|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Risk Factors Arising in the System’s Application Context** | | | | | |
| **Questions** | | | **Governance interventions and points to be addressed when assessing potential adverse effects on human rights, democracy and the rule of law** | **Response** | **Completed by** |
| **Sector or domain in which the system is being built** | 1 | Will the AI system serve an essential or primary function in a high impact or safety critical sector (e.g., transport, social care, healthcare, other divisions of the public sector, see **Appendix II** for the indicative list of other sensitive sectors/domains)?  **Example:** a hybrid AI system used to control the kinematics of a surgical robot assists doctors who are performing an emergency medical procedure. | Make sure to focus upon considerations surrounding the prevention of harm to personal, physical, mental and moral integrity and, as applicable, other protected areas/spheres from **Appendix I**. |  |  |
| 2 | Does the sector or domain in which the AI system will operate include vulnerable groups (or groups with protected characteristics) who may be significantly or disproportionately impacted by the design and use of the system?  **Example:** a risk assessment tool that is used by social workers to identify children in need of state-assisted care operates within the social care system, where children (a vulnerable group) are significantly impacted. | Make sure to focus upon considerations surrounding the potential impacts of your AI system on the right-holders from the point of view of the principle of equality and non-discrimination, with specific reference to Article 18 of the Framework Convention.  Pay close attention to risk factors referred to in questions 8 to 15 in **COBRA Resources B.** |  |  |
|  | 3 | Is the AI system likely to be accessed by or impact upon the lives of children? | Make sure to focus upon considerations surrounding the potential impacts of your AI system on the right-holders from the point of view of the principle of equality and non-discrimination, with specific reference to Article 18 of the Framework Convention.  Special consideration should also be given to the unique set of rights held by children as well as to the best mechanisms for meaningful stakeholder engagement with children prior to the design of the AI system. |  |  |
| **Existing Law and Regulatory Environment** | 4 | Have you assessed existing law and regulation in the sector or domain in which the AI system will operate and determined that it has a sufficient legal basis and can be developed and deployed lawfully? | You should determine, through expert input where appropriate, whether your project has a legal basis. |  |  |
| 5 | Is the sector or domain in which the AI system will operate historically highly regulated?  **Example:** a bank using a risk assessment tool to predict borrower default operates within the financial sector where extensive regulation pertaining to market abuse, risk management, equity law, and competition law has historically been in place. | You should determine, through expert input where appropriate, whether any relevant existing regulation in your sector or domain applies to the prospective AI system. |  |  |
| 6 | Does statute or regulation in the sector or domain in which the AI system will operate require any other types of impact assessment for the specific use-case of the AI systems you are planning to develop?  **Example:** a healthcare provider developing a risk assessment tool that uses genetic and health data is required to conduct a data protection impact assessment due to its processing of sensitive data and data concerning vulnerable subject. | Make sure to integrate the completion of other compulsory impact assessments as supplementary material and evidentiary support for your HUDERIA assessment. |  |  |
| 7 | Will the AI system perform a safety critical or high impact function independent of the sector in which it operates?  **Example:** the system being built is an AI-assisted HR recruitment tool for a civil service function. The tool is not situated within a safety critical sector; however the system could negatively impact on vulnerable groups. | Make sure to focus upon considerations surrounding the prevention of harm to personal, physical, mental and moral integrity and, as applicable, other protected areas/spheres from **Appendix I**. | **TEMPLATE** |  |
| 8 | Could the AI system be repurposed or used in ways that fall under the list of prohibited systems?  **Example:** a computer vision system used to recognise individuals’ faces as a form of identity authentication is repurposed for real-time remote biometric identification of petty criminals at a concert venue. | You should take special precautions to put in place protections, processes, and mechanisms that ensure the limitation of the system’s use to the purpose for which it was created. |  |  |
| **Scope of Deployment** | 9 | In a scenario where your project optimally scales, will the AI system directly and/or indirectly affect rights-holders and groups within local populations, national populations, global populations?  **Example (local):** a classification system used by property developers to identify sites suitable for construction within a region affects individuals and groups within local population.  **Example (national):** a classification system used by governmental organisation to determine individuals’ eligibility for social benefits affects individuals and groups within national populations.  **Example (global):** a recommender system used by an international social media platform to personalise content delivery affects individuals and groups within the global population. | In assessing the potential impacts of the system, you should pay special attention to the potential effects of the AI system on local populations and on the communities and groups that comprise them as well as social and political processes and institutions—in particular, as these relate to human rights, democracy, and the rule of law. |  |  |
