *recepció* ‘receptionist’, *biztonsági őr* ‘security guard’, *ügyész* ‘prosecutor’, *kozmetikus* ‘beautician’, *programozó* ‘programmer’, *diák* ‘student’.

We have included *diák* ‘student’ out of curiosity. Although being a student is not a job per se, but it is beyond doubt the only truly gender-neutral “occupation” there is, since it is mandatory for every child to go to school (both in Hungary and in China). We wanted to see if there would be any bias regarding this word, especially that Hungarian has a female-marked form for it – *diáklány* ‘girl student’, appending *-lány* ‘girl’ to the base noun – and could be considered to belong to the 1st type with

³Although English had a form *singress*, from Middle English *syngeresse*, it is now obsolete.

a potential male bias when considered in contrast.

2.1.3 Procedure

Hungarian participants were instructed to rate each word on a 7-point Likert scale, ranging from *completely male* to *completely female*, the ratings were then converted to numerical values from -3 to +3. The scale presented in both questionnaires followed the same logic, with 0 in the middle, hence the choices were completely male (-3); mostly male (-2); somewhat male (-1); neutral/equal (0); somewhat female (+1); mostly female (+2); completely female (+3). The exact wording of the main question of the Hungarian survey was: “Ön szerint a foglalkozás tipikusan férfi foglalkozás, vagy tipikusan női foglalkozás?” (Is this occupation typically a man’s occupation or a woman’s occupation?), cf.

2.

Kérjük jelölje minden sorban, hogy az adott szó mennyire jelöl tipikusan férfi vagy női foglalkozást. *

	Teljesen férfi	Nagyrészt férfi	Inkább férfi	Semleges/egyenlő	Inkább női	Nagyrészt női	Teljesen női
modell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
katona	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 2: A sample of the the Hungarian survey layout.

The survey was created in Microsoft Forms, and after a brief welcome message and instructions the words were presented in a simple list format, each word with a corresponding rating scale next to it, with no context. Time limit was not set, but the survey was designed to take around 5 minutes, and participants took 4 minutes 25 seconds on average to finish.

2.2 Chinese survey

2.2.1 Participants

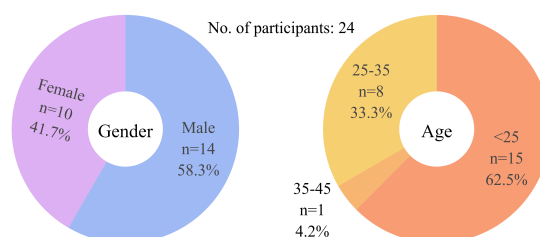


Figure 3: Demographics of the Chinese participants.

The Chinese survey was completed by 30 native Mandarin Chinese speakers, 6 of which were rejected after failing attention checks.

Wenhui: Please describe the platform used to recruit participants/how it was distributed:

Participants were recruited from **vertical** communities on WeChat in China. They were required to have been born and raised in mainland China, with Mandarin as their primary language and the primary language spoken at home. Additionally, they must have resided in China continuously for the past five years. This screening mechanism facilitated a prolonged social and cultural immersion experience.

Participants were paid a small fee for completing the questionnaire. The 24 accepted participants (n=14 male, n=10 female) were mostly university students, aged <25 (n=15), 25-35 (n=8), or 35-45 (n=1). See Figure 3 for the distribution.

2.2.2 Materials

The Chinese survey, too, contained 44 items of commonly occurring job titles in Mandarin Chinese (Simplified), with 6 attention checks, to ensure participant engagement and data quality, also in randomized order. The attention checks were 妈妈 ‘mother’, 爸爸 ‘father’, 女作家 ‘female writer’, 男作家 ‘male writer’, 女画家 ‘female painter’, 男画家 ‘male painter’. These words are inherently feminine or masculine in meaning, or explicitly determine gender by prepending 女 ‘woman’ and 男 ‘man’, helping to filter out inattentive responses. Participants who failed to rate these with the highest scores of either *completely male* or *completely female* were rejected.

