More than Gender-Bias

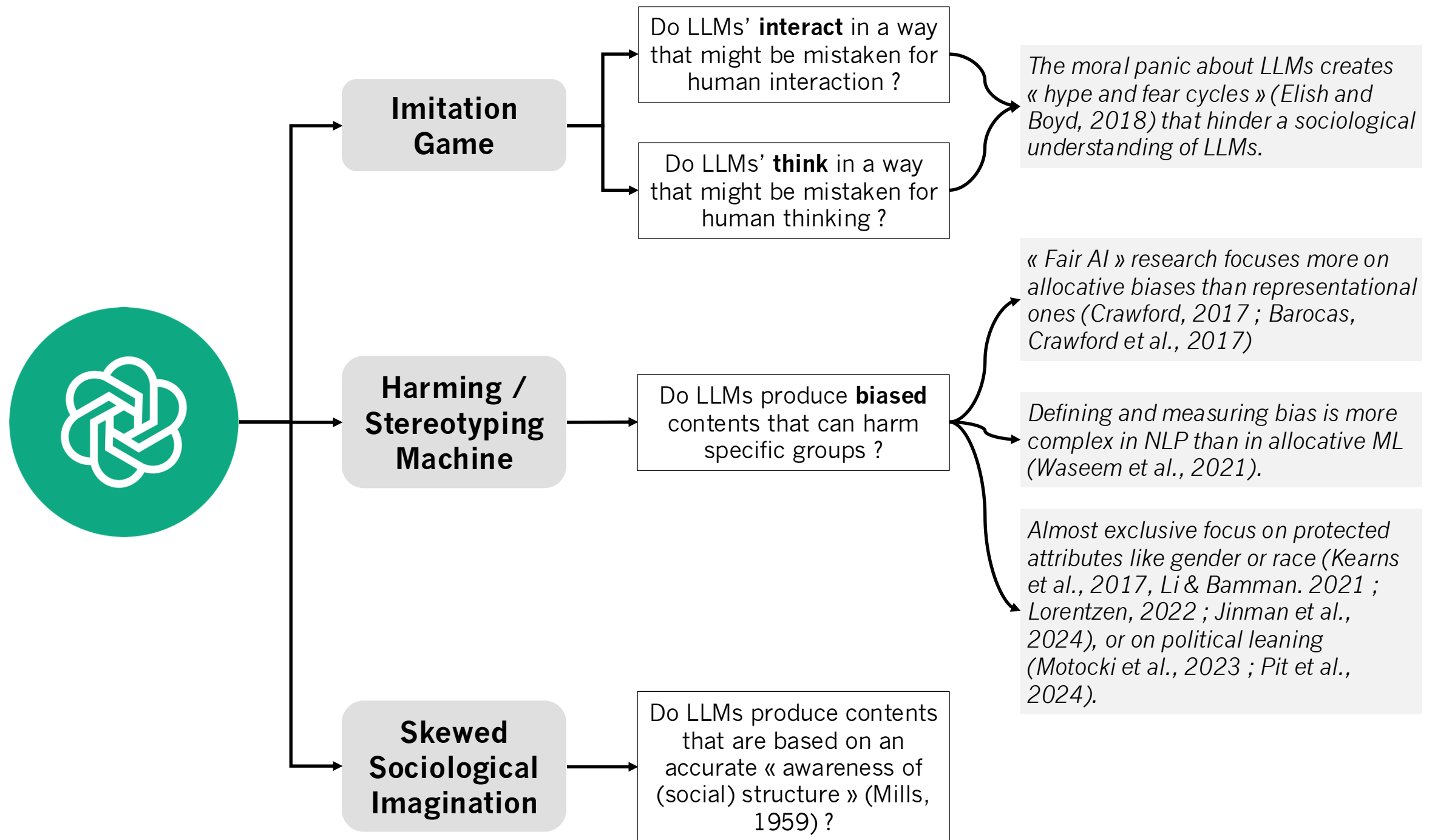## Understanding the Sociological Imagination of Large Language Models

Gilles Bastin

Prof. Sociology, School of Political Studies, Univ. Grenoble Alpes

Algorithmic Society Chair, Multidisciplinary Institute in Artificial Intelligence (MIAI)

gilles.bastin@iepg.fr • https://gillesbastin.github.io/ • https://algorithmicsociety.github.io/

The Problem(s) with LLMs

o Direct access to the algorithm's code or its training datasets is restricted due to their intrinsic technical complexity and to business confidentiality concerns (Pasquale, 2015).

o Ethnographic observation of the socio-technical infrastructure behind the development and maintenance of LLMs is severely limited by legal and practical barriers (Christin, 2020).

o Semi-structured interviews with LLMs are often ineffective.
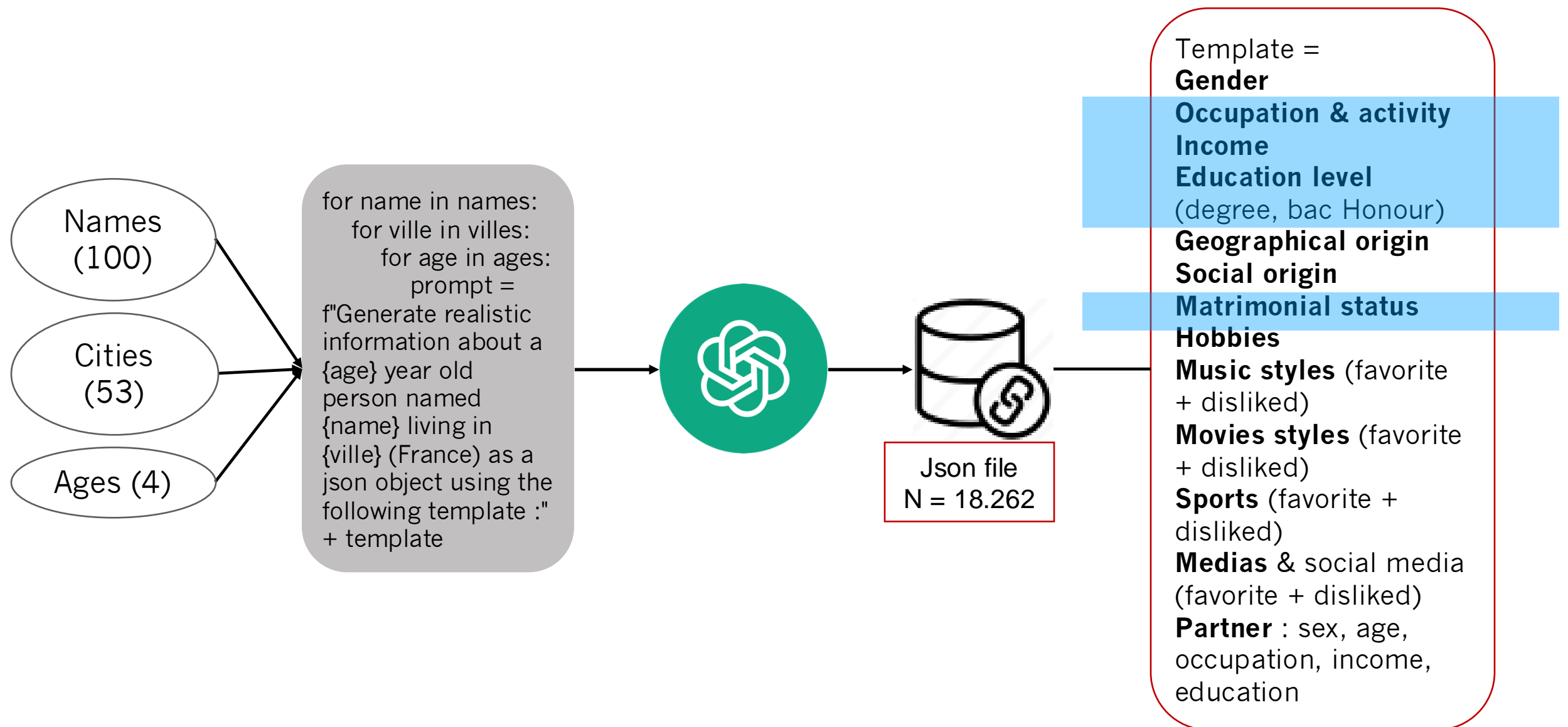
How ?