| 10 | In a scenario where your project optimally scales, will the AI system directly and/or indirectly affect between 1 and 10000 rights-holders, between 10001 and 100,000 rights-holders, between 100,001 and 1,000,000 or over 1,000,000?  **Example (1-10,000):** a conversational AI system is used within a company of 70 employees to answer employees’ frequently asked questions.  **Example (10,001-100,000):** a clustering system is used by a marketing agency to segment customers, identifying target groups for an advertising campaign targeting 12,000 potential attendees for an event.  **Example (100,001-1,000,000):** a regression-based system is used by an energy company with 800,000 customers to predict their usage and inform personalised tariffs.  **Example (over 1,000,000):** a relevance ranking system based on user interaction data is used by a social media platform to personalise content for its one billion monthly active users. | In assessing the potential impacts of the system, make sure to pay special attention to the *at scale* or *mass-level* effects of the use of the AI system on social and political processes and institutions—in particular, as these relate to human rights, democracy, and the rule of law. |  |  |
| 11 | Considering the potential direct and indirect impacts of your project on individuals, communities, and the environment, which is the widest timescale within which the AI system could affect rights-holders and groups: short term (less than a year), medium term (1 to 10 years), generationally (10 to 20 years), long-term (20 to 60 years), over lifetimes and across future generations (over 60 years)?  **Example (short term):** a conversational AI system used by a retail company to automate a limited number of customer service tasks may have short term effects on individuals and groups.  **Example (medium term):** a clustering system used by marketing agencies to segment customers, identifying target groups for an advertising campaign may have medium term effects on individuals and groups.  **Example (generationally):** a regression-based system used by energy companies to predict customers’ usage and inform personalised tariffs may have generational effects on individuals and groups.  **Example (long-term):** a classification system used to categorise immigration applications (visa, residential status, and citizenship applications) as low, mid or high risk of being fraudulent may have lifetime effects on individuals’ immigration status  **Example (future generations):** a machine vision system used to predict long-term passanger needs across modes of travel informing a transport infrastructure plan may have effects across future generations. | In assessing the potential impacts of the system, make sure to pay special attention to the longer term effects of the use of the AI system on the rights holders as well as on social and political processes and institutions—in particular, as these relate to democracy and the rule of law. |  |  |
| **Existing legacies of bias** | 12 | Do the sector(s) or domain(s) in which the AI system will operate, and from which the data used to train it are drawn, contain historical legacies and patterns of discrimination, inequality, bias, racism, or unfair treatment of minority, marginalized, or vulnerable groups? Could these patterns and historical legacies be replicated or augmented in the functioning of the system or in its outputs and short- and long-term impacts?  **Example 1:** the use of a job application screening system in a science and technology industry based on a historic hiring data replicates patterns of hiring discrimination, delivering unfavourable outcomes for vulnerable groups.  **Example 2:** a classification system used to identify individuals eligible for cervical screenings fails to identify elgibile transgender individuals. The data used to train the system did not represent this population, replicating historic inequalities in healthcare. | Make sure to focus upon considerations surrounding the potential impacts of your AI system on the right-holders from the point of view of the principle of equality and non-discrimination. |  |  |
| **Environment context** | 13 | If the design, development, and deployment of the AI system will have potentially significant impacts on the environment, will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure that the system, in both its production and use, complies with applicable environmental protection standards and supports the sustainability of the planet?  **Example:** a neural network used to predict instances of a specific disease with genetic data is hosted on a cloud server and runs constantly throughout the day over several months, processing over 100TB of data. This system contributes to the production of emissions. | You should also explore in your HUDERIA whether any human rights may be affected by proxy (see number 30 in **Appendix III**)—paying close attention to effects on vulnerable and marginalized groups. |  |  |
| **Due diligence** | 14 | Do you require your vendors and suppliers to demonstrate compliance with legal protections in place to protect rights-holders from modern slavery, human trafficking, and labour exploitation? | *In assessing the potential impacts of the system*, it is important to ensure the effective maintenance of human rights due diligence with regard to all businesses, parties, and entities directly linked to your project lifecycle. |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Risk Factors Arising in the System’s Design and Development Contexts** | | | | | |
