Ethics

It is necessary to distinguish between two ethical concerns relating to this project. Firstly, the project must in itself be conducted in such a way that it lives up to generally accepted ethical standards for scientific research. Secondly, however, it is important to consider the potential consequences of successful implementation of a model of the type that the project seeks to construct. While the two are naturally intertwined, they are nevertheless distinct issues, and will consequently be analysed as such below.

# Research ethics

In addition to the general significance of sound research ethics, the subject matter of the present project makes it even more important to consider ethical questions. Criminal records are broadly considered to be highly sensitive personal information. This paper relies entirely on data compiled by government-controlled bodies already required to fulfil ethical standards. The API documentation explains that both the Information Commissioner’s Office and Data Protection specialists at the Home Office were consulted heavily in preparation for the launch of the API.[[1]](#footnote-1) A short review of the data has thus been judged to suffice for the present purposes.

The data gathered does not contain any information that can be used directly to identify individuals. Neither is it possible from the available data to deduce the identity of individuals with any non-random success. This is partly due to the very broad categorisation of personal characteristics of the individuals searched (e.g. age category above 34), and partly due to anonymisation by the police of location data. All coordinates given are closest “anonymous location” defined as:

1. Appears over the centre point of a street, above a public place such as a Park or Airport, or above a commercial premise like a Shopping Centre or Nightclub.

2. Has a catchment area which contains at least eight postal addresses or no postal addresses at all.[[2]](#footnote-2)

# The ethics of crime prediction software

The model this project attempts to construct draws on a number of personal characteristics for suspected offenders. It is generally settled in the literature on applied ethics that predictions on the basis of e.g. gender or age are ethically sound. However, there is vast disagreement on the question of *racial* profiling. In line with the literature on the subject it is assumed that the generalisations are factually sound, and that no other, better selection tool is available. These assumptions can reasonably be challenged, but under non-conforming circumstances, the main issue for racial profiling would not essentially be ethical.[[3]](#footnote-3)

Note firstly that racial profiling is not an easily settle question even in legal terms. In the US, the courts have ruled that profiling based *partly* on race is *not* in violation of the constitution.[[4]](#footnote-4) In the UK, Baker and Phillipson (2011)[[5]](#footnote-5) argue that laws allow for stricter scrutiny and a higher threshold for the permissibility of racial profiling. This nevertheless implies *a fortiori* that it is allowed in some circumstances. As such, the present project is not legally speaking on shaky ground.

Relative to the cases considered in the established literature in applied ethics, this project distinguishes itself primarily in the use of machine learning. It is thus necessary to ask both (1) whether racial profiling *simpliciter* is ethically defensible and (2) whether there are any salient features that distinguishes machine learning based racial profiling (MLRP) from human racial profiling (HRP)?

Modern normative ethical theory tends to fall into one of three categories: Consequentialism, deontological theories, and virtue ethics. Furthermore, contractualism has attracted some attention as (broadly speaking) a moderate approach in between consequentialism and deontology. Since virtue ethics almost tautologically focuses on individual morality and character development, this is judged to be irrelevant for the present purposes.[[6]](#footnote-6) Instead, an answer is sought from the remain three theories.

Generally speaking, the Kantian focus on the dignity of the individual makes deontology the easiest moral philosophy from which to condemn racial profiling. Any argument in favour of racial profiling based on deterrent effect would further be a classic example of treating humans as means rather than ends, which is categorically forbidden in Kantian deontology.[[7]](#footnote-7) Contractualism represents an intermediate step. While Applbaum (2014)[[8]](#footnote-8) and Hellman (2014)[[9]](#footnote-9) disagree on whether racial profiling can be allowed under a contractualist framework, their disagreement hinges on *empirical* facts about the current state of affairs in American racial tensions. Thus, if racial profiling can be defeated in a consequentialist analysis, it holds *a fortiori* that it generally cannot be defended.

In a consequentialist analysis, racial profiling would simply be allowed[[10]](#footnote-10) if it produced more positive outcomes than negative ones.[[11]](#footnote-11) It is a well-known problem in consequentialism that strong pain for a few individuals is outweighed by mild pleasure for many people. Building on the argument of Kennedy (1999), it is, however, doubtful whether this would in fact be the case for racial profiling. Kennedy rightly argues that those subjected to racial profiling are *repeat players* who on several occasions will have to endure the inconvenience of “false-positive searches”. Furthermore, Hellman’s objection (2014) can be adopted for a consequentialist analysis. She holds that false-positive searches based on racial profiling do not just cause inconvenience but demean and insult the relevant individuals. In consequentialist terms, she contends that historic factors have the effect of deepening the intensity of disutility felt by the innocent targets.

