human rights and fundamental freedom to operate in practice. These threats point to the need to enhance and reinvigorate human rights discourse and protection in a data-driven age (section 3.5.1);

b) **Human-computer interaction:** Acute challenges arise in appropriately allocating and distributing responsibility between humans and machines, particularly when there is a 'human in the loop'. A recurring concern has been that, in order to ensure that complex socio-technical systems that incorporate AI always operate in the service of humanity, they should always be designed so that they can be shut down by a human operator. Yet individuals entrusted with the responsibility to supervise the operation of these systems may be understandably reluctant to intervene. This risks turning humans placed in the loop into 'moral crumple zones', largely totemic humans whose central role becomes soaking up fault, although they have only partial control of the system, and who are vulnerable to being scapegoated by tech developers and organisations seeking to avoid responsibility for unintended adverse consequences (section 3.5.2); and

c) **Interacting algorithmic systems:** Even more intractable challenges arise in seeking to identify, anticipate and prevent adverse events that arise from the interactions between complex, algorithm-driven socio-technical systems that can occur at a speed and scale that was simply not possible in a pre-digital, pre-networked age (eg the stock market 'flash crash' of 2010). The unpredictable nature of interactions between multiple algorithmic systems generates novel and potentially catastrophic risks, which we have barely begun to grasp, let alone anticipate and forestall (section 3.5.3).

All of these problems warrant further sustained attention and consideration.

While most of the discussion in Chapter 3 focuses on the responsibility of technology designers, developers and those who own and implement the systems which rely upon these technologies, the discussion in section 3.6 reminds us that it is states that bear the primary obligation to ensure that human rights are effectively protected. It draws attention to the problem of collective action that the operation of AI systems in a global networked age is likely to generate, highlighting the vital importance of a) national legislation to ensure that human rights are protected, b) the need for properly resourced national enforcement authorities with adequate enforcement powers and c) the valuable role which accessible and convenient collective complaints mechanisms, in addition to individual legal remedies, may play to ensure effective human rights protection.

The discussion then draws attention to a range of non-judicial mechanisms that have potential to help secure both prospective and historic responsibility for the adverse impacts of AI systems, including various kinds of impact assessment, auditing techniques and technical protection mechanisms (section 3.7). Technical protection mechanisms, in particular, have considerable promise. This study emphasises the need to embed these mechanisms within a governance framework that enables the relevant technical standards to be set in a transparent and participatory manner, and to ensure independent external oversight and review of their operation.

Before summarising the various findings in Chapter 3, the discussion in section 3.8 briefly considers whether our existing conceptions of human rights, and the mechanisms through which they are protected and enforced, are fit for purpose in a global and connected digital age. It suggests that the power of networked digital technologies that have emerged in recent