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

# Anonymous ACL submission

**001**

**002**

**003**

**004**

**005**

**006**

**007**

**008**

**009**

**010**

**011**

**012**

**013**

**014**

**015**

**016**

**017**

**018**

**019**

**020**

**021**

**022**

**023**

**024**

**025**

**026**

**027**

**028**

**029**

**030**

**031**

**032**

**033**

**034**

**035**

**036**

**037**

**038**

**039**

**040**

# 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 gram- matical gender markings: Hungarian and Man- darin 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 lan- guages the occupational nouns carry societal biases, despite the fact that the job titles them- selves 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 gen- eralizations. Additionally, we also compared the human raters’ responses with that of a few popular generative AI engines. Interestingly, the biases exhibited by Large Language Mod- els (LLMs) were found to be even stronger than those shown by our human participants.

# Introduction

* Check [Kaukonen et al.](#_bookmark18) ([2025](#_bookmark18)).

## 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](#_bookmark15), [2013](#_bookmark15)), and similarly to Hungarian, most occu- pations are unmarked for gender.

—Gabor: Describe word formation for job titles in Hungarian, why most jobs are unmarked for gen- der, and when they are not. Add cultural notes? See Materials.

—Wenhui: Describe word formation of job titles **041**

in Chinese, why job titles are unmarked for gender, **042**

what does adding 女 ‘woman’ and 男 ‘man’ “do”, **043**

and any related cultural aspects. **044**

… **045**

We are interested in these unmarked words, as **046**

they do not inherently possess gender bias – cer- **047**

tainly not grammatically – but according to our **048**

expectations they will be rated according to the **049**

prevailing societal stereotypes nonetheless. **050**

… **051**

# Experimental setup 052

For both languages we designed a simple survey- **053**

based experiment in which participants were asked **054**

to rate job titles on a 7-point Likert scale. Partic- **055**

ipants were instructed to make decisions on how **056**

likely is an occupation to be pursued by men or by **057**

women, according to their own perception.[1](#_bookmark1) First, **058**

we will introduce the Hungarian experiment, then **059**

the Chinese one, and then move on to the results. **060**

* 1. **Hungarian** **061**
     1. **Participants** **062**

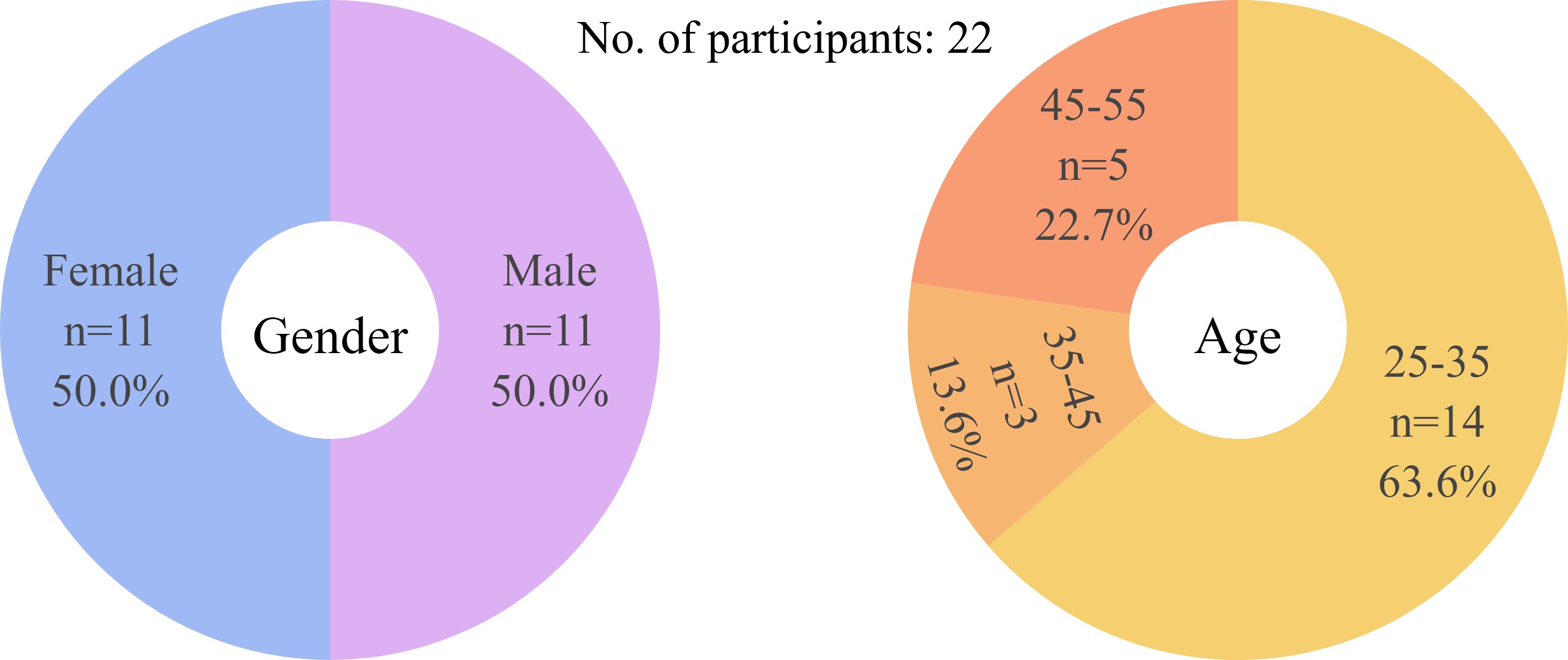


Figure 1: Demographics of the Hungarian participants.

A total of 24 native Hungarian speakers filled **063**

our questionnaire, and after validating the responses **064**

(reviewing attention checks and manually check- **065**

ing for anomalies) 2 were rejected. Participants **066**

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

**067**

**068**

**069**

**070**

**071**

**072**

**073**

**074**

**075**

**076**

**077**

**078**

**079**

**080**

**081**

**082**

**083**

**084**

**085**

**086**

**087**

**088**

**089**

**090**

**091**

**092**

**093**

**094**

**095**

**096**

**097**

**098**

**099**

**100**

**101**

**102**

**103**

**104**

**105**

**106**

**107**

**108**

**109**

**110**

**111**

**112**

**113**

**114**

**115**

**116**

**117**

were recruited online using the participant recruit- ment platform Prolific, with screeners set for lo- cation (Hungary) and first language (Hungarian); raters were compensated for their time with a small monetary reward. In the end, the Hungarian rat- ings 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](#_bookmark0) for the distri- bution.

## Materials

The Hungarian survey contained 44 items and 6 attention-check items, each a commonly occur- ring 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 (fe- male)’, and *házvezetőnő* ‘housekeeper (female)’. These words are compounds, they explicitly deter- mine the gender of the worker by appending *-nő* ‘woman’ to the base noun. If participants paid at- tention, 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 then 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 in- cluded 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 introduc- tion...:

On the surface level, the unmarked form is the base noun, and the feminine-marked form is cre- ated 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 occur- ing word-pairs in Hungarian would be for example *énekes* ‘singer’ – *énekesnő* ‘female singer’ (not in

our dataset). In cases where both versions are well **118**

established – i.e., both occur with a relatively high **119**

frequency in a balanced corpus – the unmarked **120**

word seems to carry some male bias, as the frequent **121**

use of a feminine form indicates a need and/or cus- **122**

tom for differentiation. We wanted to test if raters **123**

perceived this bias or not. In this example, the ab- **124**

solute and relative frequencies (occurrance per a **125**

million words) of the two lemmatized nouns in the **126**

Hungarian National Corpus (HNC) are 1441/9,4001 **127**

for *énekes* and 748/4,8795 for *énekesnő* ([Váradi](#_bookmark20), **128**

[2002](#_bookmark20); [Oravecz et al.](#_bookmark19), [2014](#_bookmark19)); the frequency differ- **129**

ence here is roughly half (51.9%). **130**

The stark deviations in frequencies for marked- **131**

unmarked word pairs such as the above are not an **132**

indicator for a strong gender bias – we can assume **133**

that both men and women singers would be equally **134**

represented in the Hungarian corpus – but reflect **135**

that in general, the unmarked forms are used for **136**

either males or females when talking about one’s **137**

occupation. The female-marked forms are used **138**

when there is an explicit intention to specify the **139**

gender of the individual, when it is otherwise not **140**

known from context (or proper names).[2](#_bookmark2) **141**

* + - 1. **Only unmarked form is common.** However, **142**

there are many cases where the unmarked form is **143**

the only one generally used for both genders. Take **144**

for example *ügyész* ‘prosecutor’ (8451/55,1287) **145**

vs. *ügyésznő* ‘female prosecutor’ (56/0,3653), or **146**

*fodrász* ‘hairdresser’ (944/6,1580) vs. *fodrásznő* **147**

‘female hairdresser’ (35/0,2283); the deviations in **148**

frequency here are over multiple orders of mag- **149**

nitude. In these cases, the unmarked form is the **150**

default word to describe anyone practicing the occu- **151**

pation, and appending *’-nő* ‘woman’ to it – although **152**

possible – would render it unusual, and a bit awk- **153**

ward; but still not as uncanny/forced as Modern **154**

English *singress* would be.[3](#_bookmark3) **155**

* + - 1. **Marked form is common.** Furthermore, **156**

there are cases, where the female-marked ver- **157**

sion ending in *-nő* is so ubiquitous, that it is the **158**

unmarked version that will sound a bit unusual, **159**

such as *házvezető* ‘housekeeper’ (10/0.0652) vs. **160**

*házvezetőnő* ‘female housekeeper’ (92/0,6001), or, **161**

to a small extent *takarító* ‘cleaner’ (169/1.1024) vs. **162**

*takarítónő* ‘female cleaner’ (392/2,5571). In short, **163**

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.

3Although English had a form *singeress*, from Middle En- glish *syngeresse*, it is now obsolete.

**164**

**165**

**166**

**167**

**168**

**169**

**170**

**171**

**172**

**173**

**174**

**175**

**176**

**177**

**178**

**179**

**180**

**181**

**182**

**183**

**184**

**185**

**186**

**187**

**188**

**189**

**190**

**191**

**192**

**193**

**194**

**195**

**196**

**197**

**198**

**199**

**200**

**201**

**202**

**203**

**204**

**205**

**206**

**207**

**208**

**209**

**210**

**211**

**212**

**213**

**214**

we are interested in these unmarked words and how people perceive them.

The full list of 44 items is as follows: *modell*, *ka- tona* ‘soldier’, *kórboncnok* ‘pathologist’, *vezérigaz- gató* ‘CEO’, *menedzser*, ‘manager’ *nővér*, ‘nurse’ *szakács* ‘chef’, *felszolgáló* ‘server’, *könyvelő* ‘ac- countant’, *professzor* ‘professor’, *építész* ‘archi- tect’, *tudós* ‘scientist’, *ápoló* ‘nurse’, *pénztáros* ‘cashier’, *bíró* ‘judge’, *munkás* ‘worker’, *vízi- mentő* ‘lifeguard’, *jegyárus* ‘ticket seller’, *tűzoltó* ‘firefighter’, *mérnök* ‘engineer’, *rendező* ‘direc- tor’, *takarító* ‘cleaner’, *HR-es* ‘HR specialist’, *házvezető* ‘housekeeper’, *légiutas-kísérő* ‘flight at- tendant’, *pincér* ‘waiter’, *orvos* ‘doctor’, *fodrász* ‘hairdresser’, *földműves* ‘farmer’, *gondozó* ‘care- giver’, *bolti eladó* ‘shop assistant’, *kertész* ‘gar- dener’, *titkár* ‘secretary’, *PR munkatárs* ‘PR offi- cer’, *dietetikus* ‘dietitian’, *tanár* ‘teacher’, *rendőr* ‘police officer’, *pilóta* ‘pilot’, *recepciós* ‘reception- ist’, *biztonsági őr* ‘security guard’, *ügyész* ‘prose- cutor’, *kozmetikus* ‘beautician’, *programozó* ‘pro- grammer’, *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 “oc- cupation” 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 would be any bias regarding this word, especially that Hungarian has a female-marked form for it – *diáklány* ‘girl stu- dent’, appending *-lány* ‘girl’ to the base noun – and could be considered to belong to the 1st type with a potential male bias when considered in contrast.

## Procedure

Hungarian participants were instructed to rate each word on a 7-point Likert scale, ranging from *com- pletely 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); some- what female (+1); mostly female (+2); completely female (+3). The exact wording of the main ques- tion of the Hungarian survey was: “Ön szerint a foglalkozás tipikusan férfi foglalkozás, vagy tipiku- san női foglalkozás?” (Is this occupation typically a man’s occupation or a woman’s occupation?), cf. [2](#_bookmark4).

The survey was created in Microsoft Forms, and after a brief welcome message and instructions the

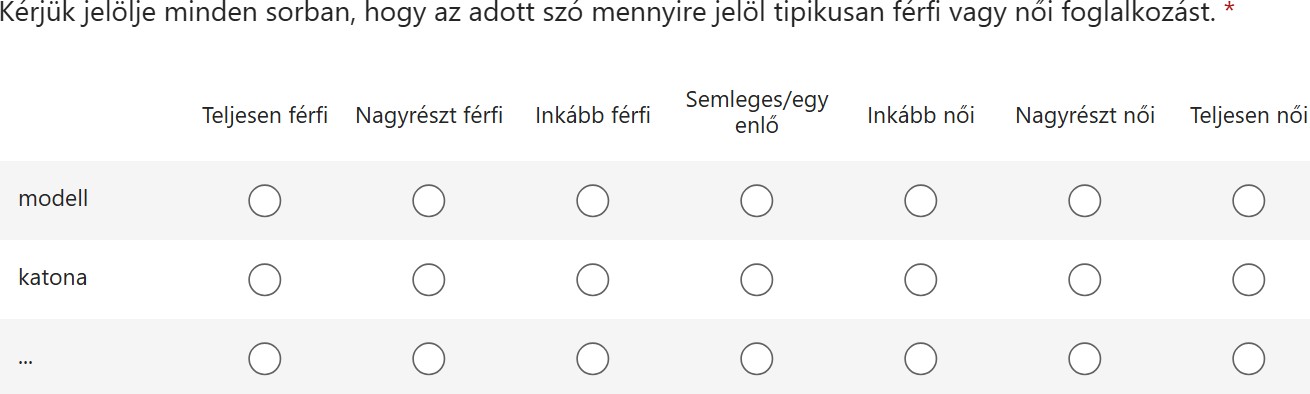


Figure 2: A sample of the the Hungarian survey layout.