Here is the full list of Chinese items: 警察 ‘police’, 秘书 ‘secretary’, 教授 ‘professor’, 护士 ‘nurse’, 高管 ‘manager’, 教师 ‘teacher’, 前台 ‘receptionist’, 工人 ‘worker’, 幼师 ‘kindergarten teacher’, 模特 ‘model’, 护工 ‘caregiver’, 保姆 ‘nanny’, 会计 ‘accountant’, 工程师 ‘engineer’, 保洁 ‘cleaner’, 法官 ‘judge’, 导购员 ‘shop assistant’, 美容师 ‘beautician’, 服务员 ‘waiter’, 乘务员 ‘flight attendant’⁴, 理发师 ‘hairstylist’, 空服员 ‘flight attendant’⁵, 售票员 ‘ticket seller’, 厨师 ‘chef’, 营养师 ‘nutritionist’, 家政员 ‘housekeeper’, 收银员 ‘cashier’, 医生 ‘doctor’, 法医 ‘pathologist’, 程序员 ‘programmer’, 保安 ‘security guard’, 导演 ‘director’, 军人 ‘soldier’, 董事长 ‘CEO’, 农民 ‘farmer’, 学生 ‘student’, 园丁 ‘gardener’, 飞行员 ‘pilot’, 人事 ‘HR personnel’, 消防员 ‘firefighter’, 科学家 ‘scientist’, 检察官 ‘prosecutor’, 救生员 ‘lifeguard’, 建筑师 ‘architect’.

⁴Also the attendant/crew on high-speed trains.

⁵Less frequent word; only on airplane.

2.2.3 Procedure

Chinese participants, too, were instructed to rate each word on a 7-point Likert scale, ranging from *completely male* to *completely female* with *neutral/equal* in the middle, and the ratings were then converted to numerical values as explained above. The exact wording of the questions in the Chinese survey was: “对于occupation这个职业, 您认为通常担任该职业的男女性别比例是多少?” (What do you think is the ratio of men to women in occupation?). The survey was created in Microsoft Forms with essentially identical instructions, participants saw each word highlighted in the question above, with no other context, and were presented with the scale directly below.

对于警察这个职业, 您认为通常担任该职业的男女性别比例是多少? *

完全由女性担任	绝大多数由女性担任	多数由女性担任	男女比例大致相当	多数由男性担任	绝大多数由男性担任	完全由男性担任
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

对于秘书这个职业, 您认为通常担任该职业的男女性别比例是多少? *

完全由女性担任	绝大多数由女性担任	多数由女性担任	男女比例大致相当	多数由男性担任	绝大多数由男性担任	完全由男性担任
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 4: A sample of the the Chinese survey layout.

2.3 Methods

For analyzing the data, we used the *scipy* library in Python, and performed one-sample *t*-tests to determine the significance of differences in the mean ratings of occupations, and independent sample (two-sample) *t*-tests to analyze the differences between different groups, such as the ratings of male and female raters and the differences between the ratings of Hungarian and Chinese raters for the same occupations. The significance level was set to $p < 0.05$, and marginally significant items ($0.05 < p < 0.1$) were also highlighted.

3 Results & Analysis

For both languages, the majority of occupations were rated with a significant gender bias. In Hungarian, 36 out of 44 occupations showed significant bias, and in Chinese, 39 out of 40 were biased. This in itself is not surprising...

3.1 Hungarian

The Hungarian data was first analyzed using a one-sample *t*-test to determine which of the occupations showed a significant bias, measured against 0 (neutral/equal). The results showed that the majority

