

# How to Build a Bot

## Course Information

Summer 2025

606 Pratt Manhattan Campus

M/Th 10:00 AM - 1:50 PM

5/27/2025 - 7/8/2025

Credits: 3

Prerequisites or other restrictions: INFO 664

## Instructor Contact Information

Filipa Calado, PhD (she/her)

Office Location: Pratt Manhattan Campus, room 602

Office Hours: by appointment (summer only)

Phone: 718-687-5194

Email: [fcalado@pratt.edu](mailto:fcalado@pratt.edu)

Zoom: <https://pratt.zoom.us/my/fcalado>

Website: <https://filipacalado.com>

## Course Description

This course offers a practical introduction to building bots in Python alongside a critical examination of algorithmic bias. Students will learn core programming skills like data collection and API interaction in order to create web crawlers and social media bots. In parallel, students will explore the ethical consequences of automated systems on social platforms, such as the amplification of misinformation and bias that perpetuate social inequalities and discrimination. This course will equip students with the critical perspectives and technical skills to analyze automated systems in a world increasingly shaped by AI and algorithmic decision-making. This course is intended for students interested in both technical development and the social impacts of automation.

## Student Learning Outcomes

Upon successful completion of this course, students will be able to:

- develop programming literacies for working with popular software in Python for web crawling and API usage.
- gain hands-on experience in computational data collection and parsing.

- identify and explain how choices in bot design, such as source selection, keyword filtering, and data processing, can introduce and/or amplify biases.
- critically assess the societal impact of bots, especially those used in social media, evaluating how bots might influence information visibility and public discourse.

## Course Format

This course will be held in hybrid format, with the first four weeks of meetings (May 29 - June 23) in person, and the last two weeks of meetings (June 26 - July 7) on zoom.

Our sessions will be split evenly between lecture, individual practice, and group work. This contrasts with the prerequisite course, INFO 664, which focused more heavily on lecture.

For each meeting, one 30-minute break will take place from approximately 12:00pm-12:30pm.

## course materials

Having a personal laptop (not a tablet) where you can install software is essential for this class.

All assignments and readings will be provided electronically and hosted on github at <https:gofilipa.github.io/how2bot>.

## Class Communication

The instructor will contact you via your pratt email (linked to Canvas). If you don't check that email frequently, please remember to do so for this class or set up mail forwarding.

The best avenue for contacting the instructor is via email, at [fcalado@pratt.edu](mailto:fcalado@pratt.edu). Response time should be within 2 business days, otherwise feel free to follow up.

## Course Schedule

### Unit 1 web crawling bots - 2 weeks

#### **May 29, session 1: intro to Python & web scraping with bs4**

Homework (due June 2) : find 2 scrapable sites

- find 2 websites to scrape. Make sure if they are scrapable with bs4. Why are you interested in this data? What could you do with it?

#### **June 2, session 2: scrapy & the scrapy shell**

Homework (due June 5): Compost Engineers chapters 1 & 2, and reading response (prompt below)

- Joana Varon and Lucía Egaña Rojas. Chapters 1 & 2 from Compost Engineers and Sus

Saberes Lentos: A Manifest for Regenerative Technologies. Coding Rights, 2024, [https://codingrights.org/docs/compost\\_engineers.pdf](https://codingrights.org/docs/compost_engineers.pdf).

- Prompt: Pick an idea from the reading that interests you (either because you agree with it, disagree with it, or are otherwise provoked by it) and explain why. 1 page.

### **June 5, session 3: blockers & XHR**

Homework (due June 9): Compost Engineers chapters 3 & 4, and reading response (prompt below)

- Joana Varon and Lucía Egaña Rojas. Chapters 3 & 4 from Compost Engineers and Sus Saberes Lentos: A Manifest for Regenerative Technologies. Coding Rights, 2024, [https://codingrights.org/docs/compost\\_engineers.pdf](https://codingrights.org/docs/compost_engineers.pdf).
- Prompt: From the authors' proposals, what do you find useful or surprising, and what do you have doubts about? 1 page.

### **June 9, session 4: selenium**

Assignment (due June 12): web scraping

- Using either scrapy or selenium, scrape some data from a website that you couldn't scrape before, and submit your python file(s) and the data on canvas.

