Dark patterns are design elements that webpages use to mislead, obscure, coerce and/or deceive users of that website into making choices that they would not normally make, without these patterns being present (McNealy, 2021). As dark patterns are becoming ever more prevalent throughout the internet, on all types of webpages, it is very likely that the vast majority of people who use the internet have encountered some form of dark pattern, whether they realise it or not.

Cognitive and psychological factors used in dark patterns can lead to compulsive buying behaviour. For example, the social validation, mentioned in the dark patterns dataset in the research from a crawl of 11k shopping websites (Mathur et al., 2019), which is classified as "social proof", manipulates the users by informing them of other people's behaviours. By doing so, it gives a feeling to users that they are shopping with others for this popular item, which can lead to extra spending than planned.

In this project, machine learning is applied to detect the dark pattern types that can be automatically/partially automatically detected based on the only the text. According to the dark pattern taxonomies in the research from Mathur et al. (2019), "Countdown Timer", "Limited-time Message", "Activity Message", "Low-stick Message", "High-demand Message" are the dark patterns we can automatically detected with our natural language processing method. Or according to the web-based dark pattern framework in the research from Curley et al. (2021), "Fake Activity", "Fake Countdown", "Ambiguous Deadlines", "Low Stock Messages", "Deceptive High Demand" are the dark patterns we can automatically detected with our natural language processing method.

1. The dark patterns that we are able to detect now using machine learning:
(Primary features, finishes before week 9)

Pattern Type	***Description***	***Detection***
Fake Activity	Informing the user about other people's activity on the website, including behavious of puchasing, viewing, visiting etc, which may not be truthful. (e.g., "3 people are viewing this item now")	Apply Natural Language Processing in machine learning / deep learning to achieve fully automatic detection of these 5 pattern types, based on the text only.
Fake Countdown	Using a countdown timer to alert users that a discount or deal is about to expire, which only purposely creates urgency for the purchase (e.g., "sale ends in 12h20m33s")	
Fake Limited-time	Giving users the impression that a deal or saleis only for a mimited amount fo time or is about to expire soon, without stating a specific deadline. (e.g., "sale ends soon", "only avaliable for a limited time")	
Fake Low-stock	Informing users about the limited availability of a product, making it more desirable to users. (e.g., "only 2 items left in stock")	

Fake High-demand	Informing users that the product is in high-demand and will sell out soon, thereby making it more attractive to users. (e.g., “this item is in high demand”, “selling fast”)	

2. The dark patterns we can try to detect using other data analysis techniques: (Secondary features, starting after week 9)

Pattern Type	***Description***	***Detection***
Confirmshaming	Invoking language and emotion (shame) to convince users not to make a certain choice, or guilting users into opting into something. (e.g., “No thanks, I don’t want to save.”)	(1) Gather the information of the buttons on the webpage. (2) Use the keyword “No” to filter for possible dark patterns content. (3) Data Analysis on the filtered content instances for detection. (detailed technique TBD)
Disguised Ads	In order to get users to click on them, advertisements appearing to be other forms of content or navigation, while actually these buttons are linked to external advertisements. (e.g., “Download” or “Next” button)	(1) Gather the information of the buttons on the webpage. (2) Apply OCR to determine the purpose of the buttons. (3) Check whether the button is linked internally within the domain or to an external site.
Trick Questions	Misleading users to make certain decisions based on the usage of confusing language, for example, using long and complicated double negative sentences. (e.g., “If you do not wish to be contacted via email, please ensure that the box is not checked.”)	(1) Gather the information of the buttons on the webpage. (2) Using Natural Language Processing for Double Negation Detection. (may be quite time consuming, so leave it at last if have time)

3. Challenging dark patterns that can be done in the future for fuether development: (Outside of our project for sure)
(To be listed after the above are done)

Curley, A., O'Sullivan, D., Gordon, D., Tierney, B., & Stavrakakis, I. (2021). "Give light, and the darkness will disappear of itself": The Design of a Framework for the Detection of Web-Based Dark Patterns. *ICDS 2021: The 15th International Conference on Digital Society*, 9.

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–32. https://doi.org/10.1145/3359183