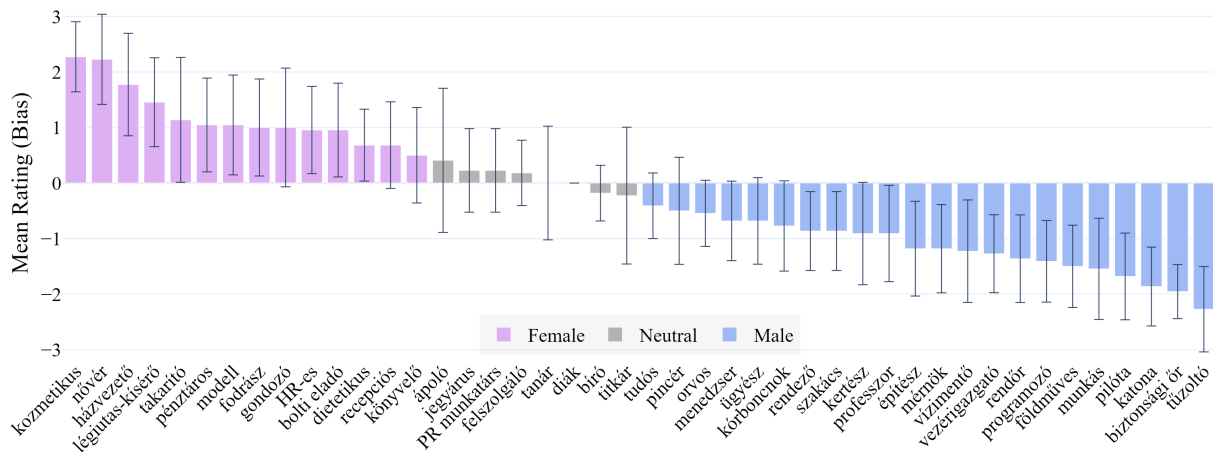


Figure 5: Mean ratings of occupational titles in Hungarian with standard deviations, significant gender bias highlighted – [explore the interactive plot](#).

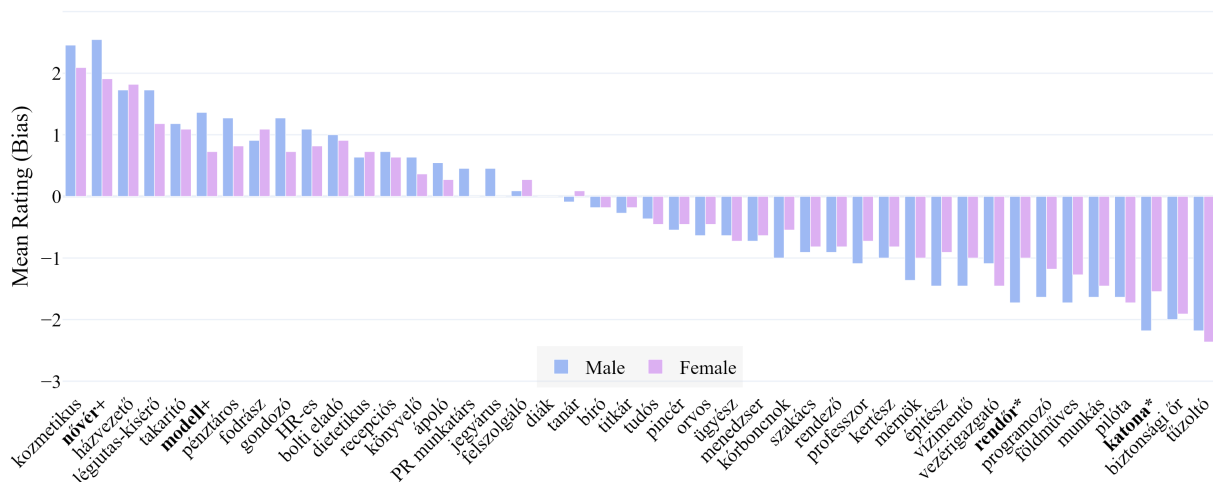


Figure 6: Mean ratings of occupational titles in Hungarian by gender, significant differences highlighted (significant*, and marginally significant+ in **bold**) – [explore the interactive plot](#).

of occupational titles – 36 out of 44 – were rated with a significant gender bias, with 14 showing female, and 22 showing male bias. See Figure 5 for a visualization of the mean ratings, with the gender biases highlighted.

3.1.1 Overall distribution of ratings

In general, occupations were rated according to expectations, following societal stereotypes and realities. Words with the highest female bias were *kozmetikus* ‘beautician’ (2.27), *nővér* ‘nurse’ (2.23), *házvezető* ‘housekeeper’ (1.77), *légiutas-kísérő* ‘flight attendant’ (1.45), and *takarító* ‘cleaner’ (1.14), while words with the highest male bias included *munkás* ‘worker’ (-1.55), *pilóta* ‘pilot’ (-1.68), *katona* ‘soldier’ (-1.86), *biztonsági őr* ‘security guard’ (-1.95), and *tűzoltó* ‘firefighter’ (-2.27).