## Unit 2 chat bots - 1.5 weeks

### **June 12, session 5: spaCy for processing text**

Homework (due June 16): ACLU tech & privacy analysis write-up

- Choose a topic from the last 6 months on the “Tech & Privacy” page on the ACLU website: <https://www.aclu.org/press-releases?issue=privacy-technology>
- Write up analysis of what is going on, and your opinion on the issue. How does the issue handle privacy rights and ethical uses of data? 1 page.

### **June 16, session 6: spacy continued, intro to transformers**

Homework (due June 19): run a task on your own data

### **June 19, session 7: transformers continued**

Assignment (due June 23): dataset proposal

- What is the dataset you'd like to create for your final project? Where would you get

the data, and how would you transform it? You can consider tools from this class (like text generation, named entity recognition, pattern matching), or you can consider other possibilities for transforming your data. 1 page, double spaced, submitted on canvas.

### Unit 3 social media bots - 1.5 weeks

#### **June 23, session 8: twitter bots**

Homework: make a plan for actions steps by next class

#### **(online) June 26, session 9: group projects**

Work on projects

#### **(online) June 30, session 10: group projects continued**

Homework (due July 3): project proposal due on Canvas

### Unit 4 project workshops & presentations – 1 week

#### **(online) July 3, final projects**

Work on projects

#### **(online) July 7, final project presentations**

### Assignments

Participation (30%)

- Includes in-class engagement, and completing and sharing homework

Unit assignments (30%)

- Average score of 3 assignments at the end of units 1-3

Final project: a bot! (40%)

- A final project that takes some data from web scraping or APIs, and uses it as the content

for a bot.

- bot to be automated and published on github.

## Recommended readings

On data gathering and web scraping:

- Dodge, Jesse, et al. “Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus.” Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, edited by Marie-Francine Moens et al., Association for Computational Linguistics, 2021, pp. 1286–305. ACLWeb, <https://doi.org/10.18653/v1/2021.emnlp-main.98>.
- Jo, Eun Seo, and Timnit Gebru. “Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning.” Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery, 2020, pp. 306–16. ACM Digital Library, <https://doi.org/10.1145/3351095.3372829>.
- Chan, Anita Say. Predatory Data: Eugenics in Big Tech and Our Fight for an Independent Future. University of California Press, 2025. library.oapen.org, <https://doi.org/10.1525/luminos.215>.
- Métraux, Julia. “Eugenics Isn’t Dead—It’s Thriving in Tech.” Mother Jones, <https://www.motherjones.com/politics/2025/01/eugenics-isnt-dead-its-thriving-in-tech/>. Accessed 14 Feb. 2025.

On machine learning:

- Alammam, Jay. The Illustrated BERT, ELMo, and Co. (How NLP Cracked Transfer Learning). <https://jalammam.github.io/illustrated-bert/>. Accessed 14 Apr. 2025.
- Alammam, Jay. The Illustrated DeepSeek-R1. 10 Feb. 2025, <https://newsletter.languagemodels.co/p/the-illustrated-deepseek-r1>.

Case studies of algorithmic bias & audits:

- Hada, Rishav, et al. “Akal Badi Ya Bias: An Exploratory Study of Gender Bias in Hindi Language Technology.” The 2024 ACM Conference on Fairness, Accountability, and Transparency, ACM, 2024, pp. 1926–39. DOI.org (Crossref), <https://doi.org/10.1145/3630106.3659017>.
- Gajjala, Radhika, et al. “Get the Hammer out! Breaking Computational Tools for Feminist, Intersectional ‘Small Data’ Research.” Journal of Digital Social Research, vol. 6, no. 2, 2, May 2024, pp. 9–26. jdsr.se, <https://doi.org/10.33621/jdsr.v6i2.193>.
- Tang, Ningjing, et al. “AI Failure Cards: Understanding and Supporting Grassroots Efforts to Mitigate AI Failures in Homeless Services.” The 2024 ACM Conference on Fairness, Accountability, and Transparency, ACM, 2024, pp. 713–32. DOI.org (Crossref), <https://doi.org/10.1145/3630106.3658935>.
- Groves, Lara, et al. “Auditing Work: Exploring the New York City Algorithmic Bias Audit Regime.” The 2024 ACM Conference on Fairness, Accountability, and Transparency, ACM, 2024, pp. 1107–20. DOI.org (Crossref), <https://doi.org/10.1145/3630106.3658959>.
- Costanza-Chock, Sasha, et al. “Who Audits the Auditors? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem.” Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery, 2022, pp. 1571–83. ACM Digital Library, <https://doi.org/10.1145/3531146.3533213>.