# The Aletheia Framework™

| AREAS | CONTEXT | ETHIC | REALISATION PRINCIPLES | | EVIDENCE |
|---|---|---|---|---|---|
| Social Impact | Benefits | AI and robotics shall be seen as delivering good. Doing good is one of the five key ethical principles of the EU guidelines for ethical AI. Good includes commercial prosperity. | 1 | Deployment of AI and robotics shall be shown to improve the well-being of employees, such as improved safety, working conditions, job satisfaction. | Increase employee morale, engagement and responsiveness |
| | | | 2 | Additional to 1. (or instead of), deployment of AI and robotics shall be supported by a business case that demonstrates it improves competitiveness and is not just 'AI for the sake of AI'. | Business positions can be enhanced. Huawei products can be promoted through social media. |
| | Human impact | AI systems should be used to enhance positive social change and enhance sustainability. | 3 | For any new deployments, it shall be clear where the human boundary/interface/interaction with the AI/robotics system starts and ends. | Make better business decisions and build a strong strategy. |
| | | | 4 | Very early analysis, in conjunction with human resources and employee representatives, shall be undertaken to identify potential job role changes or potential human resource impacts and the opportunities for retraining. | Potential to add roles such as Social Media Analyst |
| | | | 5 | Potential for upskilling opportunities shall be explored with human resources and employee representatives as soon as any impact on affected employees is established, to ensure that the organisation has the key capabilities needed to secure emerging opportunities in AI and robotics. | Potential add roles such as Social Media Analyst |
| | | | 6 | Analysis shall be undertaken to assess the impact of the deployment on the external supply chain – particularly assessing the likelihood for the technology to have a negative impact on the sustainability of elements of the external supply chain. | Not applicable |
| | | | 7 | Where there is potential for negative impact on the sustainability of the external supply chain, this shall be discussed with the external supply chain partner as soon as possible to give them maximum opportunity to adapt to remain sustainable. | Not applicable |
| | Communication | Knowledge of the human interactions with AI should be provided by key stakeholders. | 8 | Frequent communication and discussion should be had with all key stakeholders – in particular employees and employee representatives – through a variety of channels. | Customer experience and satisfaction can be improved. Brand awareness can be strongly built. |
| | Loss of skills | AI systems should be used to enhance positive social change and enhance sustainability. | 9 | Analysis shall be undertaken as to whether any loss/reduction of skills needs to be sustained and how this would be addressed. | Skills will not be depleted. Current skills and capabilities will be enhanced through social network analysis. |
| Accuracy/Trust | Safety | AI systems should be safe and secure throughout their operational lifetime. This should be verified where applicable and feasible. | 10 | A Process Failure Modes and Effects Analysis (PFMEA) shall be undertaken with specific emphasis on identifying and mitigating any hazards to human safety. | Not applicable |
| | Transparency and traceability | AI systems must provide for transparency and traceability of their design, inputs and outputs. | 11 | The provenance of the algorithms shall be clearly stated to enable any future Root Cause Analysis (RCA). | Not applicable |
| | | | 12 | The provenance of all training data shall be clearly stated to enable any future RCA. | Not applicable |
| | | | 13 | The hierarchy of decision making shall be clearly stated regarding human v AI. | Human intelligence will dominate artificial intelligence. |
| | | | 14 | It shall be clear what the insight (forecast/decision making etc.) improvement is compared with a human – forecast improvement and actual. | . |
| | Bias | AI systems must be free from bias or prejudice. | 15 | It shall be clearly stated how any training data sets have been assured to have no unintentional or unethical biases, noting that, for example, if an AI sub-system is being used to detect anomalies, the training set may need a deliberate bias to ensure sufficient amounts of anomalies occur at different rates. | Small data set can result in significant bias if outliers are existent in the data set. |
| | Validity and reliability | For AI to succeed it must be trusted. | 16 | A monitor shall be deployed in the system – essentially a **sense check** of the results comparing inputs with likely outputs for the system in question. | . |
| | | | 17 | A **continuous** automated monitor shall be deployed in the system – by using existing test/synthesised data which already has approved outputs. | An existing dataset should be used to continuously monitor the outcome. |
| | | | 18 | An **independent** check shall be deployed in the system – assessment of the same data using a completely independent assessment mechanism which is already approved. This is a validation check. | An independent check should be done by the third party with no common interest or associate |
| | | | 19 | A **process** comprehensiveness check shall be deployed in the system – have the right number of assessments taken place? | There should be an internal audit for the process comprehensiveness check. |
| | | | 20 | A **faultless** transmission of data shall be deployed in the system – use of Cyclic Redundancy Checks (or equivalent) where appropriate. | Not applicable |
| | Sparse data interpolation | For AI to succeed it must be trusted. | 21 | The sparseness of the training set of data and its impact on the validity of the output needs to be clearly stated and justified. | Not applicable |
| Governance | Data protection | For AI to succeed it must be trusted. | 22 | It shall be stated whether there is, or will be, any Personal data or not. | Not applicable-No personal data |
| | | | 23 | The legitimate purpose for using the Personal data shall be declared and confirmation provided that this has been agreed with the person or employee representative where it refers to an employee. | Not applicable-No personal data |
| | | | 24 | The architecture of the system shall protect the data from unwanted access without permission. | Not applicable-No personal data |
| | | | 25 | The architecture of the system shall have the facility to, on demand, identify an individual's personal data and update, amend or remove every trace in line with privacy requirements and individuals' rights. | Not applicable-No personal data |
| | | | 26 | No Personal data shall be sent outside of the relevant, legal zone (e.g. European Economic Area, US). | Not applicable-No personal data |
| | Export control | For AI to succeed it must be trusted. | 27 | The data flows (including access/reading of data) shall be described to, discussed with and approved by an Export Control manager to assure compliance with Export Control regulations. | Not applicable-No personal data |
| | Confidential information | For AI to succeed it must be trusted. | 28 | All confidential information shall be declared to, discussed with and the architectural protections approved by a certified IT security expert. | Not applicable-No personal data |
| | Cyber security | For AI to succeed it must be trusted. | 29 | All systems shall be assessed and approved by a certified IT security expert. | Passwords and/or authorized users to set for access |
| | Accountability | Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. | 30 | Ultimate accountability for the outcomes of the AI system needs to be clearly stated with a business owner clearly identified. | Different hierarchies of data management, maintenance and oversight should be established |
| | Responsibility for decisions | Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. | 31 | Algorithmic accountability should fall jointly on the developer and tester, or the DevOps team. They shall clearly state how they have assured confidence in the performance of their individual aspects of the AI system. | The developer team should take the full responsibility of algorithm. |
| | Risks from re-use/transfer across processes | For AI to succeed it must be trusted. | 32 | Any transfer of knowledge between AI systems shall be fully risk assessed by undertaking PFMEA (in addition to that in Realisation Principle 10) and identifying the major severity effects and causes, along with the detectability mechanisms for the proposed controls – which shall be formally reviewed before proceeding. | Risk assessment and management are critical to AI control. |

Version 1.0