Problem A

The machine learning example I will discuss is an artificial intelligence model from the paper Rubin et al. (2021). This paper details a machine learning implementation for the action selection cortico-basal ganglia thalamic (CBGT) network in the brain. The CBGT network in the brain aids in decision making in animals, particularly for biased decisions or particularly for those decisions that need to incorporate learned or conflicting information. This network uses the neurotransmitter dopamine as a method to act as an accumulation of evidence for or against a particular decision. Therefore this facilitates a model whereby rewarded actions are more promoted while punished actions are further suppressed. The machine learning model in this paper is thereby a credit assignment model where for example the model made a left turn and was rewarded dopamine aka credit, and then made a right turn and was not rewarded dopamine it would then subsequently continue to make left turns.

This model is a simulation of artificial intelligence because it mimics the methodology by which humans and other animals make decisions. This model is also flexible and able to converge quickly upon a change in the systems where the left turn is no longer beneficial and adapt to make a new decision. This convergence and change is similar to neural reorganization where animals can change their cognitive states and as a result their choices based on changes in the environment. This model is an implementation of a machine learning training method called reinforcement learning, where the goal is for an agent to learn through trial and error by getting feedback based on their actions. The example given above is a form of Q-learning where a value Q is used by the model to determine where they should perform an action. In the example the Q value would be equivalent to the dopamine reward that is given when the initial optimal decision is to turn left.

References

Rubin, J. E., Vich, C., Clapp, M., Noneman, K., & Verstynen, T. (2021). The credit assignment problem in cortico-basal ganglia-thalamic networks: A review, a problem and a possible solution. European Journal of Neuroscience, 53(7), 2234–2253. https://doi.org/10.1111/ejn.14745

Zhou, X., Tien, R. N., Ravikumar, S., & Chase, S. M. (2019). Distinct types of neural reorganization during long-term learning. Journal of Neurophysiology, 121(4), 1329–1341. https://doi.org/10.1152/jn.00466.2018

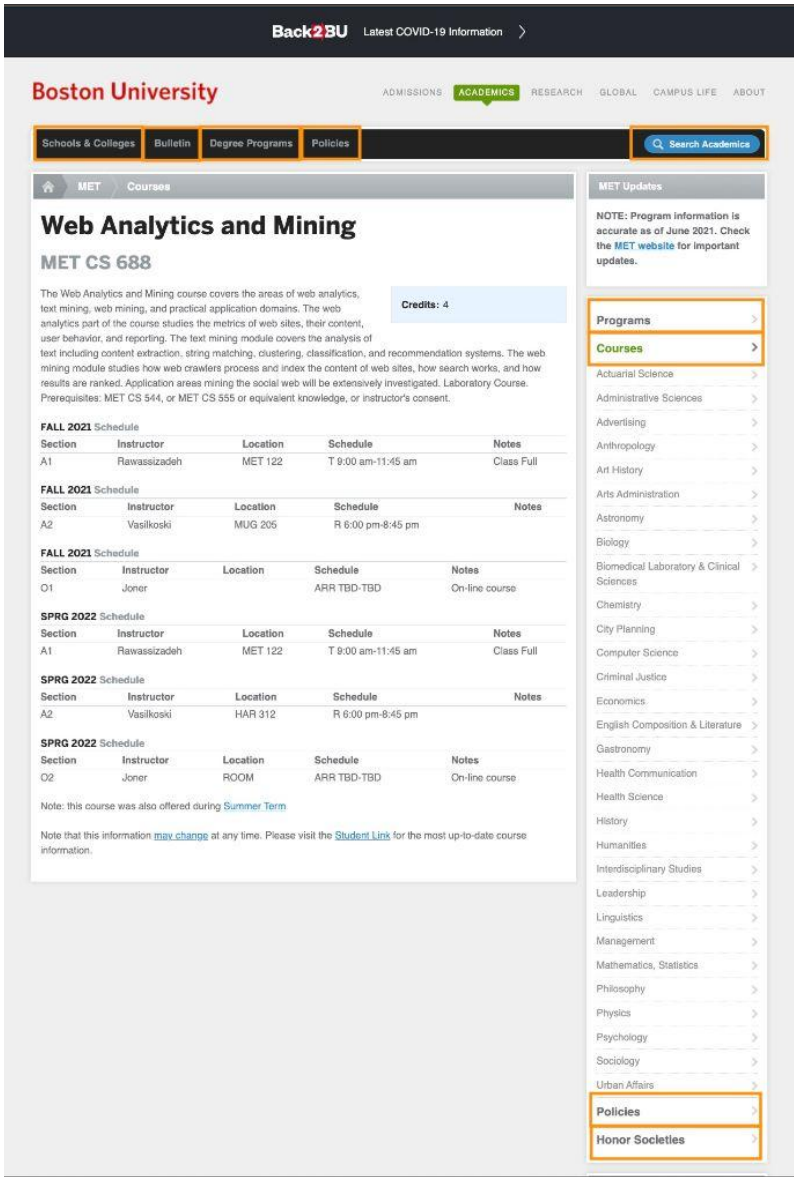
Bhatt, S. (March 19, 2018). Reinforcement Learning 101. Towards Data Science: https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292