Nővér ‘nurse’ (2.23) – is a bit special, as it liter-

ally means ‘sister’ and goes back to the time when nuns were the ones taking care of the sick; hence the word carries a strong female bias that is encoded in its literal meaning. Interestingly, it was not rated as an exclusively female job, probably because male nurses are now also common. The gender-neutral word *ápoló* ‘nurse’ for the same job was also tested, and it received a neutral rating of 0.41.

The 8 job titles that came back as not significantly biased were: *ápoló* ‘nurse’ (0.41), *jegyárús* ‘ticket seller’ (0.23), *PR munkatárs* ‘PR specialist’ (0.23), *felhasználó* ‘server’ (0.18), *diák* ‘student’ (0), *tanár* ‘teacher’ (0), *bíró* ‘judge’ (-0.18), and *titkár* ‘secretary’ (-0.23). It is worth noting that while *diák* ‘student’ was rated 0 by everyone, *tanár* ‘teacher’ had more individual variation in the ratings, leading to a higher standard deviation.

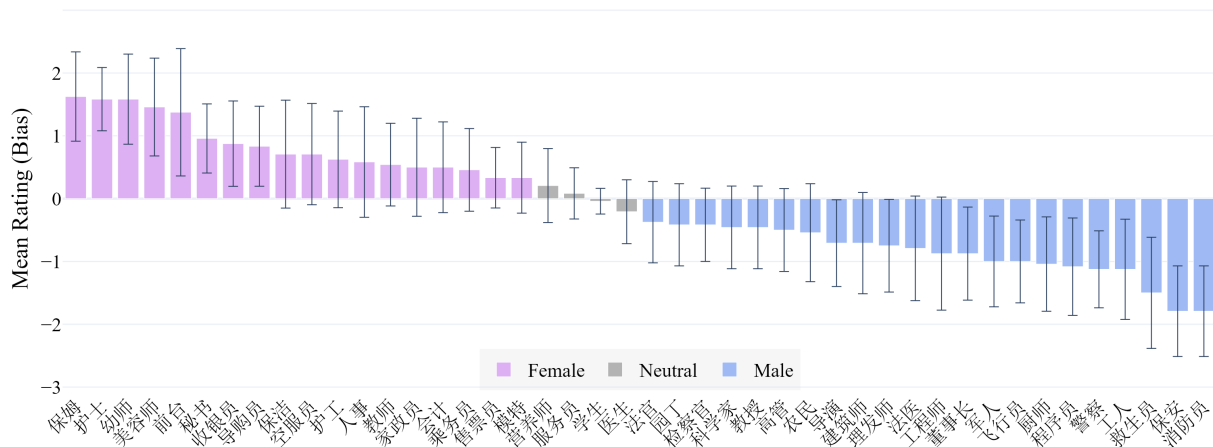


Figure 7: Mean ratings of occupational titles in Chinese with standard deviations, significant differences highlighted – [explore the interactive plot](#).