words were presented in a simple list format, each **215**

word with a corresponding rating scale next to it, **216**

with no context. Time limit was not set, but the **217**

survey was designed to take around 5 minutes, and **218**

participants took 4 minutes 25 seconds on average **219**

to finish. **220**

* 1. **Chinese survey** **221**
     1. **Participants** **222**

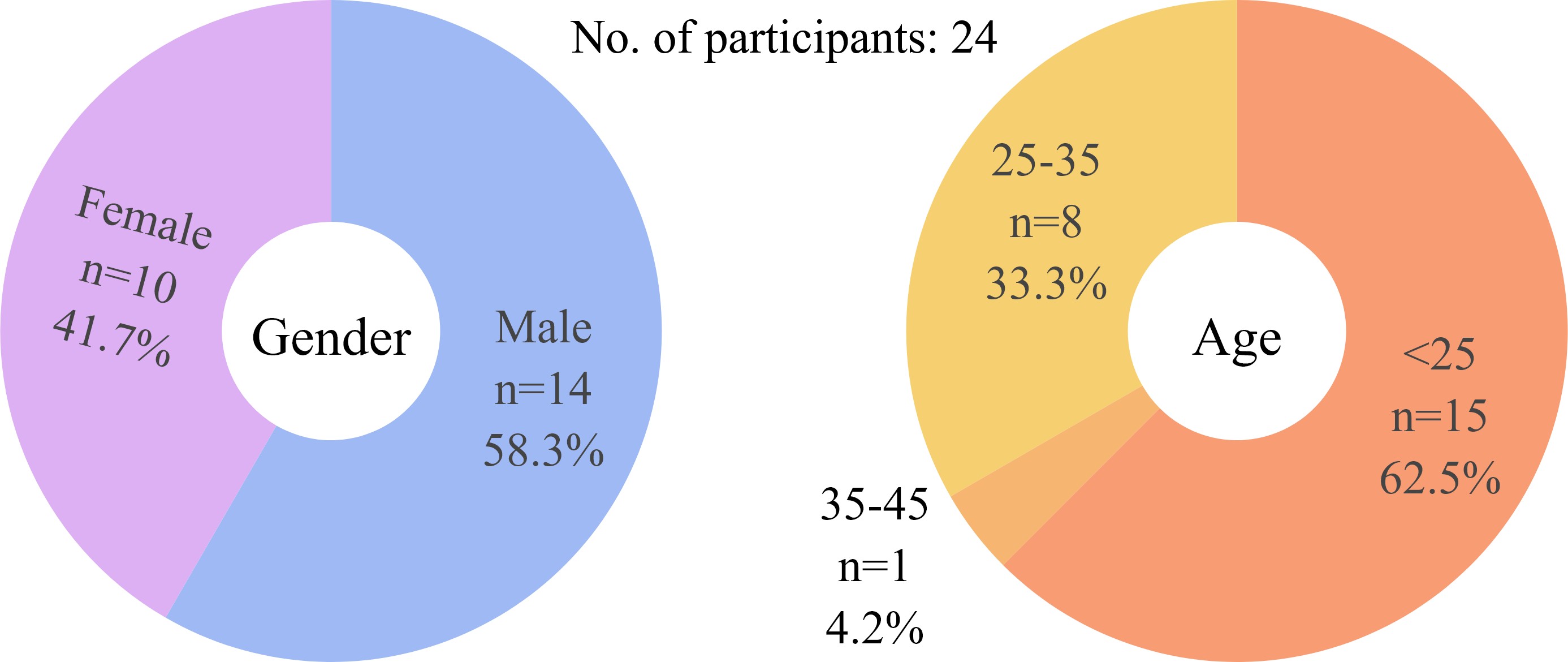


Figure 3: Demographics of the Chinese participants.

The Chinese survey was completed by 30 na- **223**

tive Mandarin Chinese speakers, 6 of which were **224**

rejected after failing attention checks. **225**

Wenhui: Please describe the platform used to **226**

recruit participants/how it was distributed. **227**

Participants were paid a small fee for completing **228**

the questionnaire. The 24 accepted participants **229**

(n=14 male, n=10 female) were mostly university **230**

students, aged <25 (n=15), 25-35 (n=8), or 35-45 **231**

(n=1). See Figure [3](#_bookmark5) for the distribution. **232**

* + 1. **Materials** **233**

The Chinese survey, too, contained 44 items of **234**

commonly occurring job titles in Mandarin Chinese **235**

(Simplified), with 6 attention checks, to ensure par- **236**

ticipant engagement and data quality, also in ran- **237**

domized order. The attention checks were 妈妈 **238**

‘mother’, 爸爸 ‘father’, 女作家 ‘female writer’, **239**

男作家 ‘male writer’, 女画家 ‘female painter’, **240**

男画家 ‘male painter’. These words are inherently **241**

feminine or masculine in meaning, or explicitly de- **242**

termine gender by prepending 女 ‘woman’ and 男 **243**

‘man’, helping to filter out inattentive responses. **244**

Participants who failed to rate these with the high- **245**

est scores of either *completely male* or *completely* **246**

**247**

**248**

**249**

**250**

**251**

**252**

**253**

**254**

**255**

**256**

**257**

**258**

**259**

**260**

**261**

**262**

**263**

**264**

**265**

**266**

**267**

**268**

**269**

**270**

**271**

**272**

**273**

**274**

**275**

**276**

**277**

**278**

**279**

**280**

**281**

**282**

**283**

**284**

**285**

*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 as- sistant’, 美容师 ‘beautician’, 服务员 ‘waiter’,乘务员 ‘flight attendant’[4](#_bookmark6), 理发师 ‘hairdresser’,空服员 ‘flight attendant’[5](#_bookmark7), 售票员 ‘ticket seller’,厨师 ‘chef’, 营养师 ‘nutritionist’, 家政员 ‘house- keeper’, 收银员 ‘cashier’, 医生 ‘doctor’, 法医 ‘pathologist’, 程序员 ‘programmer’, 保安 ‘secu- rity guard’, 导演 ‘director’, 军人 ‘soldier’, 董事长 ‘CEO’, 农民 ‘farmer’, 学生 ‘student’, 园丁 ‘gar- dener’, 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 ‘archi-

tect’.

## 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.

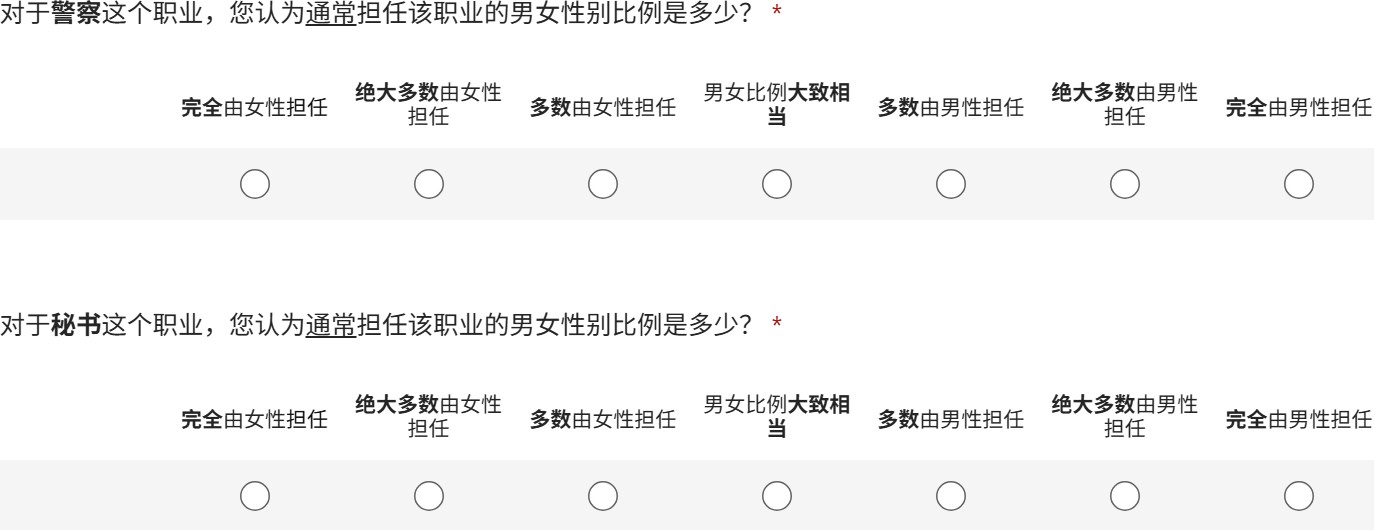


Figure 4: A sample of the the Chinese survey layout.

