MEMORANDUM FOR ACTION

TO:	**The Minister of Foreign Affairs**
CC:	**The Digital Inclusion Lab, Office of Human Rights Freedoms and Inclusion**
SUBJECT:	Governing Autonomous Vehicles: Inclusive Design and Regulation

SUMMARY:

This memorandum presents the Minister with guiding principles to help construct policies to govern the ethical design and regulation of autonomous vehicles (AVs). The issue is how Canada can balance the regulation of AVs and its positive socio-economic potential with protecting human rights. Germany and the US provide two opposite regulatory approaches that Canada could adopt. The US approach leaves almost all ethical decisions around the design and governance of AV technology to the market/industry. By contrast, the German approach incorporates a code of ethics for AVs. Since Canada recognizes that AI and AVs could generate problems around how risk is distributed to vulnerable groups, it is more aligned with the German approach. But the German approach is informed by trolley problems and thus limited. Trolley problems, when applied in action, silence equity concerns and could lead to undesirable outcomes such as algorithmic bias.

   Algorithms play a key role in how autonomous systems function. If the bias lurking inside the algorithms that make important driving-related life-or-death decisions goes unchecked, it could lead to adverse consequences, especially for poorer communities and minorities. Discriminatory outcomes not only violate human dignity but also undermine public trust in machine learning and artificial intelligence.

   Canada can better respond to human rights concerns within the context of AVs without relying on trolley problems as a starting point. Canadian policy-makers should mandate transparent processes into AV design decisions and regulate the design of inclusive AI—with the aim to eliminate algorithm bias through equality problems and principles. If Canada wants to ensure that issues of equity, inclusion, and human rights are at the forefront, it must transcend the German and American approach to regulating AVs. Canada should pursue a middle ground approach to the regulation of AI. Justification for a middle ground approach can be discerned from analogous examples. For example, Canada's privacy framework is a middle approach between the US and EU regimes—sharing US concerns about "big brother" government while also expressing concerns about the private sector abuse of personal information. While limited to the privacy context and not AI, it is reasonable to expect that Canada take a middle ground approach to regulation in this context, and in fact, it should.

RECOMMENDATION(S):

- Ensure regulation plays role in governing algorithmic ecosystems—by mandating a code of design such that design decisions show evidence of reasonable ethical considerations and assist AV designers to be more aware of unwanted biases.

- Mandate a reasonable level of transparency and auditability to improve visibility into the design process and how ethics failures are addressed by manufacturers—where such decisions can be scrutinized by regulators and the public.

Selena Lucien B.A., M.A., M.Sc., JD (Candidate 2019)
Faculty of Law, University of Ottawa

☐ I wish to discuss   ☐ I concur   ☐ I do not concur

Minister

BACKGROUND:

1.   Automated technology is evolving rapidly and has potential wide-ranging socio-economic implications. A recent report by the federal government estimates that AVs could take root in as few as 10 to 15 years. Canadian Automated Vehicles Centre of Excellence and other experts suggest that AVs could be available as soon as year 2020. The Canadian Government's 2017 budget proposed $76.7 million, part of which would fund the development of "regulations for the safe adoption of connected and autonomous vehicles." Canada, however, has yet to develop a robust regulatory framework to deal with the governance and design of AVs.

2.   Germany released the first-ever code of ethics for AVs. Unlike the US approach (which removed ethical considerations from its mandate), Germany sets out several general ethical principles as a direct response to the Trolley Problem. The Trolley Problem exposes the moral tension between actively doing versus allowing harm: is it morally acceptable to kill one to save five, or should you allow give to die rather than actively hurt one?

3.   While trolley problems help to evaluate ethical theories (i.e. utilitarianism), their practical application to ethical design of AVs is overstated and produces a limited understanding of policy challenges ahead. Framing policy based on fictitious dilemmas is problematic because it silences the structural and practical problems that genuinely underpin the ethical problems facing AVs—such as algorithms reinforcing systemic bias and discrimination. A key challenge, according to AI Now Initiative, is how stakeholders— including the companies that develop and apply machine learning systems and government—can monitor and limit bias. Experts suggest that policymakers move away from framing regulatory responses as trolley problems and mandate more transparent processes that promote inclusive AV design decisions.

