The Aletheia Framework™ 2.0

AREAS	CONTEXT	ETHIC		REALISATION PRINCIPLES	EVIDENCE – FOR THE INTELLIGENT BOROSCOPE PROJECT
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 and/or the general public, such as improved safety, working conditions, job satisfaction.	The tools optimise the on-wing component inspection by standardising the inspection process, reducing the inspection effort/ time, and assisting the inspector in processing t he data. The processed data are used for reducing maintenance frequencies through high quality data acquisition and analytics. Conventional image processing techniques have been prior investigated with performance comparison studies to conclude with selecting a customised neural network model.
			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'.	Borescope inspections take a long time and require working in unpleasant environment conditions - next to the aircraft (50 degree/ cold/winds/rain, airport, climbing inside the fancase) - and good effort to fit the adapters, position the borescope with a high manual articulation and rotating the engine stage manually. The new method reduces a) the burden by providing assistance to the operator while navigating to the target borescope view (displayed overlay features on the screen), b) the inspection and processing time by using an integrated system.
	Human impact	AI systems should be used to enhance positive social change and enhance sustainability.	3	For any deployments, it shall be clear where the human boundary/interface/ interaction is with the AI/Analytics/Robotics system; and any negative/ positive impact on human factors and/or human behaviours is fully understood and mitigated where necessary.	The developed tools are a semi-automated process and require continuous human interventions for selecting/adjusting/approving the detected features by the AI algorithm. After processing the data the inspector will make the final decision and sentence the engine. The inspector will be always watching the system while the data is captured with the tool.
			4	Early analysis, in conjunction with human resources and employees (or their representatives), shall be undertaken to identify potential job role changes or potential human resource impacts and the opportunities for retraining or redeployment.	No changes on the manpower. The method still requires two operators to perform the inspection. And under the safety and reliability compliance two operators are required to cross check and overview the process. The new method has been communicated to the wider groups and reviewed under the external marketing campaign.
			5	Potential for upskilling opportunities or redeployment shall be explored with human resources and employees (or their representatives) when any impact on affected employees is established, to ensure that the organisation has the key capabilities neewded to secure emerging opportunities in AI a nd robotics.	There are available training videos and materials for the method. In addition, an advanced course for Intelligent Borescope is under review in inclusion for training materials.
			6	Analysis shall be undertaken to assess the impact of the deployment on the supply chain – particularly assessing the likelihood for the technology to have a negative impact on the sustainability of any elements of the supply chain. The same assessment should be afforded to customers as appropriate.	The used devices in this technique have higher inspection capability and contains state-of-art features and Integration required for the inspections specs, in comparison to what is provided by other suppliers.
			7	Where there is potential for negative impact on the sustainability of the supply chain, this shall be discussed with the supply chain partner as soon as possible to give them maximum opportunity to adapt to remain sustainable. This same opportunity should be afforded to customers as appropriate.	A letter detailing the strategy will be communicated with partners. The new technology encourages other suppliers to advance in the technology.
	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 (which result/cannot be avoided) needs to be sustained, for the good of the business, and how this would be addressed.	There are no loss/reduction skills. The tools drive for upskilling and learning new methods.
Accuracy/Trust	Safety	AI systems should be safe and secure throughout their operational lifetime. This should be verified where applicable and feasible.	10	A formal risk analysis shall be undertaken with specific emphasis given to identifying and mitigating any hazards to human safety.	A risk review document has been compiled with risk scoring and actions to mitigate the raised risks.
	Transparency and traceability	AI systems must provide for transparency and traceability of their design, inputs and outputs.	11	The algorithms shall be assessed for any bias or discrimination impact and their provenance shall be clearly stated to enable any future Root Cause Analysis or troubleshooting (Note, for complex systems, it may be difficult to assess the risk of bias. A new bias assessment template has been created as part of an ecosystem of AI Assurance tools at [link]).	Development of the algorithms follows the Rolls-Royce software standards.