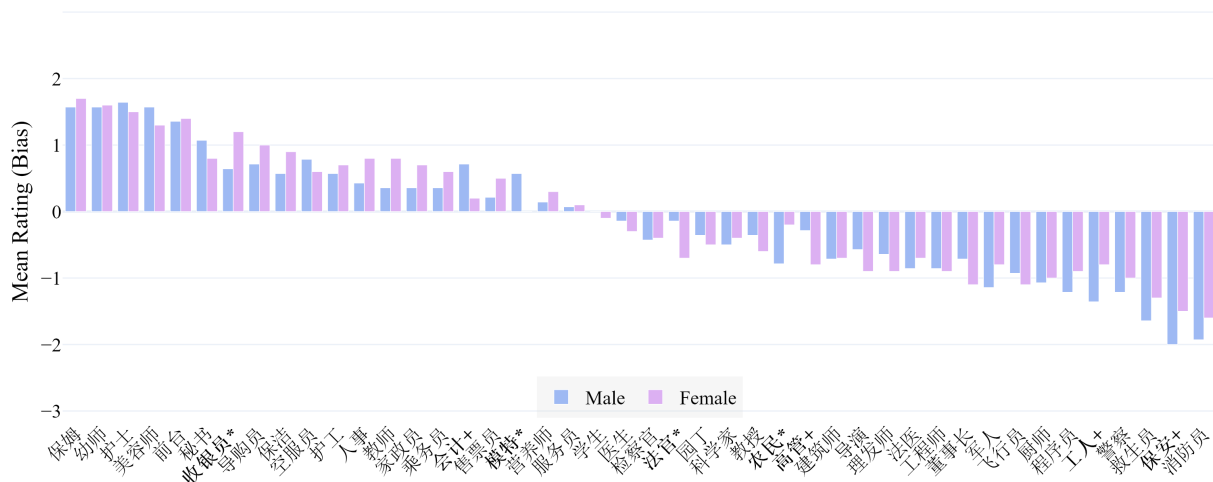


Figure 8: Mean ratings of occupational titles in Chinese by gender, significant differences highlighted (significant*, and marginally significant+ in **bold**) – [explore the interactive plot](#).

The strongest agreement were on *diák* ‘student’ (0, std=0), *biztonsági őr* ‘security guard’ (-1.95; std=0.4857), *bíró* ‘judge’ (-0.18; std=0.5011), *fel-szolgáló* ‘server’ (0.18; std=0.5885), *tudós* ‘scien-tist’ (-0.41; std=0.5903).

3.1.2 Intra-language gender differences in the Hungarian data

We also ran a two-sample *t*-test to compare the ratings of male and female participants for each occupation, and see if there was any discrepancies between the two groups. The only jobs that showed a significant difference was *rendőr* ‘po-lice officer’ (male=-1.73; female=-1.00) and *katona* ‘soldier’ (m=-2.18; f=-1.55), here the male biases were much higher by male raters. Two marginally significant different items were also found in *mod-ell* ‘model’ (m=1.36; f=0.73) and *nővér* ‘nurse’

(m=2.55; f=1.91) showing a similar trend. The results are summarized in Figure 6.

Furthermore, it seems like men’s ratings tended to have a greater absolute bias for both male- and female-coded jobs (see Figure 10 below).

3.2 Chinese

Similarly to Hungarian, we found that a majority of occupations in Chinese were also rated with sig-nificant gender bias. The results of the one-sample *t*-test showed that 39 out of 44 occupations were biased. The mean ratings are shown in Figure 7.

3.2.1 Overall distribution of ratings

Wenhui, can you help here with explaining some of our expectations for the results?:

The rating results largely aligned with our ex-pectations concerning occupational gender stereo-

types in Chinese society. Participants typically linked women to child-bearing, emotional labor, and service-oriented roles while associating men with high-intensity, high-risk physical labor, and order-maintaining roles.

In Chinese, the words with the highest female bias were 保姆 ‘domestic helper’ (1.63), 护士 ‘nurse’ (1.58), 幼师 ‘kindergarten teacher’ (1.58), 美容师 ‘beautician’ (1.46), and 前台 ‘receptionist’ (1.38). Words with the highest male bias were 警察 ‘police officer’ (-1.13), 工人 ‘worker’ (-1.13), 救生员 ‘lifeguard’ (-1.50), 保安 ‘security guard’ (-1.79), and 消防员 ‘firefighter’ (-1.79), which shows a relatively strong similarity to the Hungarian trends.

The 4 job titles that were not significantly biased were: 营养师 ‘dietitian’ (0.21), 服务员 ‘waiter/server’ (0.08), 学生 ‘student’ (-0.04), and 医生 ‘doctor’ (-0.21).

In Chinese too, there was strong agreement on 学生 ‘student’ (-0.04; std=0.2041) and 服务员 ‘waiter/server’ (0.08; std=0.4082), the remaining words in the top 5 jobs with the lowest standard deviation were 售票员 ‘ticket seller’ (0.33; std=0.4815), 护士 ‘nurse’ (1.58; std=0.5036), and 医生 ‘doctor’ (-0.21; std=0.5089).