## Methods

For analyzing the data, we used the scipy library in Python, and performed one-sample *t*-tests to deter-

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

5Less frequent word; only on airplane.

mine the significance of differences in the mean rat- **286**

ings of occupations, and independent sample (two- **287**

sample) *t*-tests to analyze the differences between **288**

different groups, such as the ratings of male and fe- **289**

male raters and the differences between the ratings **290**

of Hungarian and Chinese raters for the same occu- **291**

pations. The significance level was set to *p <* 0*.*05, **292**

and marginally significant items (0.05 < *p <* 0*.*1) **293**

were also highlighted. **294**

# Results & Analysis 295

For both languages, the majority of occupations **296**

were rated with a significant gender bias. In Hun- **297**

garian, 36 out of 44 occupations showed significant **298**

bias, and in Chinese, 39 out of 40 were biased. This **299**

in itself is not surprising... **300**

* 1. **Hungarian** **301**

The Hungarian data was first analyzed using a one- **302**

sample *t*-test to determine which of the occupations **303**

showed a significant bias, measured against 0 (neu- **304**

tral/equal). The results showed that the majority **305**

of occupational titles – 36 out of 44 – were rated **306**

with a significant gender bias, with 14 showing fe- **307**

male, and 22 showing male bias. See Figure [5](#_bookmark8) for **308**

a visualization of the mean ratings, with the gender **309**

biases highlighted. **310**

## Overall distribution of ratings 311

In general, occupations were rated according to **312**

expectations, following societal stereotypes and re- **313**

alities. Words with the highest female bias were **314**

*kozmetikus* ‘beautician’ (2.27), *nővér* ‘nurse’ (2.23), **315**

*házvezető* ‘housekeeper’ (1.77), *légiutas-kísérő* **316**

‘flight attendant’ (1.45), and *takarító* ‘cleaner’ **317**

(1.14), while words with the highest male bias in- **318**

cluded *munkás* ‘worker’ (-1.55), *pilóta* ‘pilot’ (- **319**

1.68), *katona* ‘soldier’ (-1.86), *biztonsági őr* ‘secu- **320**

rity guard’ (-1.95), and *tűzoltó* ‘firefighter’ (-2.27). **321**

*Nővér* ‘nurse’ (2.23) – is a bit special, as it liter- **322**

ally means ‘sister’ and goes back to the time when **323**

nuns were the ones taking care of the sick; hence the **324**

word carries a strong female bias that is encoded in **325**

its literal meaning. Interestingly, it was not rated as **326**

an exclusively female job, probably because male **327**

nurses are now also common. The gender-neutral **328**

word *ápoló* ‘nurse’ for the same job was also tested, **329**

and it received a neutral rating of 0.41. **330**

The 8 job titles that came back as not signifi- **331**

cantly biased were: *ápoló* ‘nurse’ (0.41), *jegyárus* **332**

‘ticket seller’ (0.23),*PR munkatárs* ‘PR specialist’ **333**

(0.23), *felszolgáló* ‘server’ (0.18), *diák* ‘student’ (0), **334**

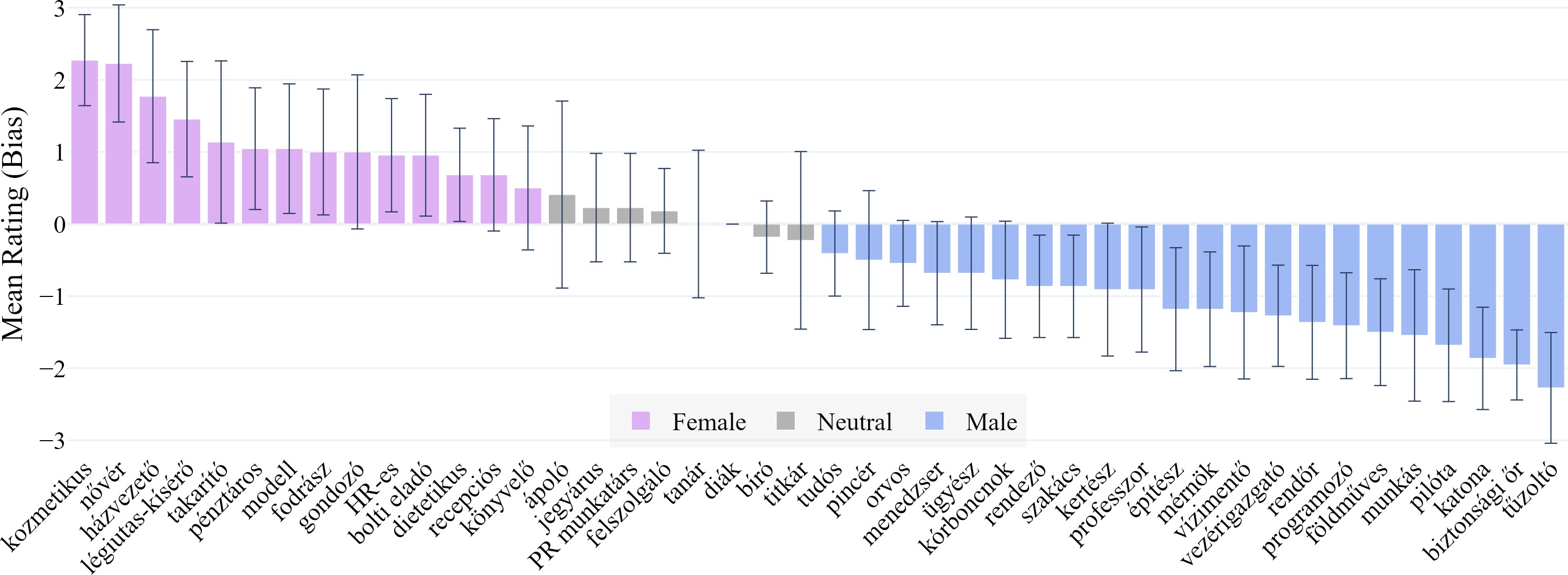


Figure 5: Mean ratings of occupational titles in Hungarian with standard deviations, significant gender bias highlighted – [explore the interactive plot](https://htmlpreview.github.io/?https%3A//github.com/partigabor/occupational-bias/blob/main/occupations_hu.html).