			12	To enable the power of data to be unlocked, all training data shall be good quality and representative and its provenance shall be clearly stated to enable any future Root Cause Analysis or troubleshooting.	Full provenance of the data has been formally captured and recorded.
			13	The hierarchy of decision making shall be clearly stated regarding human v AI.	They are assistant tools that help in extracting measurement and analysis information to help the operator in making the decision. See 3.
			14	It shall be clear what the insight (forecast/decision making etc.) improvement is compared with a human – forecast improvement and actual.	The validation and comparison results have demonstrated that the tools reduce the processing time by >75% in comparison to the old manual process. https://www.rolls-royce.com/media/our-stories/discover/2021/intelligentengine-harnessing-the-power-of-ai-to-deliver-more-intelligent-engine-inspections.aspx
	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.	The sampling campaign involved performing inspections on a diverse engine population including different lifecycles, operations, airlines, kits and inspectors.
	Validity and reliability	For AI to succeed it must be trusted.	16	A monitor shall be deployed in the system – this is a sense check of the results comparing actual outputs with likely output ranges for the system in question.	Fail safe checks are implemented in the tool to notify the inspector if the output is out of family, or if it is outside the threshold limits.
			17	A continuous automated monitor shall be deployed in the system to continuously test the system by using existing test/synthesised data, which already has known and approved outputs.	A benchmark dataset of manually processed batch has been compiled as ground truth comparison for continuous checking.
			18	An independent check shall be deployed in the system – assessment of a proportion of the same data using a completely independent assessment mechanism which is already approved. This is a validation check and could be carried out by a human.	The outcome is validated in four steps: 1- comparing the results with benchmark dataset; 2- testing the accuracy of the results; 3- running tests on new data not seen in the model training/validation sets; 4- Reproducibility and Repeatability.
			19	A process comprehensiveness check shall be deployed in the system – have the right number of assessments taken place?	A colour theme (red/green) is used to distinguish between actions in the review and approved phases. Notfication messages are used and will notify if the number of checks is less than the number of components in the set. Fail safe warning messages are also in place.
			20	A faultless transmission of data shall be deployed in the system – use of Cyclic Redundancy Checks (or equivalent) where appropriate.	The intelligent borescope uses inspection processes to store data in a structured format with metadata information. A csv log file is generated at the end of the process containing the results and details the points where auto detected or included manual adjustments.
	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.	The data collected in the sampling campaign were used to train the model. It represents variations across different levels. See 15.
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.	No Personal data used.
			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.	No Personal data used.
			24	The architecture of the system shall protect the data from unwanted access without permission - complying with the principle of 'privacy by design and by default'.	No Personal data used.
			25	The architecture of any data storage system should 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.	No Personal data used.
			26	No Personal data shall be sent outside of the relevant, legal zone (e.g. European Economic Area, US).	No Personal data used.
	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.	Data is assessed for Export Control compliance using the corporate process.
	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 an IT security expert.	All confidential information is assessed through the corporate information assurance process.
	Cyber security	For AI to succeed it must be trusted.	29	All confidential information shall be declared to, discussed with and the architectural protections approved by an IT security expert.	All systems are assessed through the corporate information assurance process
	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.	The tools are semi-automated and require interventions by the inspector to review/ adjust/approve the detected features. The final decision and the engine sentencing is done by the inspector.
	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.	Validation and confidence evidence are described in RR documents, including the steps undertaken to ensure robust performance.
	Risks from re-use/transfer across processes	For AI to succeed it must be trusted.	32	Transferring knowledge between AI systems should be risk assessed using a formal tool/method to determine where and how the system might fail. Any serious events and their causes must be identified along with the method to detect such events. – which shall be formally reviewed before proceeding.	The tools are specifically designed and tailored to a pre-defined inspection task. Any transfer of technology will require a reconfiguration to the tools processing logic and retraining of the models with different dataset.