Problem B

```
BeautifulSOUP web scraping from BU CS688 course page with tag class=level_1

[<a class="level_1" href="https://www.bu.edu/academics/schools-colleges/">Schools &amp;
Colleges</a>, <a class="level_1" href="https://www.bu.edu/academics/bulletin/">Bulletin<
/a>, <a class="level_1" href="https://www.bu.edu/academics/degree-programs/">Degree Prog
rams</a>, <a class="level_1" href="https://www.bu.edu/academics/policies/">Policies</a>,
 <a class="level_1" href="https://www.bu.edu/academics/search/">Search Academics</a>, <a
 class="level_1" href="https://www.bu.edu/academics/met/programs/">Programs</a>, <a clas
s="level_1 active" href="https://www.bu.edu/academics/met/courses/">Courses</a>, <a clas
s="level_1" href="https://www.bu.edu/academics/met/policies/">Policies</a>, <a class="le
vel_1" href="https://www.bu.edu/academics/met/honor-societies/">Honor Societies</a>]

Number of blocks extracted: 9
```



Problem C

I decided to study the hiQ Labs vs LinkedIn Corp web scraping case. In this case hiQ, which is a data analytics company scraped user data from the LinkedIn site. hiQ Labs is a data science based HR consultancy company, that aims to support companies in employee retention using data science backed knowledge. Their goal is to learn more about what keeps employees at companies in order to advise their clients in their employee retention methods. hiQ labs obtains this information by employing data scientists who use web scrapers and analyse the data, of which LinkedIn is a rich source for them to obtain data.

In May 2017, LinkedIn sent hiQ a cease and desist letter instructing them to stop scraping publicly available data from their website. LinkedIn then implemented techniques to limit hiQ's access to the publicly available data, afterwhich hiQ filed an injunction which the Ninth Circuit court then ruled in their favour. The court stated that LinkedIn did not have the right to inhibit hiQ labs from obtaining public data from their website. LinkedIn then filed a certiorari with the Supreme court, asking them to reconsider the decision made by the Ninth Circuit court. Their main argument was that even though some of their data is public hiQ labs was able to obtain far more information using bots than a human ever possibly could. The Supreme court granted them this opportunity and will be reviewing their case. Particularly there could be a consideration between public access data being able to be scraped and the limit at which a company has exceeded their access to the data.

This consideration of there being a limit to access to publicly available data was interesting to me. I immediately thought of those websites that make you prove you are a human before allowing you to proceed. While I knew this was to prevent automated bots from proceeding in the site, I thought it was due to the site wanting to protect their data. However, in my reading I learned sites will also do this to prevent automated bots from overwhelming their servers with a number of requests they would not ordinarily have. The potential for automated bots to crash their site and inhibit their users from having access to said site sounds like a fair use case for which to limit the amount of web scraping permitted on the site. In my opinion a downfall to this would be if the site then over-uses this reasoning for why they inhibit the web scraping of publicly available data.

References

Özcan, B. (Jan 21, 2022). Watch-outs for Legal and Ethical Web Scraping in 2022. AI Multiple: https://research.aimultiple.com/web-scraping-ethics/

Jouman Hajjar, A. (Jan 10, 2022). The Ultimate Guide to Web Scraping Challenges & Best Practices. AI Multiple: https://research.aimultiple.com/web-scraping-challenges/

hiQ labs website:
https://www.hiqlabs.com/new-who-we-are
https://www.hiqlabs.com/new-imwithhiq

hiQ Labs v. LinkedIn Corp. Reporters Committee:
https://www.rcfp.org/briefs-comments/hiq-labs-v-linkedin-corp/

Kesslen, M., McGee, K., & Sterba, B. (June 15, 2021). Supreme Court Grants Certiorari in Web Scraping Case hiQ v. LinkedIn. Lowenstein Sandler:

https://www.lowenstein.com/news-insights/publications/client-alerts/supreme-court-grants-certiorari-in-web-scraping-case-hiq-v-linkedin-tech-groupwhite-collar

Whittaker, Z. (June 14, 2021). Supreme Court revives LinkedIn case to protect user data from web scrapers. Tech Crunch: https://techcrunch.com/2021/06/14/supreme-court-revives-linkedin-bid-to-protect-user-data-from-web-scrapers/

Appendix

```
import sys
import requests
from bs4 import BeautifulSoup

# BU webpage for cs688
weblink = 'https://www.bu.edu/academics/met/courses/met-cs-688/'
page = requests.get(weblink)

# BeautifulSoup object
soup = BeautifulSoup(page.text, 'html.parser')

# find a class with tag "level_1"
mycls = soup.find_all("a", class_="level_1")
print("BeautifulSOUP web scraping from BU CS688 course page with
tag class=level_1\n")
print(mycls, "\n")
print(f'Number of blocks extracted: {len(mycls)}')

# print(soup("level_1"))
# sys.exit()
```