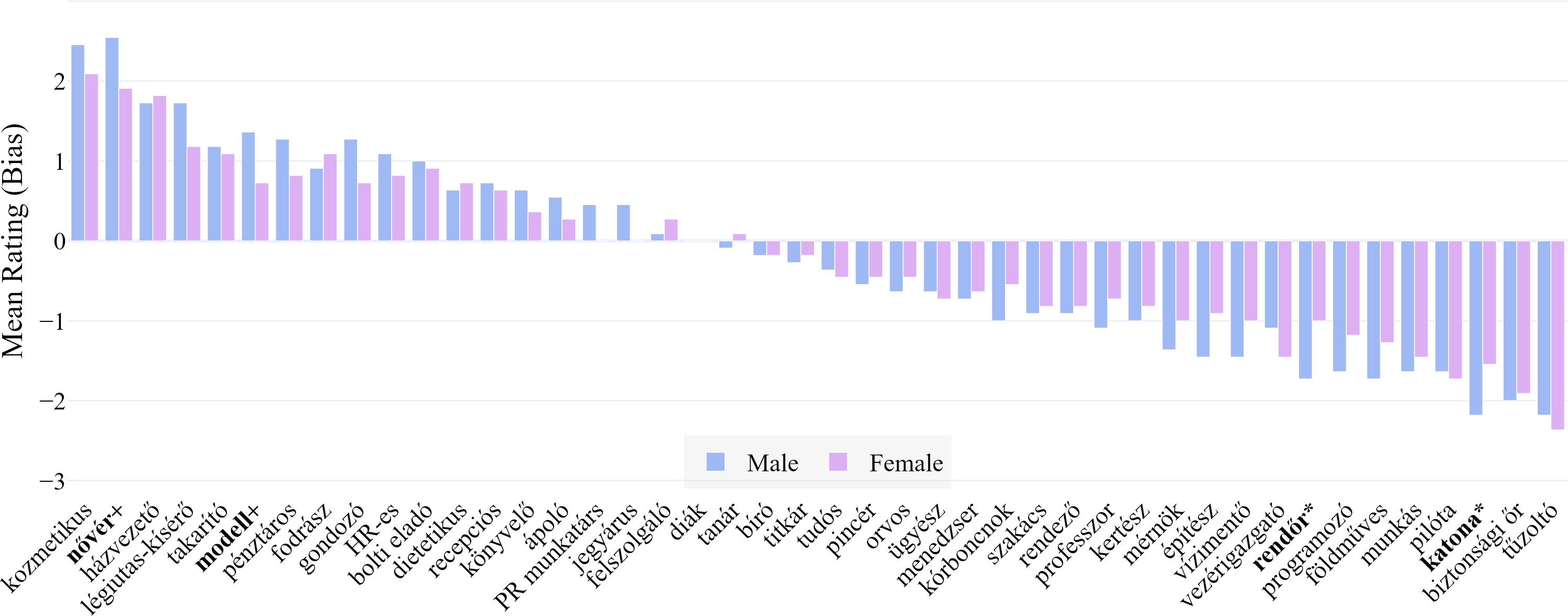


Figure 6: Mean ratings of occupational titles in Hungarian by gender, significant differences highlighted (significant\*, and marginally significant+ in **bold**) – [explore the interactive plot](https://htmlpreview.github.io/?https%3A//github.com/partigabor/occupational-bias/blob/main/occupations_hu_gender.html).

|  |  |  |
| --- | --- | --- |
| **335** *tanár* ‘teacher’ (0), *bíró* ‘judge’ (-0.18), and *titkár* | lice officer’ (male=-1.73; female=-1.00) and *katona* | **352** |
| **336** ‘secretary’ (-0.23). It is worth noting that while *diák* | ‘soldier’ (m=-2.18; f=-1.55), here the male biases | **353** |
| **337** ‘student’ was rated 0 by everyone, *tanár* ‘teacher’ | were much higher by male raters. Two marginally | **354** |
| **338** had more individual variation in the ratings, leading | significant different items were also found in *mod-* | **355** |
| **339** to a higher standard deviation. | *ell* ‘model’ (m=1.36; f=0.73) and *nővér* ‘nurse’ | **356** |
| **340** The strongest agreement were on *diák* ‘student’ | (m=2.55; f=1.91) showing a similar trend. The | **357** |
| **341** (0, std=0), *biztonsági őr* ‘security guard’ (-1.95; | results are summarized in Figure [6](#_bookmark9). | **358** |
| **342** std=0.4857), *bíró* ‘judge’ (-0.18; std=0.5011), *fel-* | Furthermore, it seems like men’s ratings tended | **359** |
| **343** *szolgáló* ‘server’ (0.18; std=0.5885), *tudós* ‘scien- | to have a greater absolute bias for both male- and | **360** |
| **344** tist’ (-0.41; std=0.5903). | female-coded jobs (see Figure [10](#_bookmark14) below). | **361** |

**345**

**346**

## Intra-language gender differences in the Hungarian data

* 1. **Chinese** **362**

|  |  |  |
| --- | --- | --- |
| **347** We also ran a two-sample *t*-test to compare the | Similarly to Hungarian, we found that a majority | **363** |
| **348** ratings of male and female participants for each | of occupations in Chinese were also rated with sig- | **364** |
| **349** occupation, and see if there was any discrepan- | nificant gender bias. The results of the one-sample | **365** |
| **350** cies between the two groups. The only jobs that | *t*-test showed that 39 out of 44 occupations were | **366** |
| **351** showed a significant difference was *rendőr* ‘po- | biased. The mean ratings are shown in Figure [7](#_bookmark10). | **367** |

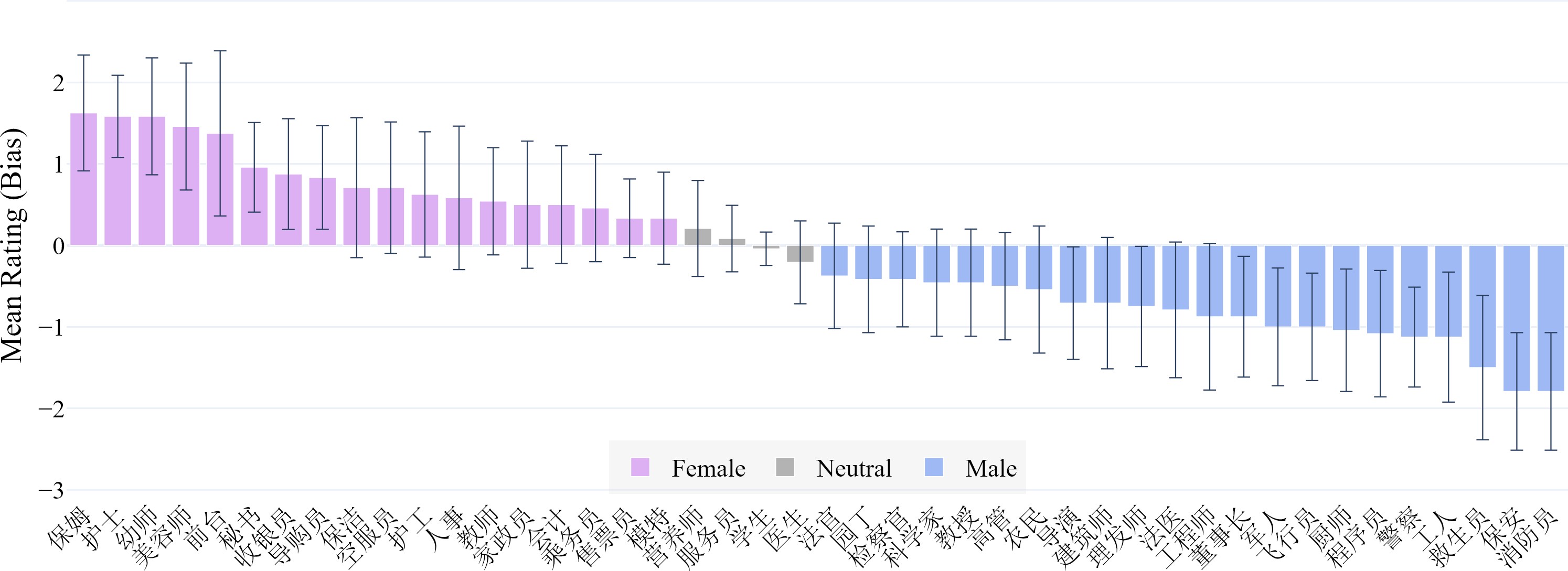


Figure 7: Mean ratings of occupational titles in Chinese with standard deviations, significant differences highlighted

– [explore the interactive plot](https://htmlpreview.github.io/?https%3A//github.com/partigabor/occupational-bias/blob/main/occupations_zh.html).