| **Questions** | | | **Governance interventions and**  **points to be addressed in when assessing potential adverse effects on human rights, democracy and the rule of law** | **Response** | **Completed by** |
| **Technological Maturity** | 1 | Will the AI system's design be based on well-understood techniques that have previously been in operation and externally validated for a similar purpose and in the same sector? | **Governance interventions**  a)  Diligent processes of testing, verifying, and externally validating the performance of the system should be put in place.  b)  System monitoring and performance evaluation protocols that are proportionate to the system’s technological maturity should be put in place.  **Points to be addressed in COBRA**  Make sure to consider, if applicable, the effect of the system’s technological maturity on the quality of its performance (the principle of reliability) when assessing the potential impacts of your AI system on the right-holders. |  |  |
| **Existing system** | 2 | If the AI system is replacing a human, technical, or hybrid system that serves the same or similar function, is a reason for replacement that the existing system is considered flawed or harmful? | **Governance interventions**  a)  Assessment has to be made as to how your AI system could potentially repeat the harmful outcomes generated by the replaced system and how to minimise any potential replicated harms.  b)  The flaws or harms generated by the previous system should be redressed and rectified across the design, development, and deployment lifecycle of the new AI system.  **Points to be addressed in COBRA**  Make sure to consider, if and where applicable, the existing system’s flaws on the quality of the AI system’s performance (the principle of reliability) when assessing the potential impacts of your AI system on the right-holders. |  |  |
| 3 | If the human, technical or hybrid system being replaced is considered critical infrastructure or serves a safety-critical or high impact function, has assessment and planning been done and made public to transparently ensure that its updating/replacement does not cause unacceptable outage or harm? | **Governance interventions**  Publicly transparent assessment and planning should be put in place which demonstrates that the updating/replacement of the previous system will not cause unacceptable outage or harm.  **Points to be addressed in COBRA**  Make sure to consider, if applicable, the implications of the replaced system’s critical nature when assessing the potential impacts of your AI system on the right-holders. |  |  |
| **Cybersecuriy context** | 4 | Could the AI system present motivations or opportunities for malicious parties to hack or corrupt it to achieve substantial financial gains, political goals, or other perceived benefits? | **Governance interventions**  Sufficient and transparently reported processes should be put in place throughout the project’s lifecycle to safeguard the system's safety, security, and robustness. These measures should be appropriately proportional to potential risks of hacking, adversarial attack, data poisoning, model inversion, or other cybersecurity threats.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 5 | Will sufficiently and transparently reported processes be implemented throughout the project’s lifecycle to ensure that measures put in place to safeguard the system's safety, security, and robustness are appropriately proportional to potential risks of hacking, adversarial attack, data poisoning, model inversion, or other cybersecurity threats? |  |  |
| 6 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to stress test the AI system for cybersecurity vulnerabilities and resilience? | **Governance interventions**  Sufficient and transparently reported processes should be put in place throughout the project’s lifecycle to stress test the AI system for cybersecurity vulnerabilities and resilience. |  |  |
| **Points to be addressed in COBRA**  Make sure to consider, if and where applicable, the implications of cybersecurity vulnerabilities when assessing the potential impacts of your AI system on the right-holders. | | |  |  |
| **Data lifecycle: Data Quality, Integrity, Provenance and Protection** | 7 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure that all data used in producing the system are sufficiently balanced and representative of the individual rights-holders and groups it is affecting? | **Governance interventions**  You should incorporate into the design and development of your AI system the inclusion of datasets, which are sufficiently balanced and representative of the individual rights-holders and groups they are affecting.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 8 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure that all data used in producing the system are accurate, reliable, relevant, appropriate, up-to-date, and of adequate quantity and quality for the use case, domain, function, and purpose of the system? | **Governance interventions**  You should incorporate into the design and development of your AI system the inclusion of data, which are accurate, reliable, relevant, up-to-date, appropriate, and of adequate quantity and quality for the use case, domain, function, and purpose of the system.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 9 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure that all data used in producing the system are attributable, consistent, complete, and contemporaneous with collection? | **Governance interventions**  You should incorporate into the design and development of your AI system the inclusion of data, which are attributable, consistent, complete, and contemporaneous with collection.