3.2.2 Intra-language gender differences in the Chinese data

The two-sample *t*-test comparing the ratings of male vs. female participants showed that there were significant differences between what people think of a typical 收银员 ‘cashier’ ($m=0.64$; $f=1.20$), 模特 ‘model’ ($m=0.57$; $f=0$), 法官 ‘judge’ ($m=-0.14$; $f=0.70$), and 农民 ‘farmer’ ($m=-0.79$; $f=-0.20$). Marginally significant differences were found for 会计 ‘accountant’ ($m=0.71$; $f=0.20$), 高管 ‘manager’ ($m=-0.29$; $f=-0.80$), 工人 ‘worker’ ($m=-1.36$; $f=-0.80$), and 保安 ‘security guard’ ($m=-2.00$; $f=-1.50$). The results can be examined in Figure 8.

3.3 Cross-linguistic comparison

The wordlists of the two datasets were almost, but not exactly the same; by performing an inner join on the two lists, we could pair 42 items together according to their meanings. Then, using a two-sample *t*-test, we checked if there were significant differences between the ratings in the two languages. The results are summarized in Figure 9.

When comparing the two sets of ratings, the first noticeable trend is that in general, the two languages have similar biases for the same occupations.

Shared items on the extreme ends of the scale include for example ‘beautician’ and ‘nurse’, as well as ‘firefighter’ and ‘security guard’. Job titles that were neutral/unbiased in both datasets were ‘server’ (*felszolgáló*; 服务员), and ‘student’ (*diák*; 学生).

In terms of significant differences, we found 16 occupations that were rated differently in the two languages. These include ‘beautician’ (Hungarian=2.27 vs. Chinese=1.46), ‘housekeeper’ ($H=1.77$ vs. $C=0.50$), ‘receptionist’ ($H=0.68$ vs. $C=1.38$), ‘flight attendant’ ($H=1.45$ vs. $C=0.46$), ‘model’ ($H=1.05$ vs. $C=0.33$), ‘dietitian’ ($H=0.68$ vs. $C=0.21$), ‘secretary’ ($H=-0.23$ vs. $C=0.96$), ‘teacher’ ($H=0.00$ vs. $C=0.54$), ‘hairstylist’ ($H=1.00$ vs. $C=-0.75$), ‘waiter’ ($H=-0.50$ vs. $C=0.08$), ‘doctor’ ($H=-0.55$ vs. $C=-0.21$), ‘gardener’ ($H=-0.91$ vs. $C=-0.42$), ‘farmer’ ($H=-1.50$ vs. $C=-0.54$), ‘pilot’ ($H=-1.68$ vs. $C=-1.00$), and ‘soldier’ ($H=-1.86$ vs. $C=-1.00$), and ‘firefighter’ ($H=-2.27$ vs. $C=-1.79$). Marginally significant differences were found for ‘professor’, ‘architect’ and ‘CEO’.

— Wenhui, this is where you can try and discuss some of these significantly different words that could be interesting for readers. For example, you can help to explain that:

The occupational title ‘secretary’ elicits significantly different gender stereotypes among Hungarian and Chinese participants, which is a noteworthy observation. Hungarian participants perceive ‘titkár’ as a male-dominated profession (-0.23), whereas Chinese participants perceive it a predominantly female occupation (0.96). This discrepancy may arise from the interaction of lexical attributes and cultural context. In Hungarian, ‘titkár’ denotes both senior political officials and regular administrative employees. Conversely, the Chinese Occupational Classification Dictionary (Occupational Code: 3-01-02-02) delineates a 秘书 *mìshū* ‘secretary’ as an office service employee involved in clerical and meeting-related responsibilities. This definition may delay the manifestation of in societal perception and gender expectations, yet it largely corresponds with public expectations of the secretary profession. The semantic scope and societal expectations of identical occupational titles in Hungarian and Chinese illustrate how cultural background influences gender bias in vocabulary.

The gender stereotype linked to ‘housekeeper’ is predominantly female in both cultures; however, the extent of this bias varies considerably (Chinese:



Figure 9: Mean ratings of common occupational titles in Hungarian and Chinese (significant*, and marginally significant+ in **bold**) – [explore the interactive plot](#).

