Introduction to Data Science 2025 — Week 5
Privacy, Data Protection & Fairness-aware AI

Roberto Carlos Villafuerte Carrillo

October 6, 2025

Overview

This short report summarizes my findings for Exercise 1 (GDPR basics) and Exercise 2 (fairness-aware AI) using simple language. I cite the relevant GDPR articles using the usual shorthand, e.g., "Art. 6(1)(a)" for Article 6, paragraph 1, item (a).

1 Exercise 1: Privacy and Data Protection

1.1 Valid consent (example: Spotify marketing emails)

When I enabled marketing emails, I saw separate switches for different types of messages (for example, marketing vs. product news). That is important because consent must be a real choice and not bundled together *(GDPR Art. 6(1)(a) and Art. 4(11); Recital 32)*. A link to the privacy policy appeared at the moment of choice, so I could see what data is used and why *(Arts. 12–13)*. It was also easy to undo my choice later (unsubscribe link or a toggle in settings), which is required *(Art. 7(3))*.

Conclusion: In my opinion this is valid consent because it is specific, informed, freely given, and easy to withdraw. If I had been forced to accept many things at once or if unsubscribe was hidden, the consent would not be valid.

1.2 Right of access (what I did and what the law says)

The GDPR gives me the right to know whether my data is processed and to get a copy plus key details like purposes, types of data, recipients, and storage time *(Art. 15)*. Replies must be clear and arrive within one month *(Arts. 12(1), 12(3))* and are free in normal cases *(Art. 12(5))*.

What I did: I used Spotify's account privacy page to request my data. I logged in, submitted the request, and later downloaded the files. This matched the GDPR rules. If a company ignored me or made it too hard, I could complain to a data protection authority *(Art. 77)* or go to court *(Art. 79)*.

1.3 Anonymisation vs. pseudonymisation

- **Pseudonymisation:** Identifiers are replaced by a code, but someone with the key can still link the data back to me. This is still personal data, so the GDPR applies *(Art. 4(5))*.
- **Anonymisation:** Data is changed so that no one can reasonably identify me anymore. Properly anonymous data is outside the GDPR *(Recital 26)*.

Easy rule: If the link to a person can be restored with reasonable effort, it is not anonymous; it is only pseudonymous.

2 Exercise 2: Fairness-aware AI

2.1 Setup (what the data-generating process was)

I simulated $n = 5000$ records. Each person has working hours `Hours` and a binary `Gender` (0 or 1). Salary is roughly

$$\text{Salary} \approx 100 \times \text{Hours} + \text{noise},$$

with small random noise. If we fit a simple linear regression of `Salary` on `Hours` only (without using gender), the slope should be close to 100 (because each extra hour pays about 100).

2.2 Three scenarios to test fairness claims

I edited the simulation to reflect three situations:

1. **(a) Direct discrimination:** For women (`Gender`=1) the salary is reduced by a constant amount (e.g., −200 euros). Hours distributions stay the same across genders.
2. **(b) No discrimination, different workloads:** Men and women have different hour distributions (e.g., men slightly more hours on average), but per-hour pay is equal.
3. **(c) Indirect discrimination:** Both (a) and (b) happen together: women get the −200 shift and also have a different hour distribution from men.

2.3 What happens to the slope?

- **Scenario (a):** The overall *slope* of the regression on `Hours` stays near 100. The model sees a vertical shift for the female group, which mainly changes the *intercept* if gender were included. Without gender in the model, the slope remains essentially the same because the "per-hour" pay still averages about 100 and the offset is uncorrelated with hours.
- **Scenario (b):** The slope also stays near 100 because pay per hour is unchanged. Different hour distributions move points left/right but do not change the hourly rate.

- **Scenario (c):** The slope *can change*. Here there is a group-specific penalty (-200) and group differences in hours. Because gender and hours are now correlated, omitting gender causes *omitted-variable bias*. The simple model tries to "explain" some of the -200 penalty using the `Hours` variable, slightly distorting the slope.

In symbols, with true model $\text{Salary} = \beta_H \text{Hours} + \beta_G \text{Gender} + \varepsilon$ and an hours-only regression, the estimated coefficient is

$$\widehat{\beta}_H \;=\; \beta_H \;+\; \beta_G \cdot \frac{\text{Cov}(\text{Gender}, \text{Hours})}{\text{Var}(\text{Hours})}.$$

In (a) and (b) the covariance term is ≈ 0, so the slope stays near $\beta_H \approx 100$. In (c) the covariance is nonzero, so the slope shifts.

2.4 How to detect indirect discrimination

To detect indirect discrimination, include the protected characteristic in the model (at least for auditing), for example:

$$\text{Salary} = \alpha + \beta_H \text{Hours} + \beta_G \text{Gender} + \varepsilon.$$

- If β_G is significantly negative (e.g., around -200), that suggests a systematic penalty for the protected group after controlling for hours.
- You can also add an interaction term $\beta_{HG}(\text{Hours} \times \text{Gender})$ to check whether the hourly rate itself differs by group.

For deployment, you might avoid using protected attributes for decisions, but you should still use them during development and auditing to test for unfair effects.

2.5 Takeaways

- The slope of an hours-only model changes only in scenario (c), where omitted-variable bias appears because gender (with a penalty) is correlated with hours.
- Auditing models with protected attributes helps reveal hidden gaps (group intercept shifts or different hourly rates) before removing those attributes for production.

References (short list)

- GDPR text: Regulation (EU) 2016/679 (General Data Protection Regulation). Commonly cited as Arts. 6, 7, 12–15, 77, 79; Recitals 26 and 32.

Chatgpt was used to generate this report based in the information in the jupyter notebook