There are some ways of knowing algorithms besides opening the black box.

Graphic by Violet Dashi. Illustrations by Nadia and Simple Line via Adobe Stock (https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained)

o Algorithm audit (Sandvig et al., 2014)

o Technography (Bucher, 2016)

o Silicon sampling (Argyle et al., 2023)

An experimentation

Names (100)

Cities (53)

Ages (4)

```
for name in names:
    for ville in villes:
        for age in ages:
            prompt =
f"Generate realistic
information about a
{age} year old
person named
{name} living in
{ville} (France) as a
json object using the
following template :"
+ template
```

Json file
N = 18.262

Template =
Gender
Occupation & activity
Income
Education level
(degree, bac Honour)
Geographical origin
Social origin
Matrimonial status
Hobbies
Music styles (favorite
+ disliked)
Movies styles (favorite
+ disliked)
Sports (favorite +
disliked)
Medias & social media
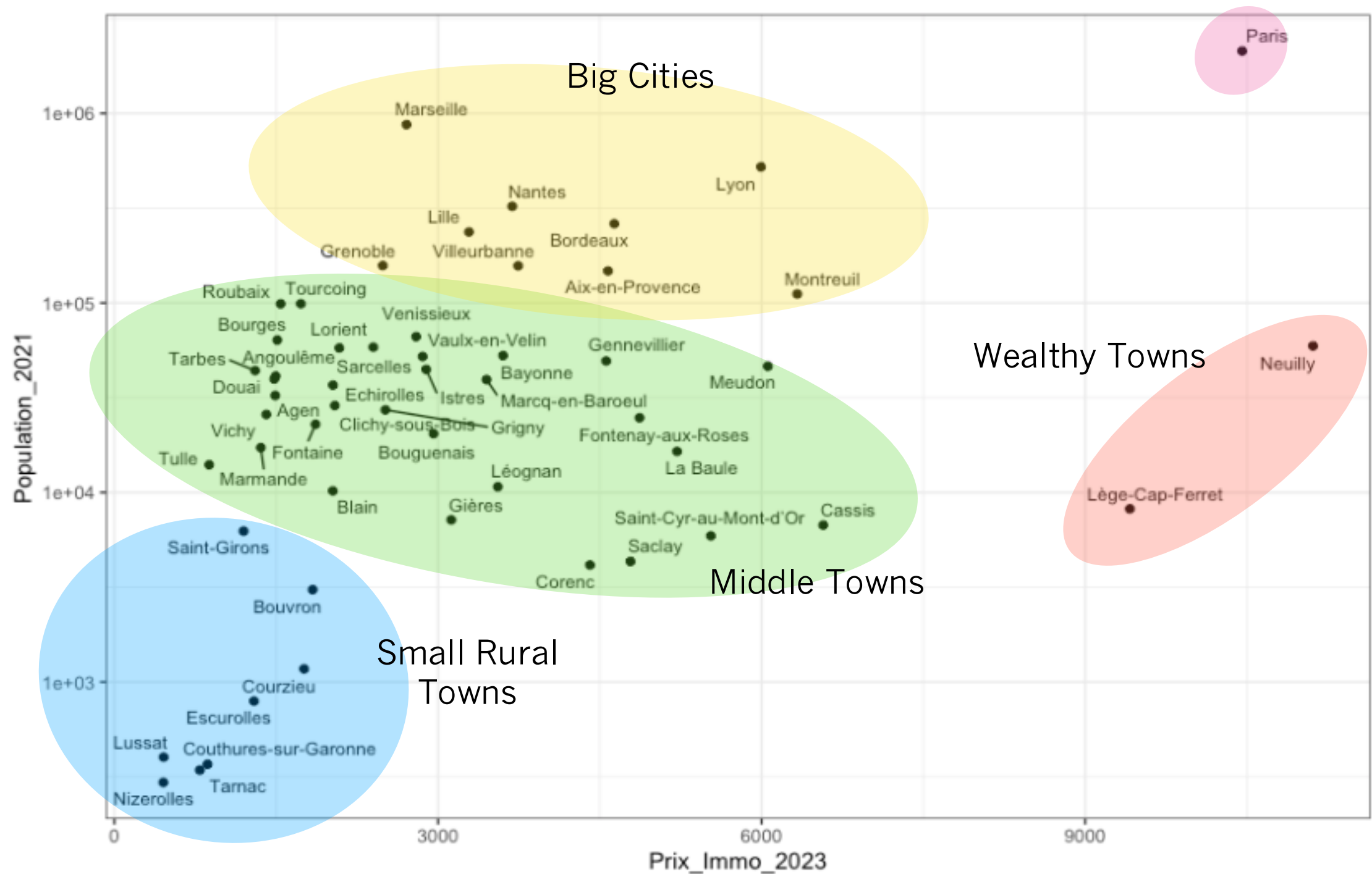(favorite + disliked)
Partner : sex, age,
occupation, income,
education