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 10 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure the proper recording, traceability, and auditability of the provenance and lineage of all data used in producing the system, and any other data involved in the dynamic learning, tuning, or re-training of the system across its lifecycle? | **Governance interventions**  You should incorporate into the design and development of your AI system the inclusion mechanisms and processes to ensure the proper recording, traceability, and auditability of the provenance and lineage of all data used to train, test, and validate the system, and any other data involved in the dynamic learning, tuning, or re-training of the system across its lifecycle.  You should also determine how to demonstrate this in your assurance case. |  |  |
| **Data lifecycle: Means and Methods of Data Collection** | 11 | Where there is human involvement in the data lifecycle, will transparent and publicly accessible measures be implemented to ensure mitigation of potential measurement errors or biases in collection, measurement, and recording processes? | **Governance interventions**  You should incorporate into the design and development of your AI system transparent and publicly accessible measures to ensure mitigation of the potential for measurement errors or biases in collection, measurement, and recording processes where there is human involvement in data collection.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 12 | In the event that collected or procured datasets have missing or unusable data, will the methods used for addressing these deficiencies be transparent and made accessible to relevant stakeholders? | **Governance interventions**  Methods and measures should be put into place to address issues arising from collected or procured datasets, which have missing or unusable data.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 13 | Where personal data are used in the production of the AI system, will information be made available to impacted rights-holders and other relevant stakeholders about the consent or legitimate basis to use that data for the purpose of the system? | **Governance interventions**  Measures should be put in place to make information available to impacted rights-holders and other relevant stakeholders about the consent or legitimate basis to use personal data for the purpose of the system. |  |  |
| 14 | If consent or the legitimate basis to use personal data is implied, will rights-holders and other relevant stakeholders be consulted to identify acceptability of the data use or concerns that need to be addressed? | **Governance interventions**  If the project involves the use of personal data based on implied consent, how you can consult impacted rights-holders and other relevant stakeholders to identify acceptability of the data use or concerns that need to be addressed. |  |  |
| **Data lifecycle: Data Types** | 15 | Will the AI system use dynamic data, collected and processed in real time (or near real time), for continuous learning? | **Governance interventions**  You should address the question how to manage the related risks and demonstrate risk management measures in particular, those involving the assurance of goals and properties of safety, security, reliability, robustness, data quality, and data integrity as well as, where appropriate, non-discrimination and bias mitigation. |  |  |
| 16 | Do the domain in which the data are collected or procured, and the type of the data collected or procured, pose risks of rapid or unexpected distributional shifts or drifts that could adversely impact the accuracy and performance of the system? | **Governance interventions**  In addition to measures in respect of question 15 above, you should also involve domain experts to determine the potential sources of distributional shifts or drifts and build processes of dynamic assessment, re-assessment, external validation, and monitoring into your project lifecycle. |  |  |
| 17 | If the AI system will use unstructured data or a combination of structured and unstructured data, will the project lifecycle incorporate mechanisms and processes to ensure that the inferences generated from that data by the system are reasonable, fair, and do not contain lurking proxies or correlations that are discriminatory and inequitable? | **Governance interventions**  You should manage the risks that these data may generate inferences that are unreasonable, inequitable, or contain lurking discriminatory proxies or correlations. |  |  |
| **Data Lifecycle: Data Labelling and Annotating Practices** | 18 | Will processes of labelling and annotating the data used to produce the AI system be transparently reported and made accessible for audit, oversight, and review by appropriate authorities and relevant stakeholders? | **Governance interventions**  You should determine how to make the processes of labelling and annotating the data that will be used to produce your AI system transparent and accessible for audit, oversight, and review by appropriate authorities and relevant stakeholders.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 19 | Where human labellers and annotators are involved, will sufficient and transparently reported processes be put into place to mitigate potential labelling or annotation biases, especially in cases where these activities concern social and demographic categories that can import patterns of discrimination and proxies for protected characteristics? | **Governance interventions**  You should determine how to mitigate any potential labelling or annotation biases that may arise in the production of your AI system.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 20 | If data labelling or annotation is partly or fully automated, will sufficient and transparently reported processes of human oversight be implemented to mitigate the negative impact of biases generated by automated labelling or annotation? Have cases where the dataset includes social and demographic categories that can import patterns of discrimination and proxies for protected characteristics been considered and addressed? | **Governance interventions**  You should determine how to mitigate any potential labelling or annotation biases generated by automated data labelling or annotation.  You should also determine how to demonstrate this in your assurance case. |  |  |
