

Web Scraping: Ethics and Legality (Week 1 Task)

1. What are “Terms of Use,” and why do they matter in web scraping?

“Terms of Use” (or Terms of Service) are legally binding rules set by a website.

They specify whether automated access (like scraping) is permitted.

Violating them can result in bans, account suspension, or legal action.

Respecting them reduces legal risk and shows professionalism.

Example: LinkedIn prohibits automated scraping in its Terms of Use.

2. Is web scraping legal?

Web scraping exists in a legal gray area, depending on jurisdiction, intent, and compliance.

Legal when: Data is public, scraping does not harm servers, and used for educational or research purposes.

Potentially illegal when: It violates Terms of Use, scrapes personal data, or bypasses restrictions like CAPTCHAs or logins.

Case: hiQ Labs v. LinkedIn (2019) ruled scraping public LinkedIn profiles did not violate the CFAA.

3. What are the ethical concerns of web scraping?

Server Strain: Overloading sites with requests disrupts service.

Data Ownership: Public data may still have copyright or database rights.

Privacy: Collecting sensitive or personal data without consent is unethical.

Transparency: Using data without attribution or misleadingly is unethical.

Respect for Site Rules: Ignoring robots.txt erodes trust.

Conclusion

Web scraping is powerful but must be done within legal and ethical limits. Always check Terms of Use and robots.txt, avoid harming servers, respect privacy, and use data responsibly. Done properly, scraping supports research and innovation without crossing ethical or legal lines.

References

- 1 Harvard Law Review (2020). The Law and Ethics of Web Scraping. <https://harvardlawreview.org/>
- 2 Electronic Frontier Foundation (2019). Web Scraping and the CFAA: hiQ v. LinkedIn. <https://www.eff.org/deeplinks/2019/09/ninth-circuit-rules-scraping-public-data-not-crime>
- 3 Towards Data Science (2021). Is Web Scraping Legal? <https://towardsdatascience.com/>