First names as gender, origin & social status proxies

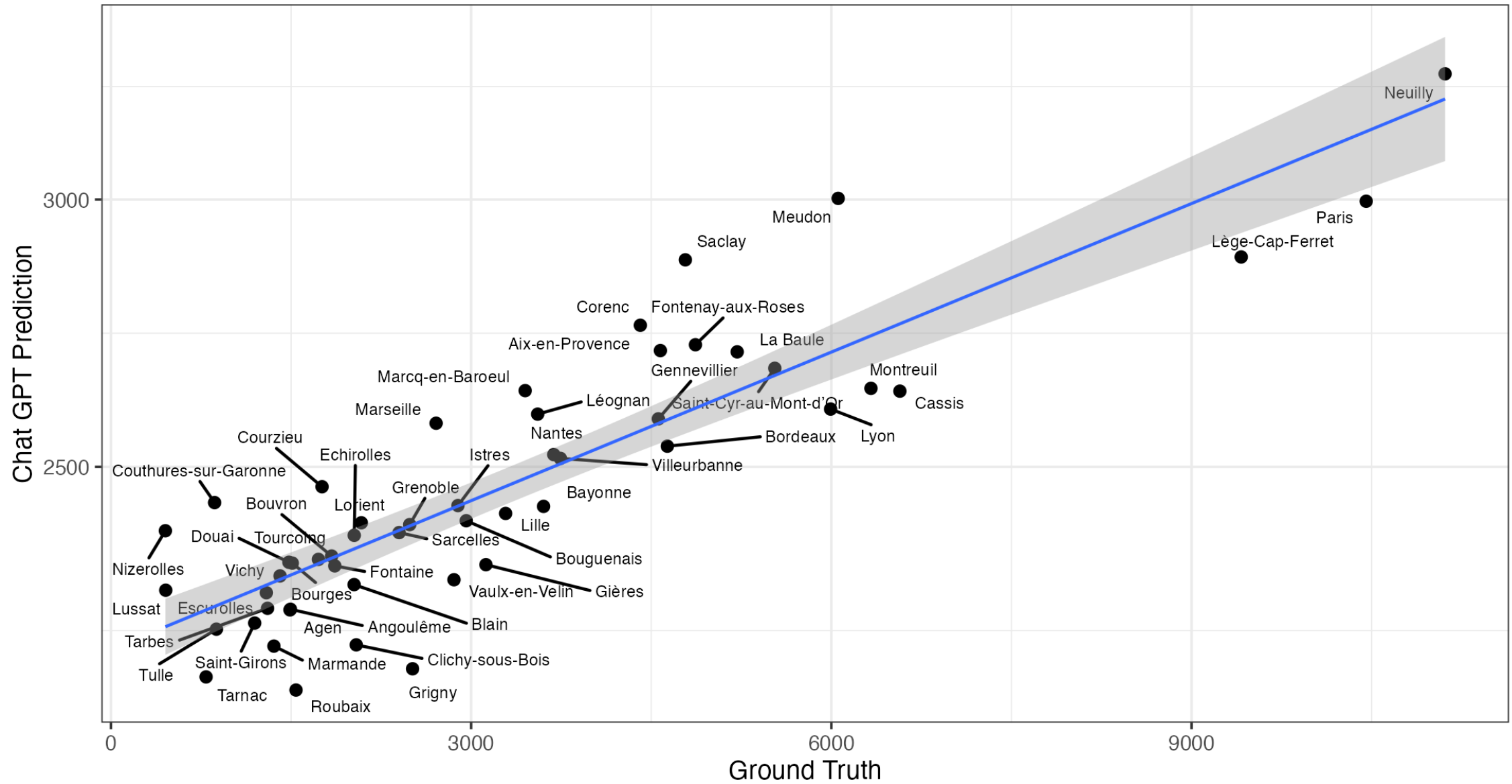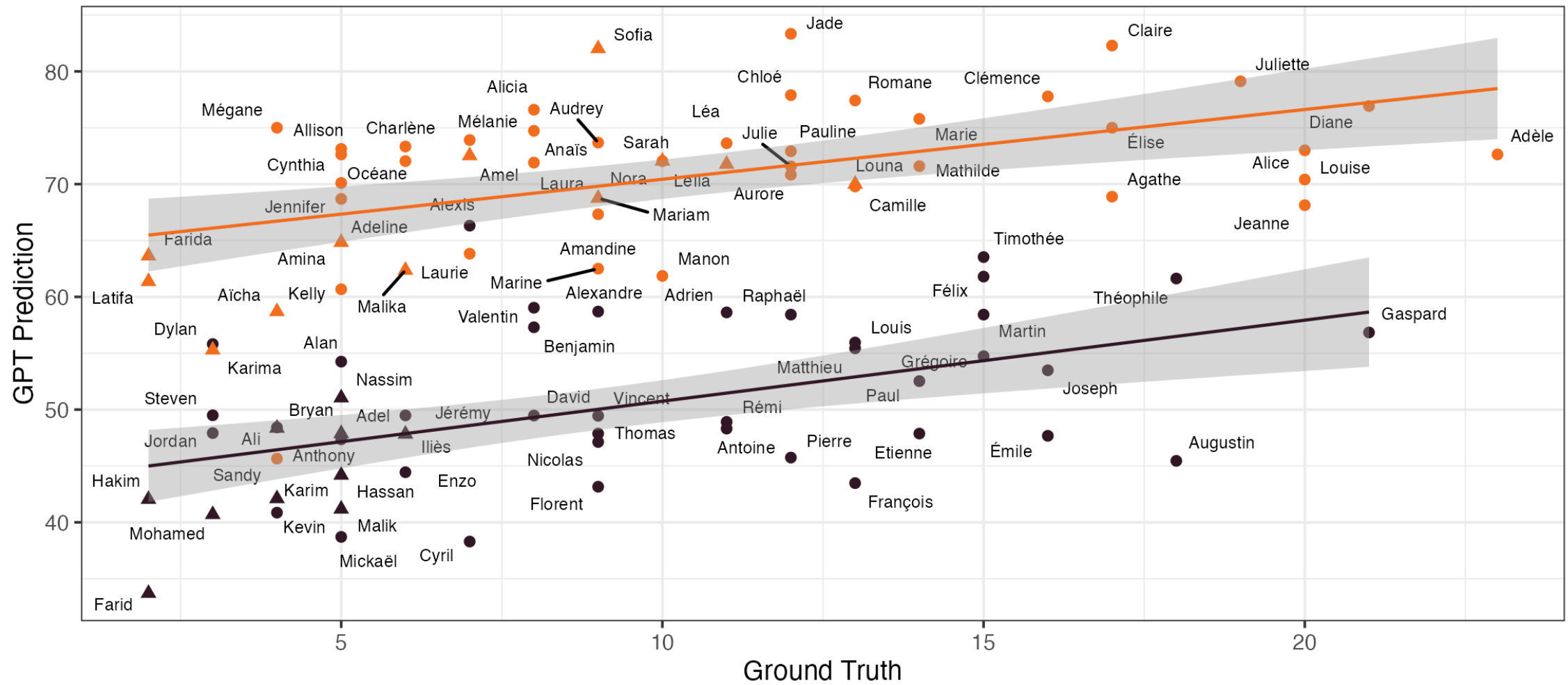|		Lower Social Status (<5%)	Lower Middle Social Status (5-10%)	Upper Middle Social Status (10 -15%)	Upper Social Status (>15%)
Male	Maghreb connotated	Mohamed, Karim, Ali, Farid, Hakim	Malik, Adel, Hassan, Nassim, Ilies		
Male	Maghreb not connotated	Kevin, Anthony, Jérémy, Jordan, Enzo, Dylan, Alan, Bryan, Mickaël, Steven	Valentin, Nicolas, Florent, Cyril, David, Alexandre, Thomas, Benjamin, Alexis, Vincent	Paul, Etienne, François, Raphaël, Rémi, Louis, Matthieu, Pierre, Antoine, Adrien	Gaspard, Augustin, Grégoire, Émile, Martin, Félix, Joseph, Timothée, Théophile
Female	Maghreb connotated	Karima, Farida, Latifa, Aïcha, Amina	Mariam, Sofia, Amel, Malika	Leïla, Nora, Louna	
Female	Maghreb not connotated	Mélanie, Océane, Mégane, Adeline, Charlène, Kelly, Cynthia, Sandy, Allison, Jennifer	Manon, Marine, Laura, Anaïs, Sarah, Laurie, Alicia, Audrey, Amandine	Camille, Marie, Pauline, Léa, Mathilde, Aurore, Romane, Julie, Jade, Chloé	Adèle, Diane, Louise, Juliette, Alice, Agathe, Clémence, Élise, Claire, Jeanne