| **Data Lifecycle: Dataset Linkage** | 21 | Is there a possibility of deanonymizing or identifying rights-holders through data linkage with existing data, publicly available datasets, or data that could be easily obtained? | **Governance interventions**  You should determine whether your AI system may deanonymize or identify rights-holders through data linkage with existing data, publicly available datasets, or data that could be easily obtained and, if so, how to manage the risks of this potential deanonymization or identification. |  |  |
| **Data Lifecycle**  **(conclusion)** | **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues concerning data lifecycle governance of the assessed AI system when assessing the potential impacts of your AI system on the right-holders and in particular regarding the overall performance of the system, the principle of reliability and the principle of privacy and personal data protection and the principle of equality and non-discrimination. | | |  |  |
| **Project Design Context: Decision to Design and Definition of Problem and Outcome** | 22 | Will an evaluation be carried out as to whether building the AI system is the right approach given available resources and data, existing technologies and processes, the complexity of the use-contexts involved, and the nature of the policy or social problem that needs to be solved? | **Points to be addressed in COBRA**  You should determine whether building the AI system is the right approach given the potential adverse effects on human rights, democracy and the rule of law, available resources and data, existing technologies and processes, the complexity of the use-contexts involved, and the nature of the policy or social problem that needs to be solved. |  |  |
| 23 | Will processes of formulating the problem to be solved by the AI system and of defining its target variable (or measurable proxy) be opened to input from stakeholder engagement and public scrutiny? | **Governance interventions**  You should determine the proportionate level of stakeholder involvement, and, if and where appropriate, seek public input accordingly.  This public input should include (1) determining the reasonableness, fairness, equity, and justifiability of the translation of the project's objective into the statistical and mathematical frame and (2) determining the alignment of that translation with the potential impacts of the system on human rights, democracy, and the rule of law.  You should also determine how to demonstrate this in your assurance case. |  |  |
| **Model Development Context** | 24 | Where feature engineering, whether automated or carried out by humans, involves the grouping, disaggregating, or excluding of input features related to protected or potentially sensitive characteristics (e.g. decisions about combining or separating categories of gender or ethnic groups) or proxies for these, will the production of the AI system incorporate processes to mitigate emergent forms of bias? Will processes be incorporated to make the rationale behind these decisions transparent and accessible to impacted rights-holders and other relevant stakeholders? | **Governance interventions**  You should determine how to incorporate measures to mitigate any emergent forms of bias in the event that the feature engineering stage of an AI system’s production, whether automated or carried out by humans, involves the grouping, disaggregating, or excluding of input features related to protected or potentially sensitive characteristics or proxies for these.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 25 | If the algorithmic model(s) or technique(s) used by the AI system have a non-deterministic, probabilistic, evolving, or dynamic character that prevents or hinders the system's intended functionality from being formalized into specific and checkable design-time requirements (or that impairs commonly accepted methods of formal verification and validation), will the system interact with rights-holders in ways that could adversely impact their human rights? | **Governance interventions and points to be addressed in COBRA**  You should determine (1) whether the algorithmic model(s) or technique(s) used by the AI system will have a non-deterministic, probabilistic, evolving, or dynamic character that prevents or hinders the system's intended functionality from being formalized into specific and checkable design-time requirements (or that impairs commonly accepted methods of formal verification and validation), and (2) whether the system will interact with rights-holders in ways that could adversely impact their their human rights. |  |  |
| 26 | If the algorithmic model(s) or technique(s) used by the AI system have a complex, high-dimensional, or non-linear character that impairs or prevents the interpretability and explainability of the system, will the system interact with rights-holders in ways that could adversely impact their human rights? | **Points to be addressed in COBRA**  If so, you should address this risk factor in your assessment of questions 10 and 11 in **COBRA Resources C**. |  |  |