0.5 vs. Hungarian: 1.77). In the results of Chinese participants, it is pertinent to compare the terms 家政员 *jiāzhèngyuán* meaning ‘housekeeper’ and 保姆 *baómǔ* meaning ‘nanny’, which are analogous in meaning yet vary in register. While the term ‘housekeeper’ (0.5) demonstrates less gender bias than the colloquial term ‘nanny’ (1.63), it remains categorised as a female-dominated profession due to the traditional female-associated traits of household tasks and emotional management suggested by the central term 家政 *jiāzhèng* ‘housekeeping’. This domain-specific scoring disparity indicates that although the formal and informal characteristics of occupational terminology mitigate gender stereotypes, societal gender stereotypes regarding occupations may be intricately linked to the semantic implications of language. Particular terminology related to domestic or emotional labour is more prone to evoke conventional female gender bias perceptions, thus reinforcing the gendered classification of professions.

The disparities in gender perception between ‘farmer’ and ‘worker’ are particularly significant. Data from Chinese participants indicates that ‘farmer’ (-0.54) demonstrates a comparatively weaker male orientation than ‘worker’ (-1.13), whereas in Hungarian, the two are nearly identical (farmer: -1.50 vs. worker: -1.55). This discrepancy can be attributed to China’s societal and economic development trajectory. Throughout the era of collective farming in China (1950s–1970s), the work-point accounting system and production brigades promoted women’s involvement in agricultural labour, while slogans such as “Women hold

up half the sky” (妇女能顶半边天) enhanced the visibility of women in this sector (Jacka, 1997a,b). Despite the Household Contract Responsibility System established in the 1980s reinforcing the dominance of male household heads and consequently marginalising women’s land rights (Zhu, 2009), the significance of women in agricultural production has progressively become apparent [data is needed here to show the proportion of women in the Chinese agriculture industry]. During this period, township enterprises and factories employed female labour; however, they confined women to light industrial positions through technical qualification prerequisites (Liu, 2007). This led to a gendered labour division in which women in industry participated in unskilled roles, while men retained technical monopolies (Bossen, 2002). This historical process has resulted in the ongoing divergence of gender perceptions between the agricultural and industrial sectors. Despite contemporary ideals of occupational gender equality, gender stereotypes in lower-tier professions continue to demonstrate significant historical persistence.

— housekeeper is not exactly the same as 家政员 (housekeeper in Hungarian is equal to a maid, major tasks are cleaning and cooking. I heard that 家政员 is more serious.)

— secretary has more than one meaning (I don’t know Chinese, but Hungarian yes, one is a secretary of a political party (leadership position), other is a personal assistant (subordinate).)

— whatever else you can say about the significantly different words.

— the top 10 most different occupations in order are: ‘hairdresser’, ‘housekeeper’, ‘flight attendant’, ‘farmer’, ‘soldier’, ‘secretary’, ‘beautician’, ‘pilot’, ‘model’, and ‘receptionist’.

The most striking dissimilarity was the word for ‘hairdresser’ (*fodrász* vs. 理发师), which shows a strong female bias in Hungarian (1.00) but a definite male bias in Chinese (-0.75). We believe that this is a neat example for a cultural difference, in Hungary hairdressers are perceived to be predominantly female, while in China the profession is perceived to be more male-dominated. It is not difficult to find evidence for the latter from the press, even in English-language media.⁶

3.4 Discussion

The first obvious thing to notice when looking at Figure 9 is that the overall trends in gender bias are quite similar between the two languages, despite some notable differences in specific occupations. This suggests that stereotypes are comparable and – mostly – consistent across developing societies.

It would be interesting not only to explore the reasons behind this difference, but to compare actual figures on the gender distribution of hairdressers in the two countries, and see if the ratings reflect reality. (Is there any available workforce statistics on gender distribution per profession?) — many service jobs?

3.5 Gender bias by language

An interesting feature is that Hungarian raters tended to rate occupations with a stronger bias than Chinese speakers. Out of 42 shared occupations, 31 were rated with a higher bias in Hungarian, while only 10 were rated with a higher bias in Chinese (excluding the 1 item with equal value). This is a threefold difference, and warrants further exploration to fully explain.

Tie in Kaukonen et al. here

3.6 Gender bias by language users’ gender

An interesting divergence arose when comparing the two datasets and the differences between the ratings by gender. While in Hungarian, biases – both male and female – were stronger in the ratings of men, in Chinese women rated with a stronger bias on average, especially regarding female biases. You can compare the contrast in Figure 10.