CONSIDERATIONS:

4.   Algorithms play a central role in how autonomous systems function and if left unchecked could lead to significant harm. Even if machine learning algorithms are trained on good data sets, their design could encode discrimination in numerous ways: choosing the wrong data, building a decision-making model with in advertent discriminatory features, absence of human oversight; unpredictable systems; or unchecked and intentional discrimination. Algorithmic bias is particularly worrisome for AVs because they do not need to involve a human "in the loop" who can detect and compensate for biases in the algorithmic model. By not involving the human, AVs might be regarded as neutral or impartial, but they are highly susceptible to biased algorithms. There are numerous examples of high-profile demonstrations of systemic racial bias in algorithms used to predict recidivism risk. These predictive algorithms are promoted as "more objective" and yet they exhibit systemic biases against particular racial groups (ProPublica, 2016).

5.   Glimpses of what might be on the horizon within the context of AVs is a string of errors of AI programs that uncannily favour the dark biases of humanity. For example, Google Image recognition labelled black people as gorillas, Snapchat and Instagram filters fail to recognize black faces, Apple Watch/Heart monitors/fitness trackers struggle to detect darker skin. Experts suggest that the development and design of algorithmic products, including AVs, must diversity of inputs, especially of the norms and values of specific populations affected by the output of AI systems.

6.   Similar to Google and Snapchat, the German approach presumes, that machine learning can distinguish between people and animals. Algorithms play a key role in determining what we see in the world, as well as how people see us. For instance, currently, AVs "driven" by algorithms are more likely to run over people with dark skin because it fails to detect/recognize them as human. The trolley problem does not address these dangerous algorithmic outcomes. It is not only the technologies that raise regulatory questions but also the perceptions and assessments of their social impacts by consumers and governments. Discriminatory outcomes not only violate human rights, but they also subvert public trust in machine learning. If public opinion becomes negative, it could lead to reactive regulation and stymie innovation and economic growth. Unless academics, technologists and other stakeholders establish a concrete process to hold algorithms and AV companies behind them accountable, Canadians are at risk.

7.   A fundamental barrier to algorithmic accountability is ensuring companies invest serious capital and personnel to ensure legal and ethical compliance. Government involvement in dictating AV design is risks stifling innovation, creativity, and technological progress. Further, there is no consensus around acceptable ethical decision-making or what it means for a company to provide their code to make algorithmic decisions more transparent, or what government board would be in charge to review them. Some argue that since algorithms are proprietary, so the answer to "who to put in jeopardy the driver or the passerby" is unknown and carries the bias of the individuals who design the algorithm-based system. Thus, the regulatory difficulty lies with how to obtain an algorithm's inner workings, especially when it is commercially sensitive and highly lucrative for companies to keep it a secret.

8.   By contrast, others argue that citizens have a right to an explanation about the decisions that affect them including the right to not be discriminated against. Therefore, the government as the direct representation of the public, has to be in a position to comprehend the decision-making processes of algorithms, and where necessary, to correct them. Canada's regulatory approach can be assisted by communities of researchers, businesses, and developers who share a mutual goal to design systems that maximize human benefit. For example, FATML (Fairness, Accountability and Transparency in Machine Learning), is one of many groups that can assist Canada in its commitment to regulate the design algorithms that are accountable and human-centric.

COMMUNICATIONS IMPLICATIONS/ACTIONS:

9.   Proposed course of action is likely to generate mostly positive media coverage. First, the popular press frequently publishes about the widespread unease about algorithmic bias in commonly used applications such as Snapchat, Instagram, Google, and FaceApp. In particular, the press generally highlights how algorithms transgress ethical redlines when their application involves minority groups in cases of hiring, lending, and face recognition; and relies on these examples to call for algorithmic regulation. Second, AI companies, such as Google (manufacturer of AVs), stress the importance of transparency in machine learning to avoid or diminish bias. Thus, the implementation of a balanced approach to regulation is likely expected to be met favourably by stakeholders. The recent Innovation Superclusters Initiatives by the Government of Canada includes an AI supercluster. The technological advancements stemming from this massive investment also highlights the need for an AI regulatory framework.