| 27 | Where complex or potentially opaque models are under consideration, will processes of model selection include appropriate and transparent considerations of the AI system's explainability by taking into account:  a. The normal expectations of intelligibility and accessibility that accompany the function the system will fulfil in the sector or domain in which it will operate;  b. The availability of more interpretable algorithmic models or techniques in cases where the selection of an opaque model poses risks to the physical, psychological, or moral integrity of rights-holders or to their human rights  c . The availability of the resources and capacity that will be needed to responsibly provide supplementary methods of explanation (e.g. simpler surrogate models, sensitivity analysis, or relative feature important) in cases where an opaque model is deemed appropriate and selected? | **Governance interventions**  You should determine how to incorporate measures to include explainability considerations in your model selection process.  You should also determine how to demonstrate this in your assurance case.  **Points to be addressed in COBRA**  If so, you should address this risk factor in your assessment of questions 10 and 11 in **COBRA Resources C**. |  |  |
| **Model Output and Model Implementation Context** | 28 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure that the inferences generated from the model’s learning mechanisms are reasonable, fair, equitable, and do not contain discriminatory correlations or influences of lurking or hidden proxies for discriminatory features that may act as significant factors in the generation of its output? | **Governance interventions**  You should determine how to ensure that the inferences generated from the model’s learning mechanisms are reasonable, fair, equitable, and do not contain discriminatory correlations or influences of lurking or hidden proxies for discriminatory features that may act as significant factors in the generation of its output.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 29 | Will sufficient and transparently reported processes of external peer review and evaluation by independent domain and technical experts be included in the evaluation, verification, and validation of the AI model? | **Governance interventions**  You should determine how to ensure that processes of external peer review and evaluation by independent domain and technical experts be included in the evaluation, verification, and validation of the AI model.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 30 | Will processes for evaluating the trained AI model include transparently reported external validation? | **Governance interventions**  You should determine how to ensure that processes of external validation are included in the evaluation of the trained AI model. |  |  |
| 31 | Will processes of monitoring the AI system during its operation involve regular re-evaluations of performance that are sufficient to keep pace with real world changes that may cause concept drifts and shifts in underlying data distribution | **Governance interventions**  You should determine how to ensure that processes of monitoring your AI system during its operation involve regular re-evaluations of performance that are sufficient to keep pace with real world changes that may cause concept drifts and shifts in underlying data distributions. |  |  |
| 32 | When performance metrics for the AI system are considered, determined, and reported, will the prioritization of error types (e.g. false positives/negatives) be:   1. Informed by the specific context of the use case and by the potential effects of differential error rates on affected sub-populations (in particular, on vulnerable or protected groups)? 2. Clearly and accessibly presented, so that the rationale behind the chosen metrics is made explicit and understandable in plain, non-technical language? | **Governance interventions**  You should determine how to ensure that the selection and reporting of error types are informed by the specific context of the use case and the potential effects of differential error rates on affected sub-populations. |  |  |
| 33 | When performance metrics for the AI system are considered, determined and presented, will the prioritization and reporting of metrics beyond accuracy (e.g. sensitivity, precision, specificity) be informed by the specific context of the use case and its performance needs (e.g. a system whose effective identification of rare events is more critical than its overall accuracy rate)? | **Governance interventions**  You should determine how to ensure that the selection and reporting of performance metrics for your AI system beyond accuracy (e.g. sensitivity, precision, specificity) are informed by the specific context of the use case and its performance needs. |  |  |
| **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues when assessing the potential impacts of your AI system on the right-holders and in particular regarding the overall performance of the system, the respect of the principle of reliability and the respect for equality and non-discrimination. | | |  |  |

#### 

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Risk Factors Arising in the System’s Deployment Context** | | | | | |
| **Questions** | | | **Governance interventions and points to be addressed in when assessing potential adverse effects on human rights, democracy and the rule of law** | **Response** | **Completed by** |
