# Responsible AI Technologies, Artefacts, Systems and Applications
# 50 Principles

The following Principles are intended to be applied by the entities responsible for all phases of AI research, invention, innovation, dissemination and application. They were derived by consolidating elements from 30 international sources on 'Ethical Analysis and IT' and 'Principles for AI'.

## 1. Evaluate Positive and Negative Impacts

1.1 Conceive and design only after ensuring adequate understanding of purposes and contexts
1.2 Justify objectives
1.3 Demonstrate the achievability of postulated benefits
1.4 Conduct impact assessment
1.5 Publish sufficient information to stakeholders to enable them to conduct impact assessment
1.6 Conduct consultation with stakeholders and enable their participation
1.7 Justify negative impacts on individuals ('proportionality')
1.8 Consider alternative, less harmful ways of achieving the same objectives

## 2. Complement Humans

2.1 Design as an aid, for augmentation, collaboration and inter-operability
2.2 Avoid design for replacement of people by independent devices, except in circumstances in which artefacts are demonstrably more capable than people, and even then ensuring that the result is complementary to human capabilities

## 3. Ensure Human Control

3.1 Ensure human control over AI-based artefacts and systems
3.2 In particular, ensure control over autonomous behaviour of AI-based artefacts and systems
3.3 Respect each person's autonomy, freedom of choice and self-determination
3.4 Ensure human review of inferences and decisions prior to acting on them
3.5 Respect people's expectations in relation to personal data protections, incl.:
- awareness of data-usage
- consent
- data minimisation
- public visibility and consultation, and
- relationship of data-usage to the data's original purpose

3.6 Avoid deception of humans
3.7 Avoid services being conditional on the acceptance of AI-based artefacts and systems

## 4. Ensure Human Safety and Wellbeing

4.1 Ensure people's physical health and safety ('nonmaleficence')
4.2 Ensure people's psychological safety, by avoiding negative effects on any individual's mental health, inclusion in society, worth, standing in comparison with other people, or emotional state
4.3 Ensure people's wellbeing ('beneficence')
4.4 Mitigate negative consequences
4.5 Avoid violation of trust
4.6 Avoid the manipulation of vulnerable people, including taking advantage of individuals' tendency to addiction, e.g. to gambling

## 5. Ensure Consistency with Human Values and Human Rights

5.1 Ensure compliance with human rights laws
5.2 Be just / fair / impartial and treat individuals equally
5.3 Avoid unfair discrimination and bias, not only where it is legally procribed but also where it is publicly unacceptable

5.4 Avoid restrictions on freedom of movement

5.5 Avoid interference with privacy, family, home or reputation

5.6 Avoid interference with the rights of freedom of information, opinion and expression

5.7 Avoid interference with the right of freedom of assembly

5.8 Avoid interference with the right of freedom of association

5.9 Avoid interference with the rights to participation in public affairs and access to public service

**6. Embed Quality Assurance**

6.1 Invest in quality assurance

6.2 Ensure effective, efficient and adaptive performance of intended functions

6.3 Ensure security safeguards against inappropriate modification to and deletion of sensitive data

6.4 Ensure justification of the use of sensitive data

6.5 Ensure data quality and data relevance

6.6 Deal fairly with people (faithfulness, fidelity)

6.7 Avoid invalid and unvalidated techniques

6.8 Test for result validity

6.9 Impose controls in order to ensure that safeguards are operative and effective

6.10 Conduct audits of safeguards and controls

**7. Deliver Transparency and Auditability**

7.1 Ensure that the fact that the process is AI-based is transparent to all stakeholders

7.2 Ensure that the means whereby inferences are drawn, decisions made and actions are taken are logged and can be reconstructed

7.3 Ensure people are aware of inferences and how they were reached

**8. Exhibit Robustness and Resilience**

8.1 Deliver and sustain appropriate security safeguards against compromise of intended functions arising from both passive threats and active attacks

8.2 Deliver and sustain appropriate security safeguards against inappropriate access to sensitive data arising from both passive threats and active attacks

8.3 Conduct audits of justification, proportionality, transparency, mitigation measures and controls

8.4 Ensure resilience, in the sense of prompt and effective recovery from incidents

**9. Ensure Accountability for Legal and Moral Obligations**

9.1 Ensure that the responsible entity is apparent or can be readily discovered by any party

9.2 Ensure that effective remedies exist, in the form of complaints processes, appeals processes, and redress where harmful errors have occurred

**10. Enforce, and Accept Enforcement of, Liabilities and Sanctions**

10.1 Ensure that complaints, appeals and redress processes operate effectively

10.2 Comply with external complaints, appeals and redress processes and outcomes, including, in particular, provision of timely, accurate and complete information relevant to cases

## References

Clarke R. (2019a) 'Principles and Business Processes for Responsible AI' Computer Law & Security Review 35, 4 (2019) 410-422, PrePrint at http://www.rogerclarke.com/EC/AIP.html

Clarke R. (2019b) 'Principles for Responsible AI' Xamax Consultancy Pty Ltd, August 2019, at http://www.rogerclarke.com/EC/RAIC.html