Political violence and a machine’s ‘superior ethical performance’
*
Some ethical questions for engineers (and us):

Do we know enough about how humans actually behave in their use of violence?

What does ‘behaving ethically’ mean? What does behaving ‘more’ or ‘less’ ethically mean?

Which warriors – fighting where, fighting when, fighting for what reason – will we assess when measuring human ethical performance?

Will we be tempted to underestimate humans’ ethical performance so it is more easily surpassed by machines?

How much time should we allow for AI decision-making to occur under real-world conditions while we wait to become satisfied that this matches or exceeds the ethical standard of human decision-making?

Are our definitions of key concepts (e.g. ‘civilian’) the same other people’s definitions?

What is the value of a human life? Is very human life of equal value?

What if there is a problem with the justification for resorting to violence in the first place?

If too much bad human behaviour is the problem, and AI’s better-than-human behaviour is a possible solution, is it a *better* solution than trying to improve human behaviour or reduce the risk of bad behaviour occurring?

If humans are ideally expected to be ‘100% ethical’ (perfect), why is it good to set for AI a lower standard (better-than-human)?

Can AI be praised and blamed? Can AI be meaningfully punished?