| **Privacy and Data Protection** | 1 | If the AI system will be processing personal data, will the project lifecycle incorporate transparently reported mechanisms that demonstrate its compliance with the applicable data protection and privacy laws? | **Governance interventions**  You should determine, through input from data protection experts, the measures that your project needs to take to process personal data lawfully and in accordance with the applicable data protection rules (such as Convention 108+), including the completion of a Data Protection Impact Assessment (DPIA).  You should also determine how to demonstrate this in your assurance case. |  |  |
| 2 | If the AI system will be processing personal data, will the project lifecycle incorporate transparently reported mechanisms that demonstrate its respect for the rights of data subjects and conformity to the additional obligations of data controllers and processors (such as, for instance, those set out in Articles 9 and 10 of Convention 108+)? | **Governance interventions**  You should determine, through input from data protection experts, the measures that your project needs to take to process personal data in accordance with obligations of data controllers and processors (for instance, as set out in Articles 9 and 10 of Convention 108+).  You should also determine how to demonstrate this in your assurance case. |  |  |
| 3 | If the AI system is designed with the purpose or function of individual-targeted curation, profiling, prediction, or behavioural steering, will affected rights-holders be able to obtain from the data controller sufficient information concerning:   1. The use of their personal data and the categories used in the system's processing. 2. An explanation of the rationale behind the output of the processing in plain, non-technical language. 3. The purpose of the curation, profiling, prediction, classification, or behavioural steering. 4. The categories of persons or bodies to whom personal data, the profile or the result of the processing may be communicated? | **Governance interventions**  You should determine, through input from data protection experts, the measures that your project needs to take to process sensitive data as defined in the applicable data protection rules (such as Convention 108+), in compliance with appropriate safeguards. |  |  |
| **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues when assessing the potential impacts of your AI system on the right-holders and in particular regarding the overall performance of the system, the respect of the principle of reliability and the respect for privacy and personal data protection. | | |  |  |
| **Non-discrimination and bias** | 4 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure that the AI system, in both its production and use, mitigates possible sources of bias and discriminatory patterns in each of the following:   1. The datasets used to train the system; 2. The decisions made to build the system; 3. The way the problem to which the system responds is understood, formulated, and framed; 4. The way the target variable and its measurable proxy are defined; 5. The way the system's algorithmic model(s) is selected, and its parameters tuned and adjusted; 6. The way that model is trained, tested, and validated; 7. The way the system is implemented, and the way users are trained to deploy it; 8. The choices made about monitoring, updating, repurposing, or deprovisioning the system?   **TEMPLATE** | **Governance interventions**  You should determine, how to incorporate sufficient and transparently reported processes to ensure that your AI systems, in both its production and use, mitigates possible sources of bias and discriminatory patterns.  You should also determine how to demonstrate this in your assurance case. |  |  |
| 5 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure that the system, in both its production and use, promotes diversity and inclusiveness:   1. In the composition of innovation team building the system; 2. In the expertise, insight, and knowledge drawn upon to develop it; 3. In the individuals and groups able to access its benefits? | **Governance interventions**  You should determine, how to ensure that your AI system, in both its production and use, promotes diversity and inclusiveness.  You should also determine how to demonstrate this in your assurance case. |  |  |
| **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues when assessing the potential impacts of your AI system on the right-holders and in particular regarding the overall performance of the system and the respect for the principle of equality and non-discrimination. | | |  |  |
| **Human rights protection of system designers and implementers** | 6 | Will the project lifecycle incorporate sufficient processes to ensure that the deployment of the system does not harm the human rights of implementers? | **Governance interventions and points to be addressed in COBRA**  You should carry out a thorough consideration of the possible effects of the AI system on the human rights and of its implementers and ensure that the deployment of the system does not harm them. |  |  |
| **Level of automation/Level of human involvement and choice** | 7 | Will implementers of the AI system be sufficiently trained so that they are able to fully understand:   1. the strengths and limitation of the system and its outputs; 2. the potential conditions of situational complexity, uncertainty, anomaly, or system failure that may dictate the need for the exercise of human judgment, common sense, and practical intervention. | **Governance interventions**  You should determine how to ensure that implementers of the AI system are sufficiently trained to fully understand both the strengths and limitation of the system (and its outputs) and the potential conditions of situational complexity, uncertainty, anomaly, or system failure that may dictate the need for the exercise of human judgment, common sense, and practical intervention.  **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues when assessing the potential impacts of your AI system on the right-holders and in particular regarding the respect for the principle of human dignity and individual autonomy, the principle of transparency and oversight and the principle of accountability and responsibility. |  |  |