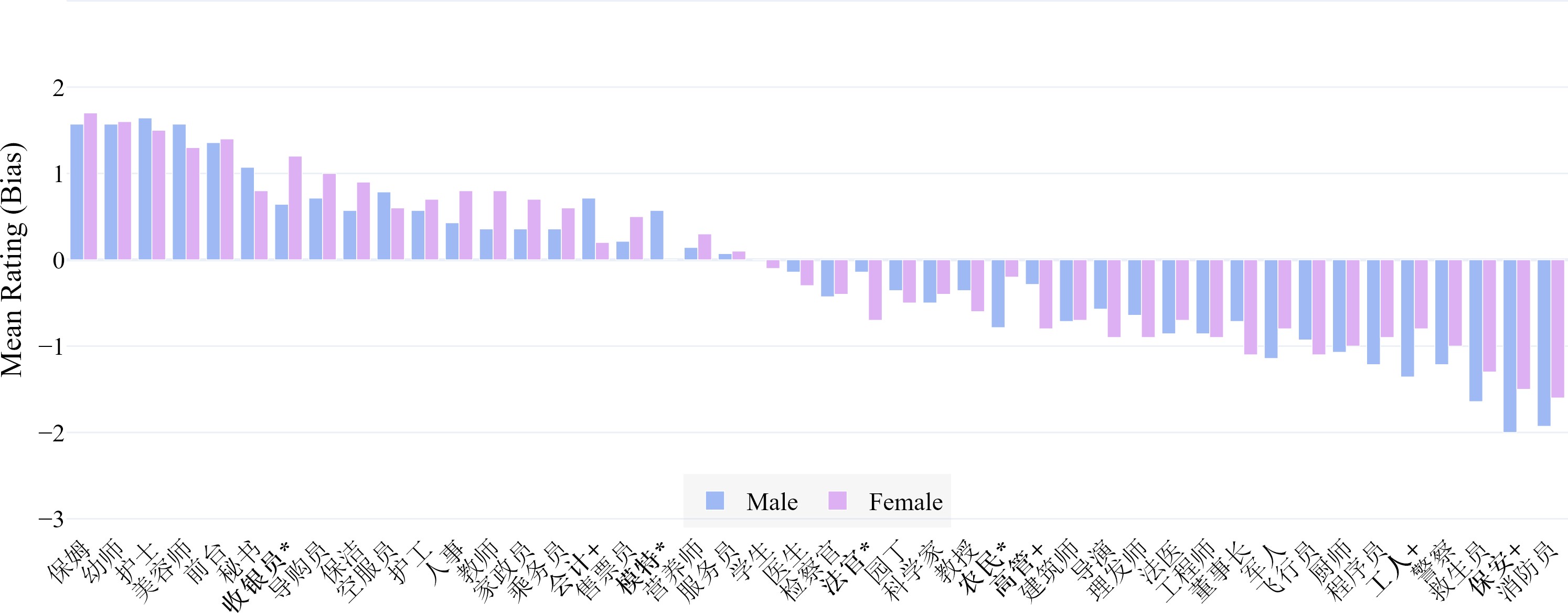


Figure 8: Mean ratings of occupational titles in Chinese by gender, significant differences highlighted (significant\*, and marginally significant+ in **bold**) – [explore the interactive plot](https://htmlpreview.github.io/?https%3A//github.com/partigabor/occupational-bias/blob/main/occupations_zh_gender.html).

**368**

**369**

**370**

**371**

**372**

**373**

**374**

**375**

**376**

**377**

**378**

**379**

**380**

**381**

**382**

**383**

**384**

## Overall distribution of ratings

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

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 ‘reception-

ist’ (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 Hungar-

ian 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 **385**

学生 ‘student’ (-0.04; std=0.2041) and 服务员 ‘wa- **386**

ter/server’ (0.08; std=0.4082), the remaining words **387**

in the top 5 jobs with the lowest standard devia- **388**

tion were 售票员 ‘ticket seller’ (0.33; std=0.4815), **389**

护士 ‘nurse’ (1.58; std=0.5036), and 医生 ‘doctor’ **390**

(-0.21; std=0.5089). **391**

## Intra-language gender differences in the 392

**Chinese data** **393**

The two-sample *t*-test comparing the ratings of male **394**

vs. female participants showed that there were sig- **395**

nificant differences between what people think of a **396**

typical 收银员 ‘cashier’ (m=0.64; f=1.20), 模特 **397**

‘model’ (m=0.57; f=0), 法官 ‘judge’ (m=-0.14; **398**

f=0.70), and 农民 ‘farmer’ (m=-0.79; f=-0.20). **399**

Marginally significant differences were found for **400**

会计 ‘accountant’ (m=0.71; f=0.20), 高管 ‘man- **401**

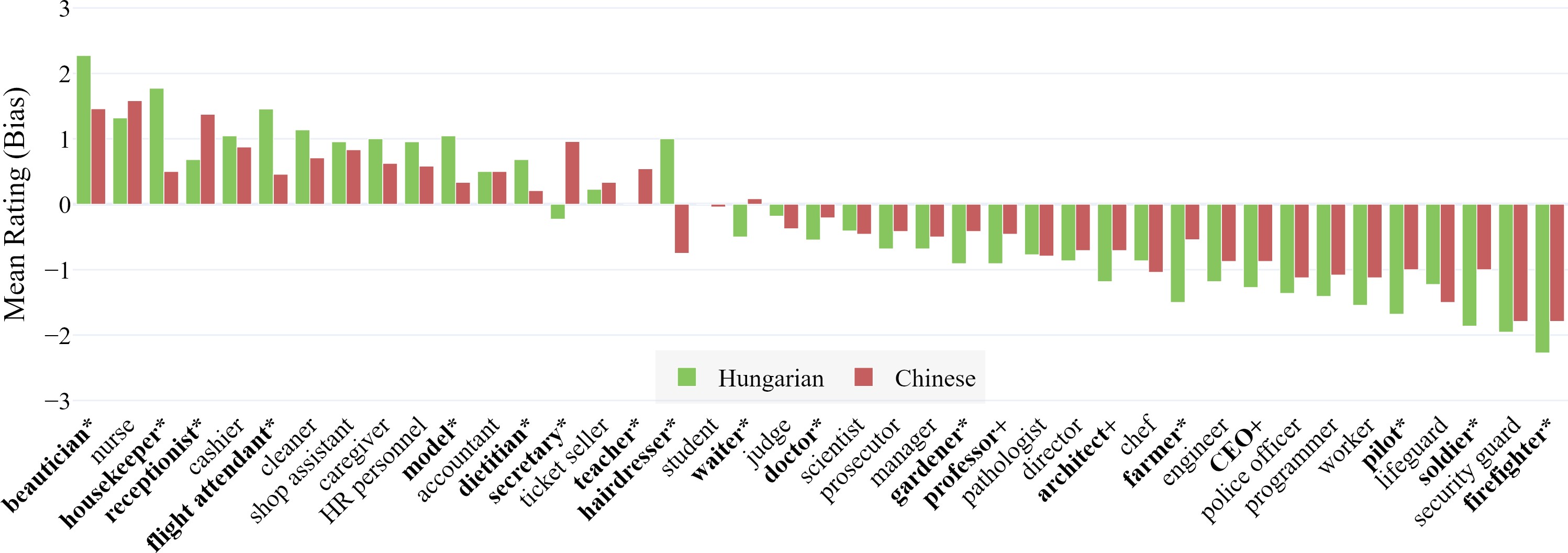


Figure 9: Mean ratings of common occupational titles in Hungarian and Chinese (significant\*, and marginally significant+ in **bold**) – [explore the interactive plot](https://htmlpreview.github.io/?https%3A//github.com/partigabor/occupational-bias/blob/main/occupations_comparison.html).

**402**

**403**

**404**

**405**

**406**

**407**

**408**

**409**

**410**

**411**

**412**

**413**

**414**

**415**

**416**

**417**

**418**

**419**

**420**

**421**

**422**

**423**

**424**

**425**

**426**

**427**

**428**

**429**

**430**

**431**

**432**

**433**

**434**

ager’ (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](#_bookmark11).

## 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 accord- ing to their meanings. Then, using a two-sample *t*-test, we checked if there were significant differ- ences between the ratings in the two languages. The results are summarized in Figure [9](#_bookmark12).

When comparing the two sets of ratings, the first noticeable trend is that in general, the two lan- guages have similar biases for the same occupations. Shared items on the extreme ends of the scale in- clude for example ‘beautician’ and ‘nurse’, as well as ‘firefighter’ and ‘security guard’. Job titles that were neutral/unbiased in both datasets were ‘server’ (*felszolgáló*, ‘���’), and ‘student’ (*diák*, ‘��’).

In terms of significant differences, we found 16 occupations that were rated differently in the two languages. These include ‘beautician’ (Hun- garian=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), ‘hairdresser’ (H=1.00 vs. C=-0.75), ‘waiter’ (H=-0.50 vs.