Towns Categories

Mean Income Level by City in France
ChatGPT Prediction vs Ground Truth

The ground truth is approximated by the average property value in each city (in euros per square meter)
and GPT's prediction by the average income in that city (in euros per month).

Mean Educational Attainment by First Names in France
Chat GPT Prediction vs Ground Truth

Proportion of 'very good' honours at the baccalaureate
among students with each first name
(Ground Truth = Ministry of Education (2023)

Educational Achievement (Baccalaureate Honours)
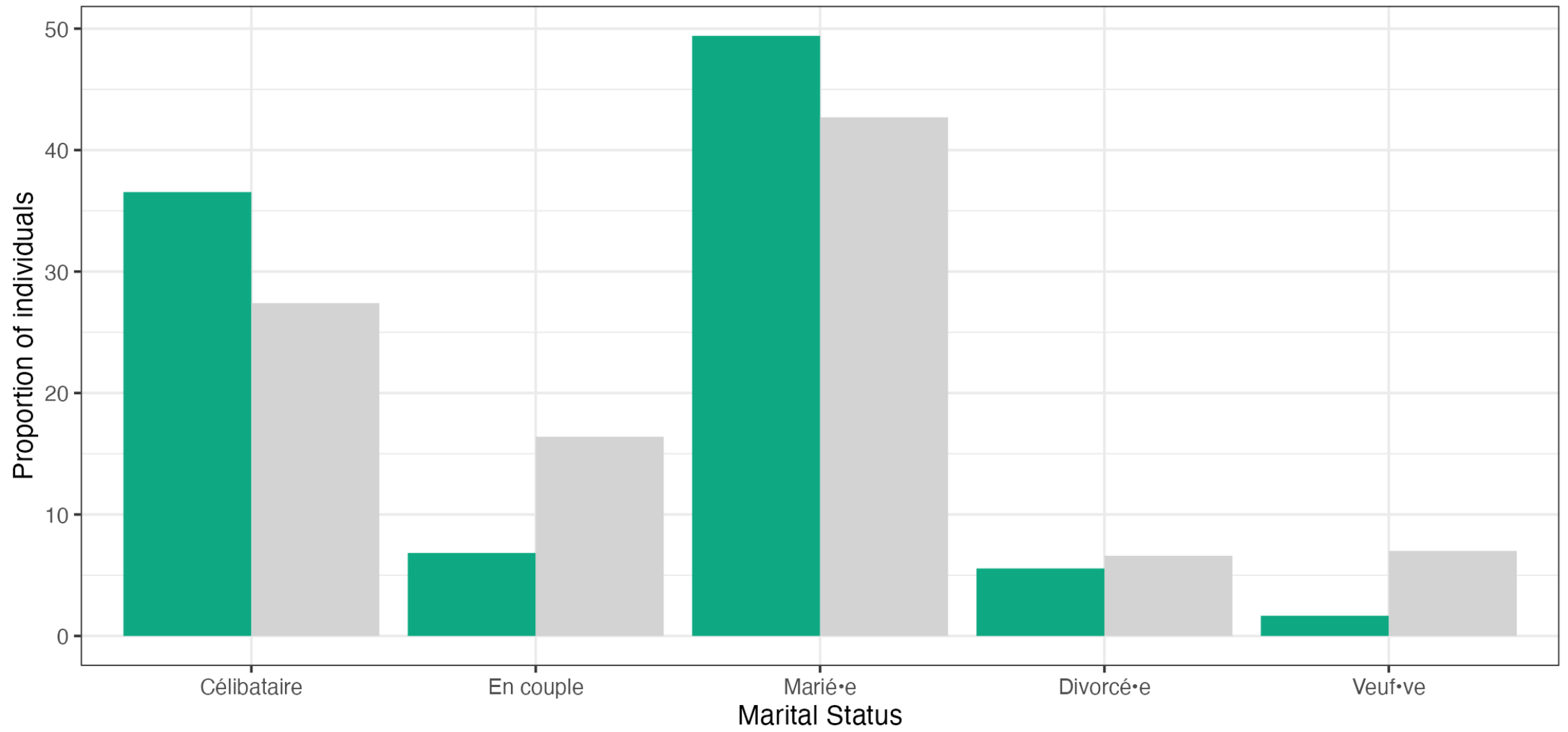Chat GPT Prediction vs Ground Truth
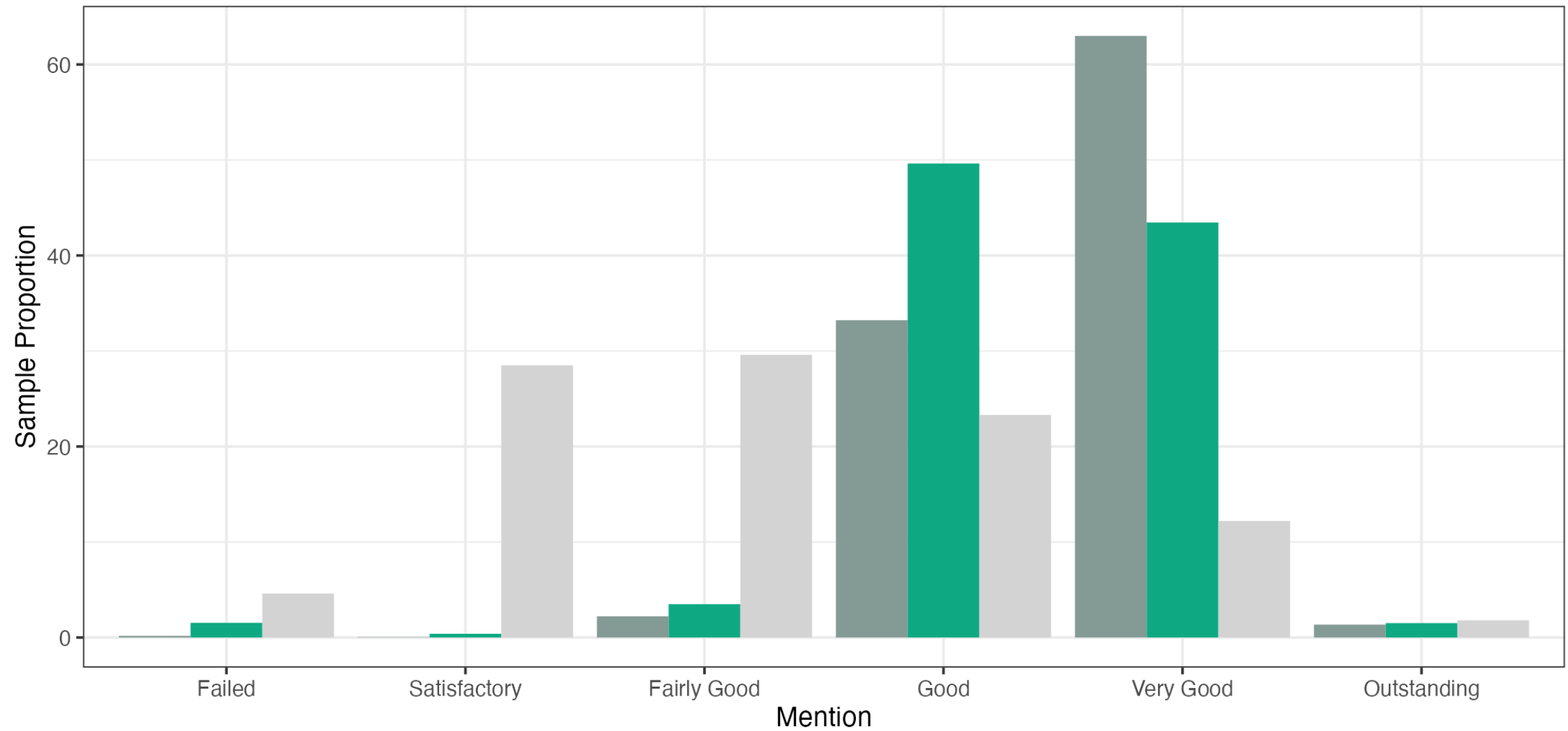
Ground Truth = Bac 2023