While the argument given above shows the depth of displeasure felt, it can also be argued that the positive consequences from racial profiling are limited. Note here that this project *only* deals with street-level stop and searches, and that the ethical discussion is limited to cases where no other approach is feasible. It holds almost *ex hypothesis* that the criminals who can *only* be stopped with this type of selection mechanism are minor offenders. Consider terrorism, major drug operations, or organised violent crime. For all of these types of crime, it is almost *a priori* true that their very nature make them enforceable through other means. As such, the benefits to the average citizen are arguably fantastically small.

The analysis above naturally forms the basis for the conclusion that racial profiling *simpliciter* is not morally defensible. This is undisputedly a controversial conclusion that would both find support and resistance in the existing literature. However, the argument greatly illustrates the salient features that determine in any individual case whether racial profiling can be allowed.

First, Kennedy’s argument: All other things equal, circumstances where false-positive stops are *more* likely to be endured by the same individuals make racial profiling *less* ethically defensible. Similarly for Hellman: In countries and cultures with *more* negative stereotypes about racial minorities, racial profiling is *less* morally defensible. Finally, under circumstances where *more* high-risk offenders are likely to escape apprehension without the use of racial profiling, racial profiling is *more* ethically defensible. This would e.g. mean that racial profiling (all other things equal) is *more* morally defensible in the UK than the US due to a different history of segregation and *more* defensible in risk-scenarios with greater likelihood of lone wolf terrorists than otherwise.

The difference between MLRP and HRP does not significantly influence neither the first nor the third point above. It is less clear how it impacts Hellman’s dignity objection. It could be argued that the use of machine learning removes the risk of generalisations carrying any historical or social meaning. However, this would miss the fact that these historical and social factors have shaped the data on which the machine is trained. Thus, using machine learning risks masking historical injustices as mathematical truths. In addition to this, MLRP based on biased data risks creating a vicious circle which in fact would *deepen* the racial injustice. If the police continuously retrains the model and the training data is created by a racially biased model, it would necessarily self-reinforce its bias. As such, the presence of historical or instutionalised injustices are especially important when considering MLRP as opposed to HRP.

1. <https://data.police.uk/about/#anonymisation> [↑](#footnote-ref-1)
2. <https://data.police.uk/about/#location-anonymisation> [↑](#footnote-ref-2)
3. Both Applbaum (2014) and Hellman (2014) make similar assumptions. See also Altman, Andrew, "Civil Rights", The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.) and references herein. [↑](#footnote-ref-3)
4. US vs. Weaver (1990), 8th circuit court of Appeals, No. 89-2887 [↑](#footnote-ref-4)
5. Baker, A., & Phillipson, G. (2011). Policing, profiling and discrimination law: US and European approaches compared. Journal of Global Ethics, 7(1), 105-124. [↑](#footnote-ref-5)
6. See e.g. Baron, Pettit, and Slote (1997) [↑](#footnote-ref-6)
7. <https://plato.stanford.edu/entries/kant-moral/> [↑](#footnote-ref-7)
8. Arthur I. Applbaum (2014) “Bayesian Inference and Contractualist Justification on Interstate 95”. In Cohen and Wellman, eds. *Contemporary debates in applied ethics*, 2nd edn., 219-231 (Wiley-Blackwell). [↑](#footnote-ref-8)
9. Deborah Hellman (2014) “Racial profiling and the meaning of racial categories”. In Cohen and Wellman, eds. *Contemporary debates in applied ethics*, 2nd edn., 232-243 (Wiley-Blackwell). [↑](#footnote-ref-9)
10. Potentially required. The discussion of supererogatory acts in consequentialism is beyond this discussion. [↑](#footnote-ref-10)
11. Negative consequences, displeasure, disutility, pain etc. (and the parallel for pleasure) are used interchangeably below. It is immaterial to the present discussion which welfare theory is used. [↑](#footnote-ref-11)