C=0.08), ‘doctor’ (H=-0.55 vs. C=-0.21), ‘gar-

dener’ (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 dif- **435**

ferences were found for ‘professor’, ‘architect’ and **436**

‘CEO’. **437**

— Wenhui, this is where you can try and dis- **438**

cuss some of these significantly different words **439**

that could be interesting for readers. For example, **440**

you can help to explain that: **441**

— housekeeper is not exactly the same as 家政员 **442**

(housekeeper in Hungarian is equal to a maid, major **443**

tasks are cleaning and cooking. I heard that 家政员 **444**

is more serious.) **445**

— secretary has more than one meaning (I don’t **446**

know Chinese, but Hungarian yes, one is a secretary **447**

of a political party (leadership position), other is a **448**

personal assistant (subordinate).) **449**

* whatever else you can say about the signifi- **450**

cantly different words. **451**

* the top 10 most different occupations in order **452**

are: ‘hairdresser’, ‘housekeeper’, ‘flight attendant’, **453**

‘farmer’, ‘soldier’, ‘secretary’, ‘beautician’, ‘pilot’, **454**

‘model’, and ‘receptionist’. **455**

… **456**

The most striking dissimilarity was the word for **457**

‘hairdresser’ (*fodrász* vs. 理发师), which shows a **458**

strong female bias in Hungarian (1.00) but a defi- **459**

nite male bias in Chinese (-0.75). We believe that **460**

this is a neat example for a cultural difference, in **461**

Hungary hairdressers are perceived to be predom- **462**

inantly female, while in China the profession is **463**

perceived to be more male-dominated. It is not dif- **464**

ficult to find evidence for the latter from the press, **465**

even in English-language media.[6](#_bookmark13) **466**

… **467**

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

**468**

**469**

**470**

**471**

**472**

**473**

**474**

**475**

**476**

**477**

**478**

**479**

**480**

**481**

**482**

**483**

**484**

**485**

**486**

**487**

**488**

**489**

**490**

**491**

**492**

**493**

**494**

**495**

**496**

**497**

**498**

**499**

**500**

**501**

**502**

**503**

**504**

## Discussion

The first obvious thing to notice when looking at Figure [9](#_bookmark12) 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 rea-

sons 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?

## 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 explo- ration to fully explain.

Tie in Kaukonen et al. here

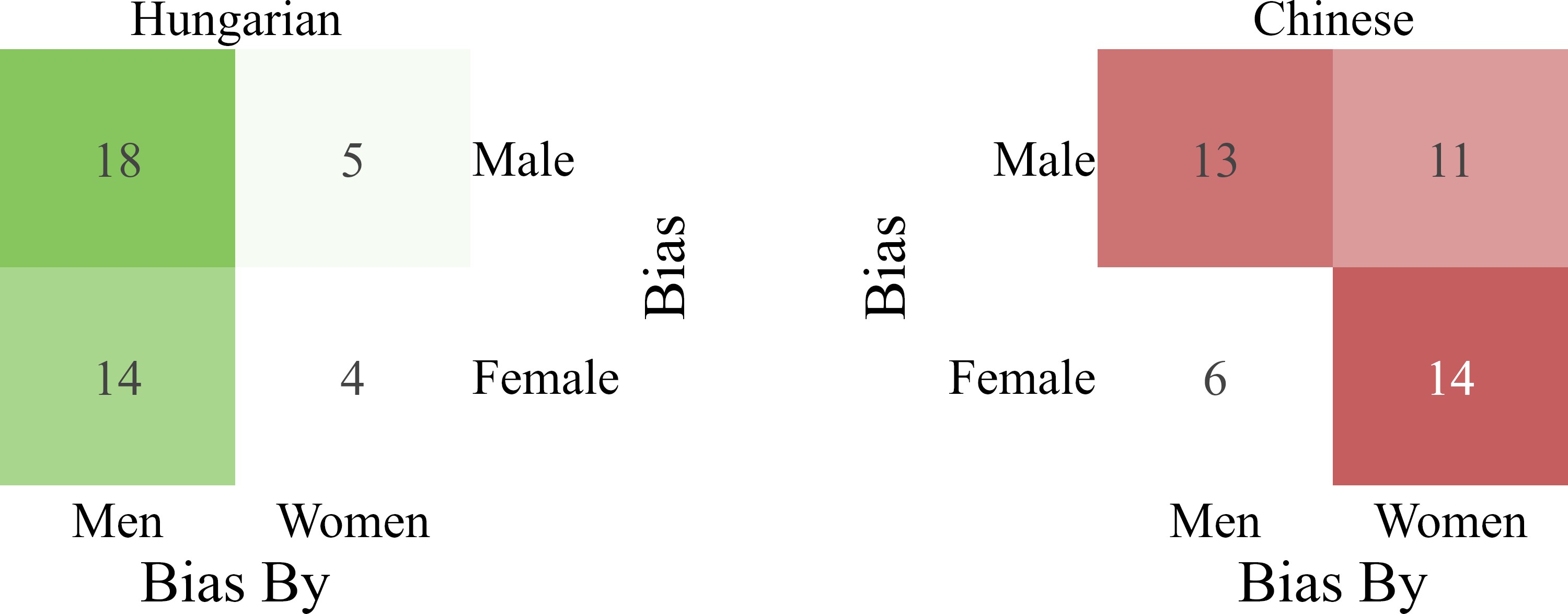


Figure 10: Confusion matrix of the Hungarian and Chi- nese ratings, showing the differences in the ratings of male and female participants.

## 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](#_bookmark14).

# Comparing the human ratings with AI-generated ratings

Gabor is still working on this part...

Now, we were also curious how the human rat- ings compare to the ratings of Large Language

Models (LLMs) used in popular AI agents, and so **505**

we basically repeated the same experiment using **506**

AI instead of human raters. We prompted Copilot **507**

(+Think Deeper), ChatGPT (+Reason), Gemini (2.5 **508**

Flash), and Deepseek (Deepthink R1) to elicit a rat- **509**

ing on the same job titles in both languages, using **510**

the same scale as the human raters. The instructions **511**

were in Hungarian and in Chinese, respectively, **512**

with a pretext: “You are participating in an exper- **513**

iment and your answers will help our research.”. **514**

We asked the AI agents 10 separate times, and took **515**

the mean as a baseline for the results for each job **516**

title. The mean AI ratings were then superimposed **517**

on the human raters’ plots to allow for a convenient **518**

comparison, you can see these in Figures [11](#_bookmark16) and **519**

[12](#_bookmark17). **520**

The rationale behind this was to simulate how **521**

the general public and non-experts would turn to **522**

AI to do the job of human raters, and to show that **523**

these practices can be highly problematic, as the **524**

biases encoded in the LLMs are expected to be **525**

even stronger than those of the human raters. Men- **526**

tion real studies on this and cite. We think that the **527**

results perfectly illustrate the dangers of using AI **528**

agents instead of human raters, and only perpetuate **529**

social biases. **530**

1. **Conclusion** **531**

In conclusion, our study provides evidence for occu- **532**

pational gender bias in both Hungarian and Chinese **533**

contexts, using data from native speakers. We have **534**

demonstrated that gender stereotypes are present **535**

in job titles, even if the words themselves are un- **536**

marked for gender, and this is due to societal norms **537**

and cultural influences, reflecting deep-seated bi- **538**

ases. **539**

… **540**

For the sake of reproducibility and open sci- **541**

ence, we have made all the data available... please **542**

find all raw data and code in this repository: **543**

<https://github.com/partigabor/occupational-bias>. **544**

ANONYMIZE THE FILES BEFORE **545**

SUBMISSION! **546**

**References** **547**

畢新偉 Xinwei Bi. 2013. [「她」字的來源與女性](https://www.airitilibrary.com/Article/Detail?DocID=P20200814001-201304-202009300017-202009300017-136-142) **548**