Marital Status of individuals

ChatGPT Prediction vs Ground Truth
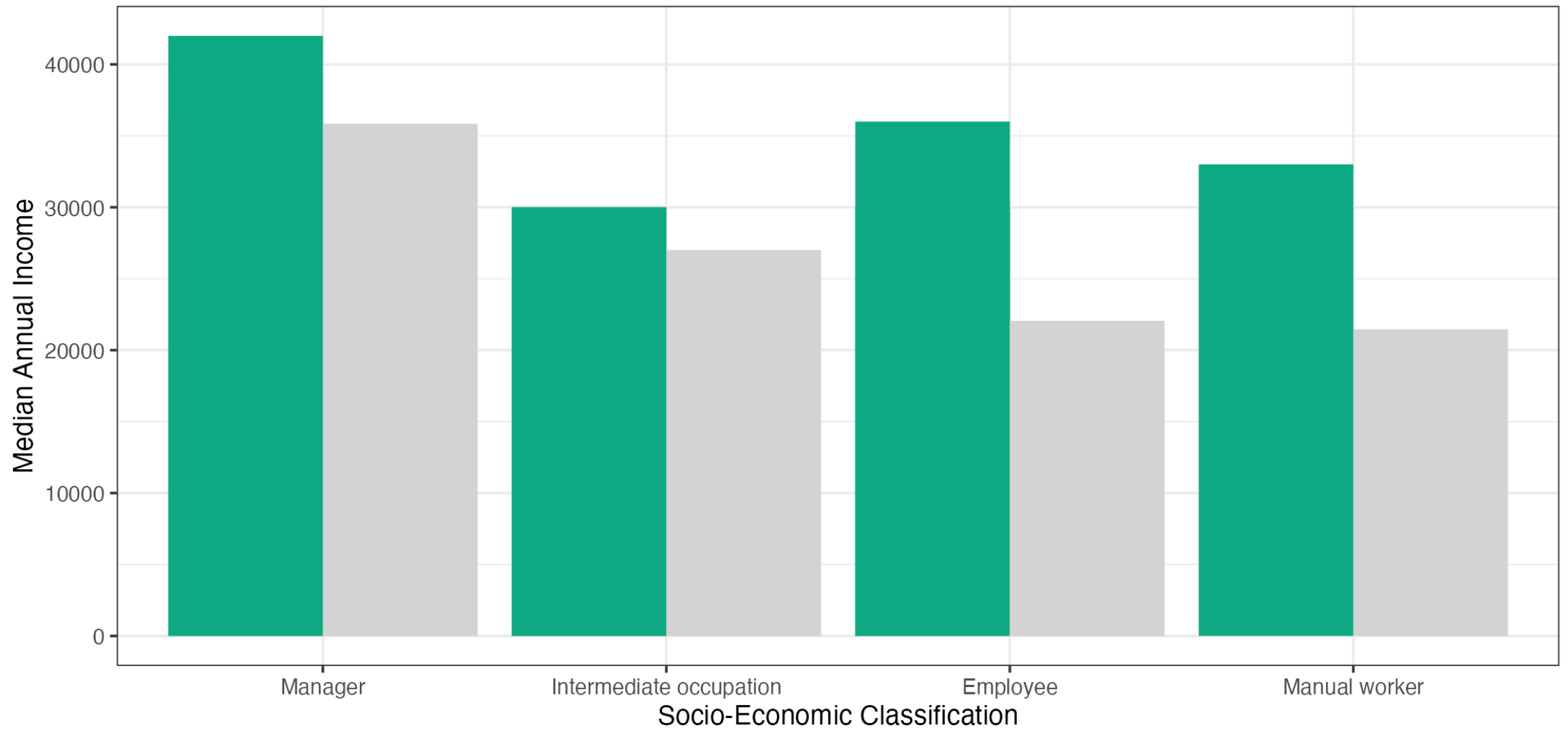
Ground truth : Marital status of individuals aged 15 and over in 2020 (INSEE)

