Research Proposal Review

Considering your thoughts on your chosen area of interest for your project:

- Which of the methods described in this week's reading would you think would suit your purpose?
- Which data collection methods would you consider using?

Note that you may find that you could be using a mixture of both the research methods and the data collection methods. These considerations should be included in your presentation of the Project Proposal in Unit 10.

My working question (subject to change) for the research proposal is "Can machine learning be used to ethically reduce recidivism following imprisonment in England and Wales?"

This will be a descriptive study incorporating mixed methods:

Starting with a literature review. Ideally this will be UK focused, but could include wider studies if there is insufficient data in the UK alone.

1. A rapid assessment of available literature to explore the use of machine learning models to predict recidivism, and especially and ethical issues encountered.
2. Documentary research to discover the legislative and ethical considerations of the use of machine learning in the criminal justice system of England and Wales.
3. The literature review will iteratively spiral in to focus on a specific topic. I'm expecting to move towards a prediction model that could be used by the probation service to target interventions post release, but this will be determined by the materials discovered.

Once a use case has been identified and explored by the literature review, experimentation will be used to test one or more machine learning methods to determine if a useful model can be created within the ethical boundaries now defined.

1. Ideally genuine historical data will be used, although souring it will no doubt be challenging. There is anonymised data available from the Ministry of Justice here: https://www.gov.uk/guidance/ministry-of-justice-data-first. However it is a requirement to register first as a researcher and I'm not sure if I can meet all of the requirements to gain access. If I cannot access those data sets that other publicly available data from, for example, Kaggle will need to be used. Models will still be able to be build and concepts likely proven, but the findings will be less robust.

There should be no ethical concerns with the use of the MoJ data because it is already anonymised. Similarly, data from Kaggle will only be used if it is anonymised or synthesised, thereby presenting no ethical risks with the data itself.

There could be opportunities for further qualitative research by means of focus groups with the probation service, if access can be arranged, so seek input as to the practicality of implementing whatever recommendations are proposed.