[主體性⸺評《「她」字的文化史》 [The Origin](https://www.airitilibrary.com/Article/Detail?DocID=P20200814001-201304-202009300017-202009300017-136-142) **549**

[of Chinese Characters and Female Subjectivity: A](https://www.airitilibrary.com/Article/Detail?DocID=P20200814001-201304-202009300017-202009300017-136-142) **550**

[Review of “The Cultural History of the Chinese](https://www.airitilibrary.com/Article/Detail?DocID=P20200814001-201304-202009300017-202009300017-136-142) **551**

[Character ‘She’ ”]](https://www.airitilibrary.com/Article/Detail?DocID=P20200814001-201304-202009300017-202009300017-136-142). 二十一世紀 *[21st century]*, **552**

(136):136–142. **553**

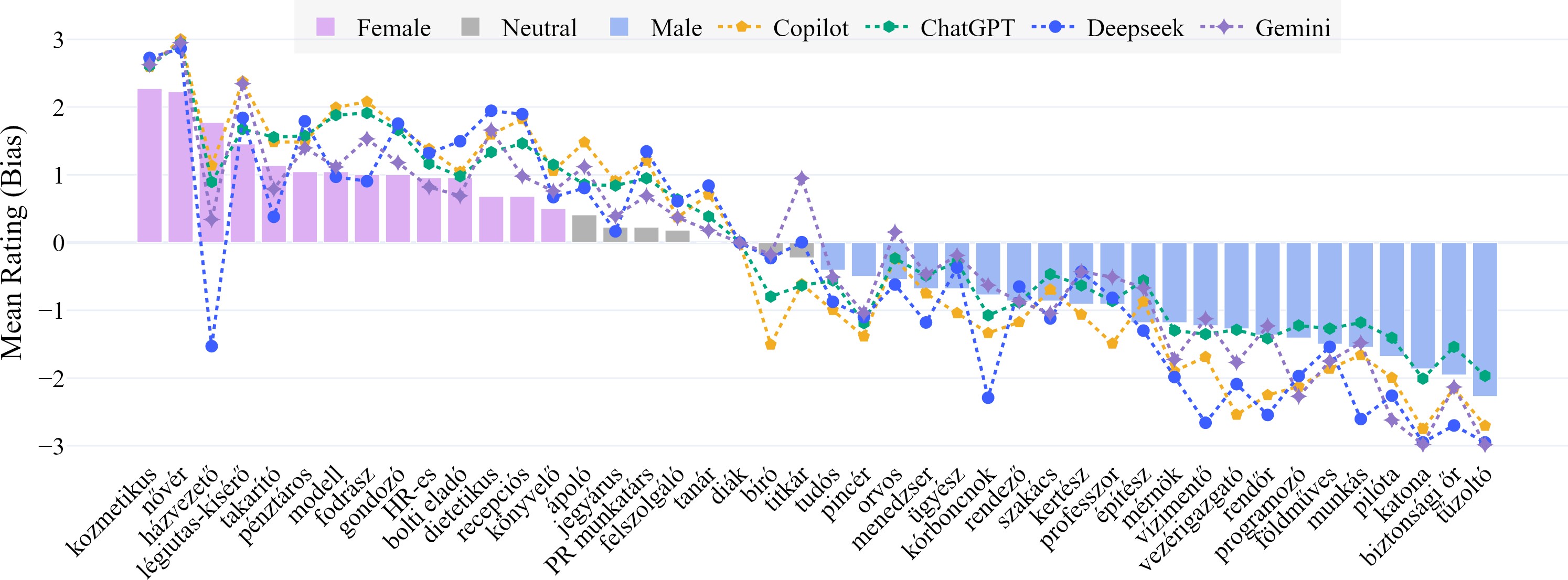


Figure 11: AI ratings for Hungarian occupations [— explore the interactive plot](https://htmlpreview.github.io/?https%3A//github.com/partigabor/occupational-bias/blob/main/occupations_hu_with_ai.html).

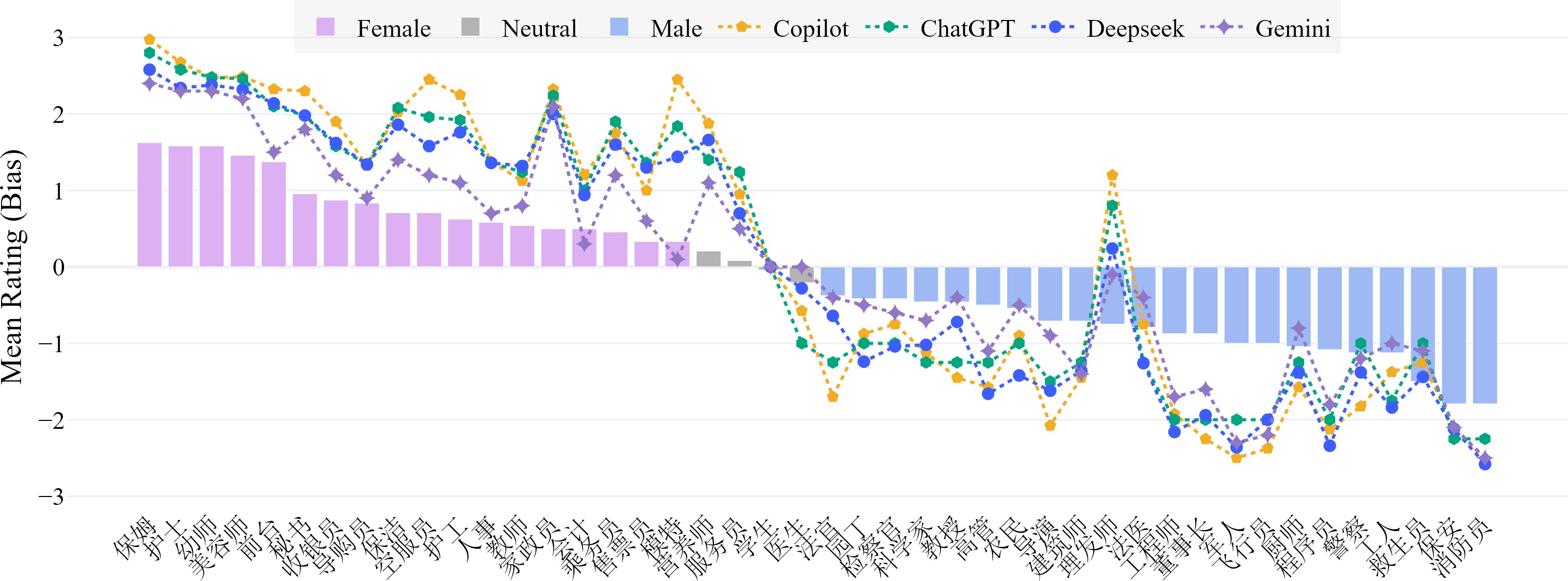


Figure 12: AI ratings for Chinese occupations [— explore the interactive plot](https://htmlpreview.github.io/?https%3A//github.com/partigabor/occupational-bias/blob/main/occupations_zh_with_ai.html).

**554** Elisabeth Kaukonen, Polina Oskolskaia, Liina Lind-

**555** ström, and Raili Marling. 2025. [Gender, language](https://doi.org/10.3389/fcomm.2024.1454022)

**556** [and labour: Gender perception of Estonian and Rus-](https://doi.org/10.3389/fcomm.2024.1454022)

**557** [sian occupational titles](https://doi.org/10.3389/fcomm.2024.1454022). *Frontiers in Communication*,

**558** 9.

**559** Csaba Oravecz, Tamás Váradi, and Bálint Sass. 2014.

**560** The Hungarian Gigaword Corpus. In *LREC*, vol-

**561** ume 9, pages 1719–1723.

**562** Tamás Váradi. 2002. The Hungarian National Corpus.

**563** In *LREC*, pages 385–389.