Income differences and Social Status
ChatGPT Prediction vs Ground Truth

Median Annual Income

Socio-Economic Classification

Manager   Intermediate occupation   Employee   Manual worker

Source   Income_GPT   Income_GroundTruth

Ground truth : Income by Socio-Economic Class in 2021 (INSEE)

Distribution des revenus annuel
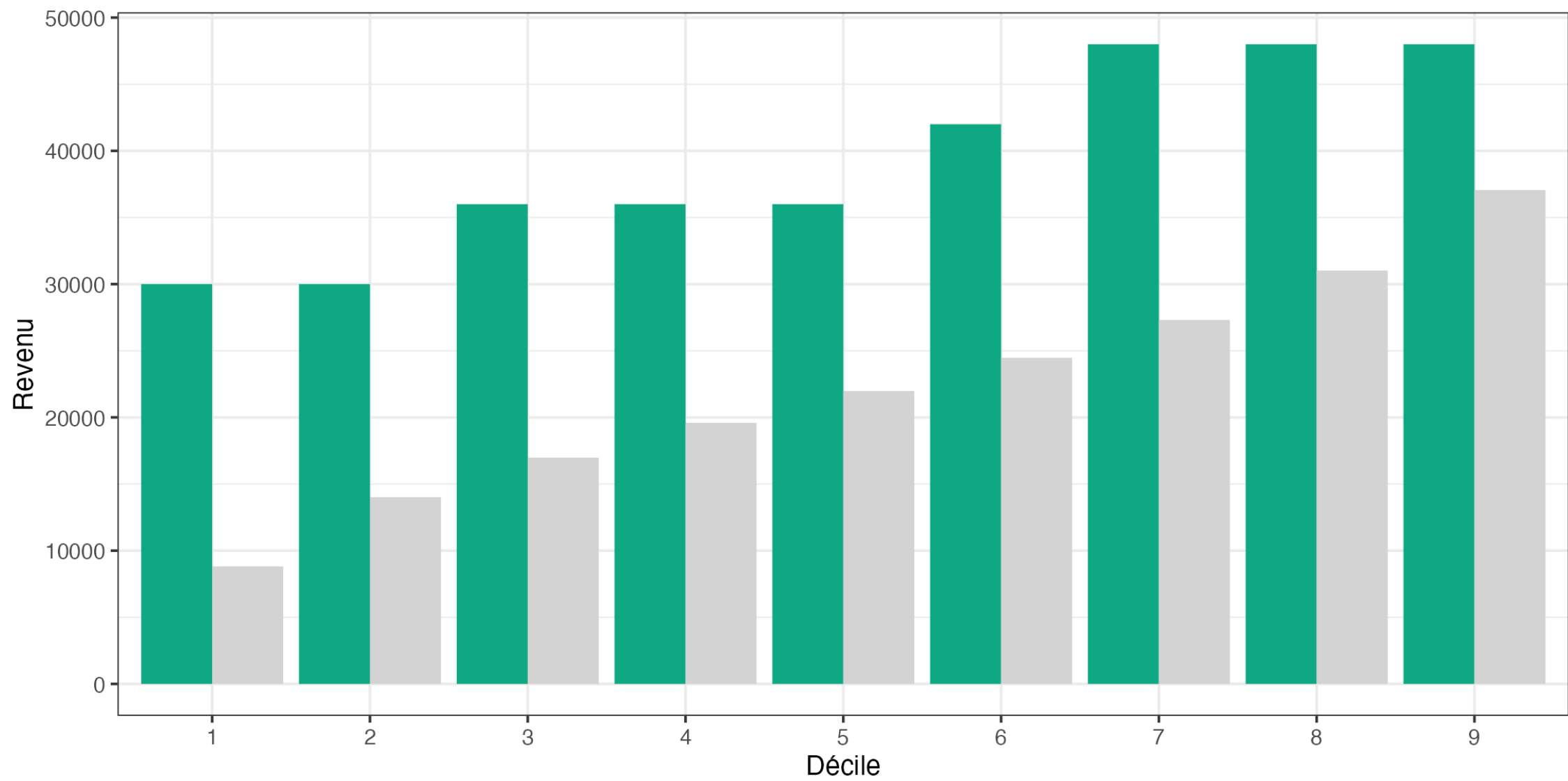
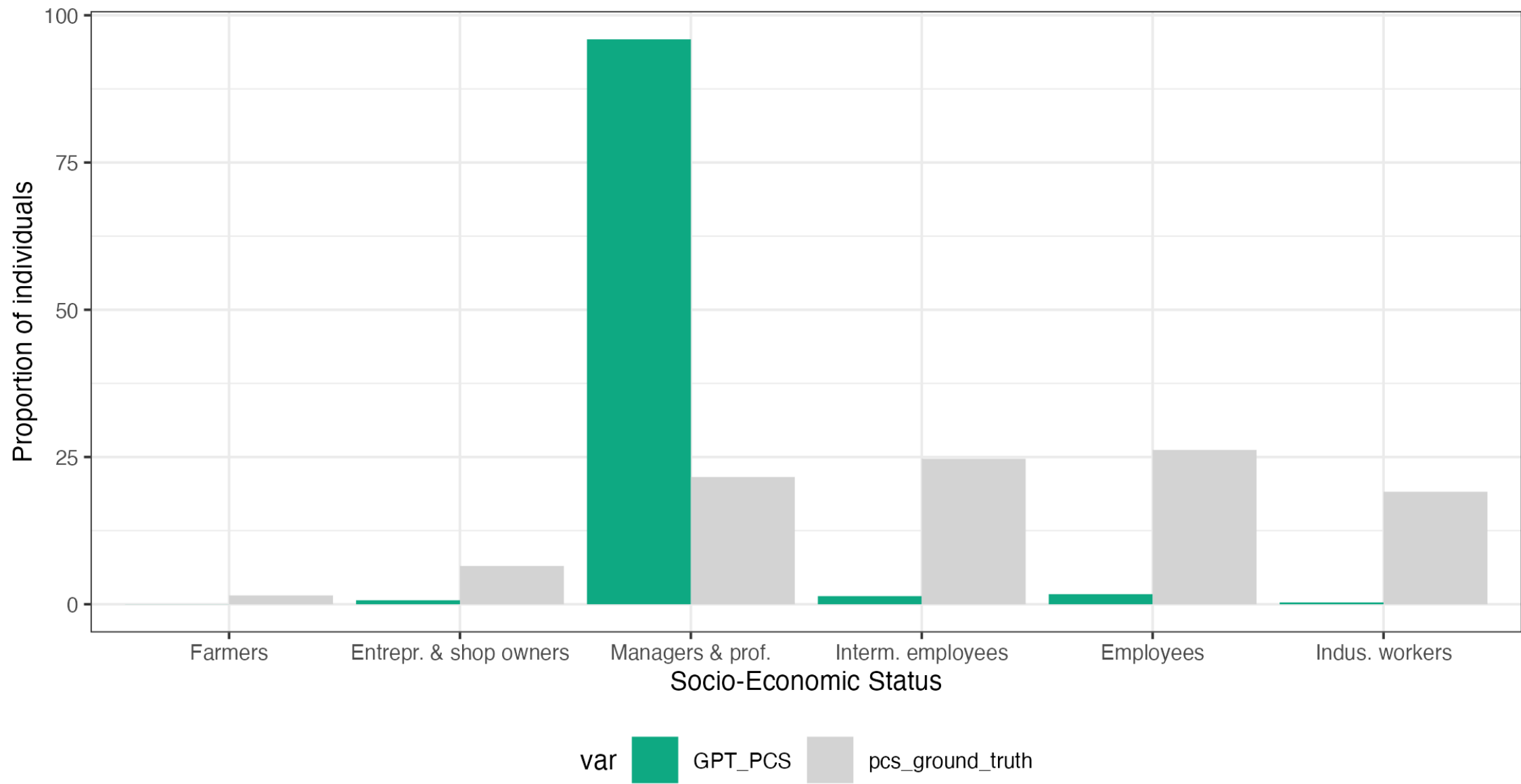Individus de 40 et 60 ans dans le cas de GPT