⁶<https://www.chinadailyhk.com/hk/article/603100>

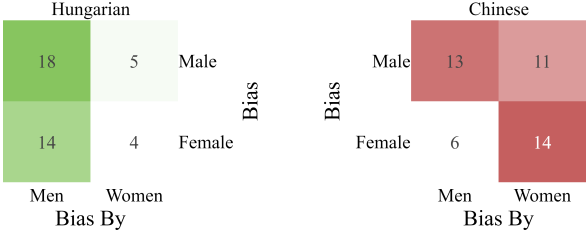


Figure 10: Confusion matrix of the Hungarian and Chinese ratings, showing the differences in the ratings of male and female participants.

4 Comparing the human ratings with AI-generated ratings

Gabor is still working on this part...

Now, we were also curious how the human ratings compare to the ratings of Large Language Models (LLMs) used in popular AI agents, and so we basically repeated the same experiment using AI instead of human raters. We prompted Copilot (+Think Deeper), ChatGPT (+Reason), Gemini (2.5 Flash), and Deepseek (Deepthink R1) to elicit a rating on the same job titles in both languages, using the same scale as the human raters. The instructions were in Hungarian and in Chinese, respectively, with a pretext: “You are participating in an experiment and your answers will help our research.”. We asked the AI agents 10 separate times, and took the mean as a baseline for the results for each job title. The mean AI ratings were then superimposed on the human raters’ plots to allow for a convenient comparison, you can see these in Figures 11 and 12.

The rationale behind this was to simulate how the general public and non-experts would turn to AI to do the job of human raters, and to show that these practices can be highly problematic, as the biases encoded in the LLMs are expected to be even stronger than those of the human raters. Mention real studies on this and cite. We think that the results perfectly illustrate the dangers of using AI agents instead of human raters, and only perpetuate social biases.

5 Conclusion

In conclusion, our study provides evidence for occupational gender bias in both Hungarian and Chinese contexts, using data from native speakers. We have demonstrated that gender stereotypes are present in job titles, even if the words themselves are unmarked for gender, and this is due to societal norms

of Reform, Cambridge Modern China Series, pages 120–142. Cambridge University Press, Cambridge.

Tamara Jacka. 1997b. [Domestic work](#). In *Women's Work in Rural China: Change and Continuity in an Era of Reform*, Cambridge Modern China Series, pages 101–119. Cambridge University Press, Cambridge.

Elisabeth Kaukonen, Polina Oskolskaia, Liina Lindström, and Raili Marling. 2025. [Gender, language and labour: Gender perception of Estonian and Russian occupational titles](#). *Frontiers in Communication*, 9.

William Labov. 1972. *Sociolinguistic Patterns*. University of Pennsylvania Press.

Charles N Li and Sandra A Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. University of California Press.

Jieyu Liu. 2007. *Gender and Work in Urban China: Women Workers of the Unlucky Generation*. Routledge.

Csaba Oravecz, Tamás Váradi, and Bálint Sass. 2014. The Hungarian Gigaword Corpus. In *LREC 2014*, volume 9, pages 1719–1723.

Jerome L. Packard. 2000. *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge University Press.

Dagmar Stahlberg, Friederike Braun, Lisa Irmen, and Sabine Sczesny. 2011. Representation of the sexes in language. In *Social communication*, pages 163–187. Psychology Press.

Qi Su, Pengyuan Liu, Wei Wei, Shucheng Zhu, and Chu-Ren Huang. 2021. [Occupational gender segregation and gendered language in a language without gender: Trends, variations, implications for social development in China](#). *Humanities and Social Sciences Communications*, 8(133):1–10.

J. R. Sun. 1997. 语言与性别 [*Language and Gender*]. Jiangsu Education Press.

Tamás Váradi. 2002. The Hungarian National Corpus. In *LREC 2002*, pages 385–389.

Ling Zhu. 2009. [Gender Inequality in the Land Tenure System of Rural China](#). In *China's Economy*, volume Volume 1 of *Series on Developing China ? Translated Research from China*, pages 21–36. WORLD SCIENTIFIC.