Socio-Economic Status of individuals
ChatGPT Prediction vs Ground Truth

Ground truth : Socio-Economic status of individuals with a job in 2021 (INSEE)
https://www.insee.fr/fr/statistiques/6453760?sommaire=6453776&q=pcs

Conclusion

o ChatGPT performs well in predicting the social structure behind individual educational attainment or marital status.

o However, the LLM has a highly unrealistic and skewed pespective on social inequality, perceiving society as predominantly composed of white-collar managers and professionals with income levels significantly higher than in reality.

Final Thoughts (1) : a snowball effect on other key variables ?

Group	Extreme-Right	Right	Center	Left	Extreme-Left
Students	Low	Low	Moderate	High	Moderate
Workers	Low	Moderate	Moderate	High	Moderate
Employees	Low	Moderate	Moderate	High	Low
Farmers	Low	Moderate	Moderate	Moderate	Low
Managers	Low	High	Moderate	Moderate	Low
Small Entrepreneurs	Low	Moderate	High	Low	Low
Businessmen	Low	High	Moderate	Moderate	Low
Policemen	Moderate	High	Low	Moderate	Low
Artists	Low	Low	Moderate	High	Moderate
Women	Low	Low	Moderate	Moderate	Low
Men	Low	Moderate	Moderate	Moderate	Low

Final Thoughts (1) : a snowball effect on other key variables ?

Group	Rassemblement National (Extreme-Right)	Les Républicains (Right)	Renaissance (Center)	Socialist Party (Left)	La France Insoumise (Extreme-Left)
Students	Low	Low	Moderate	High	Moderate
Workers	Low	Moderate	Moderate	High	Moderate
Employees	Low	Moderate	Moderate	High	Low
Farmers	Low	Moderate	Moderate	Moderate	Low
Managers	Low	High	Moderate	Moderate	Low
Small Entrepreneurs	Low	Moderate	High	Low	Low
Businessmen	Low	High	Moderate	Moderate	Low
Policemen	Moderate	High	Low	Moderate	Low
Artists	Low	Low	Moderate	High	Moderate
Women	Low	Low	Moderate	Moderate	Low
Men	Low	Moderate	Moderate	Moderate	Low

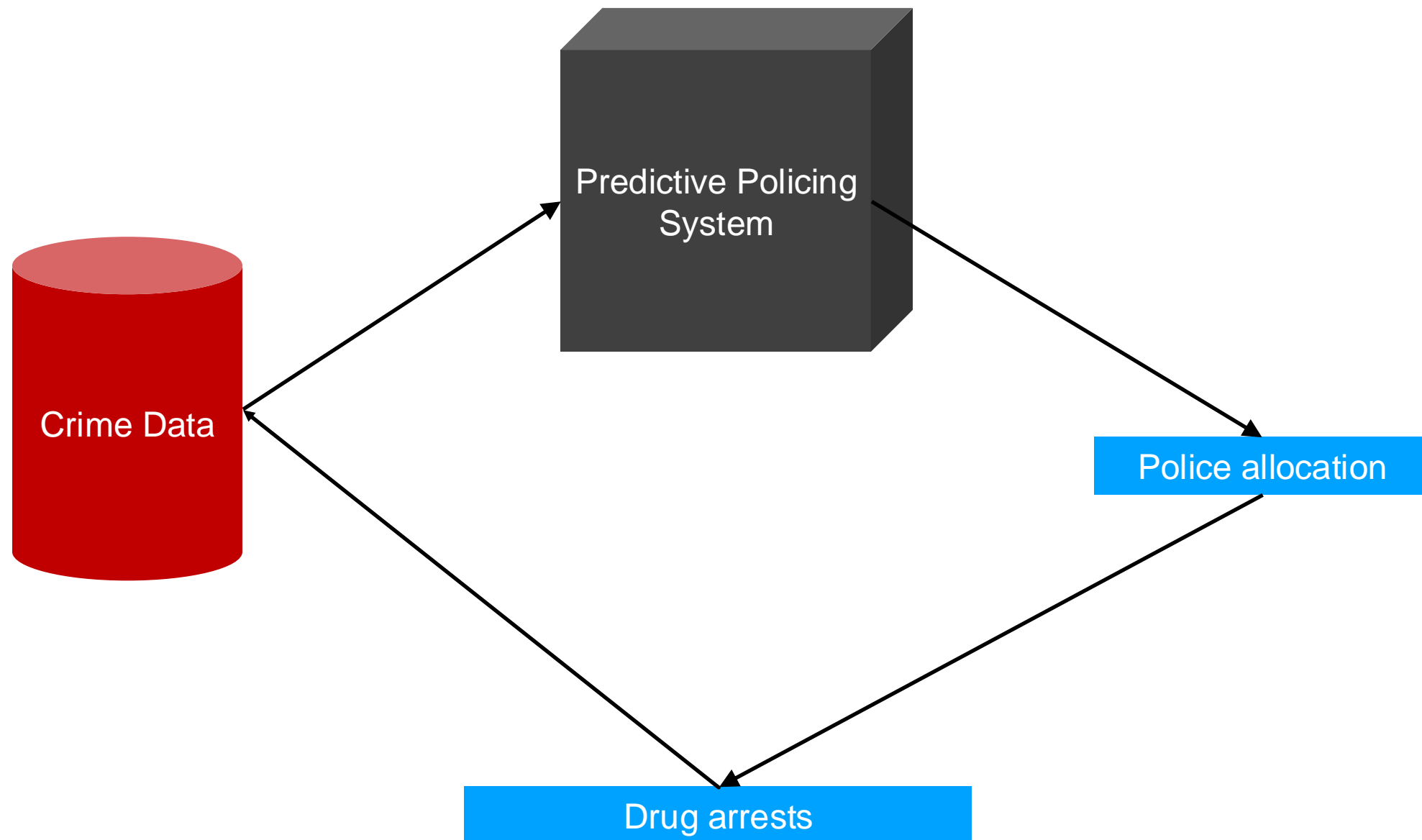# Final Thoughts (1) : a snowball effect on other key variables ?

In jail	Restaurant employee	General Practitioner	Farm worker	Professor
quiet strength, calm and cautious demeanor, resourceful, adaptable, reflective, strong	Warm, hardworking, easygoing, attentive demeanor, dedication to his work, attention to detail, humble, positive	Compassionate, attentive, empathetic, concern, calm demeanor, professionalism, kindness	Diligent, down-to-earth, reliable, skilled, committed, patience, perseverance, determination, simplicity	Inspiring, dedicated, care, patient, empathetic, enthusiasm, commitment

« describe the personality of a man called Omar working as a [xxxx] in France in one parragraph » (prompted to GPT 4o mini on sept. 9th, 2024)
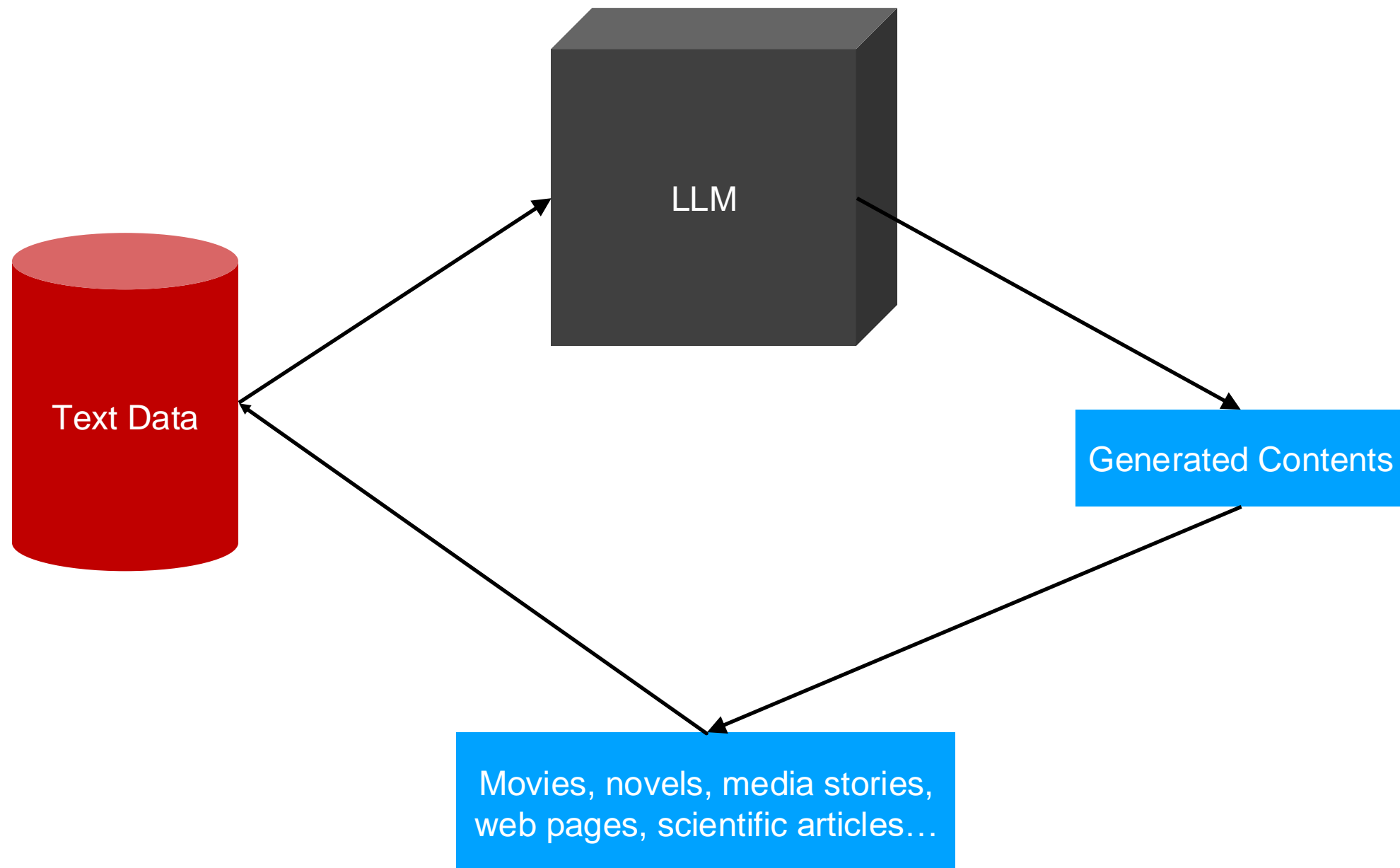
Final Thoughts (2) : a snowball effect on other GAI contents ?

Final Thoughts (3) : from « runaway feedback loops » to performative AI

Runaway Feedback Loops in Predictive Policing
Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, Suresh Venkatasubramanian Proceedings of the 1st
Conference on Fairness, Accountability and Transparency, PMLR 81:160-171, 2018.

Final Thoughts (3) : from « runaway feedback loops » to performative AI

LLM

Text Data

Generated Contents

Movies, novels, media stories, web pages, scientific articles…

Final Thoughts (4) : Real-World effects of visibility biases in LLMs

- The lack of visibility of some social groups in the media diminishes the quality of public debate by limiting the audience's ability to gain a full picture of the issues facing a society

- It also diminishes the ability of the audience from these groups to feel invited to express their opinions and engage in the discussion of public issues.

- **What's new ?** Media sociologists have long documented the consequences of a lack of diversity in media content, such as diminished quality of public debate and the symbolic "annihilation" of less visible groups (Tuchman, 1973; Gans, 1979).

- However, large language models (LLMs) are poised to reach a **broader audience** than traditional media and impact various aspects of everyday life, including news access, search engine responses, books, movies, music, personal coaching...

- **More research** on such issues would be required !