| **Human in the loop** | 8 | If the AI system has a high level of automation or operational 'autonomy' and interacts with rights-holders in ways that could adversely impact their physical, psychological, or moral integrity or harm their human rights, will mechanisms of human control and intervention (e.g. human-in-the- loop or human-on-the-loop) be incorporated into the implementation of the system? | **Governance interventions**  You should determine how to ensure that, if your AI system has a high level of automation and operational 'autonomy' and interacts with rights-holders in ways that could adversely impact their integrity or otherwise harm their human rights, processes of human control and intervention are appropriately integrated into the system’s deployment.  **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues when assessing the potential impacts of your AI system on the right-holders and in particular regarding the respect for the principle of human dignity and individual autonomy, the principle of transparency and oversight and the principle of accountability and responsibility. |  |  |
| **Out-of-the-scope uses** | 9 | Will sufficient processes be in place to prevent the system malfunction, misuse, or abuse, in particular, where the breakdown or failure of the system, the use of the system out-of-the-scope of its intended purpose, or its malicious misapplication could harm human rights of any impacted rights-holder? | **Governance interventions**  You should determine how to ensure that sufficient protections are in place  to prevent the use of the system out-of-the-scope of its intended purpose, or its malicious misapplication.  **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues when assessing the potential impacts of your AI system on the right-holders and in particular regarding the respect for the principle of transparency and oversight and the principle of accountability and responsibility. |  |  |
| **Proximity to decision-making or action** | 10 | What is the proximity of the application of the AI system to the relevant activity(ies) (decision-making or action) which may have the identified impacts on human rights?  Possible categories include: taking or reviewing the relevant decision or action/substantially informing or influencing the relevant decision or action/merely informing decision making or action/support function without meaningful influence on the decision making or action/none of the above. | **Governance interventions**  You should determine whether sufficient and adequate protections, in accordance in particular with the existing applicable human rights standards, are in place to mitigate any risks resulting from the manner of the involvement of the AI system in the relevant decision-making or actions.  **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues when assessing whether the potential impacts of your AI system on the right-holders have been properly mitigated. |  |  |
| **Accountability, access to remedies and oversight** | 11 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure end-to-end accountability across the production and use of the AI system? Namely, will it ensure that the system:  a.  Is auditable by design, allowing for the end-to-end traceability and oversight of its processes of production and use.  b.  Establishes a continuous chain of human responsibility for all roles involved in the project lifecycle to allow for end-to-end answerability in the event that the human rights of affected individuals have been negatively impacted.  c.  Enables designated public authorities and third parties, where appropriate, to assess its compliance with existing legislation, regulation, and standards instruments across the entire project lifecycle? | **Governance interventions**  You should determine how to incorporate transparently reported processes to ensure sufficient accountability across the production and use of the AI system. clear and accessible way.  You should also determine how to demonstrate this in your assurance case.  **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues when when assessing the potential impacts of your AI system on the right-holders and in particular regarding the respect for the principle of transparency and oversight and the principle of accountability and responsibility. |  |  |
| 12 | Will sufficient and transparently reported processes be implemented throughout the project’s lifecycle to ensure that affected persons whose human rights have been adversely impacted by the AI system have actionable redress and effective remedy by:  a.  Providing sufficient and meaningful information that indicates when the system is being used and how and where to complain in the event of an adverse impact on human rights;  b.  Facilitating access of affected rights-holders to sufficient and meaningful information about the processes behind the design, development, and deployment of the system and about the rationale underlying the outcomes of its processing;  c.  Employing algorithmic models that are appropriately interpretable or explainable (especially regarding discriminatory proxies or inferences that may be embedded in trained machine learning systems) given the risks to human rights, democracy, and the rule of law they may pose? | **Governance interventions**  You should determine how to incorporate transparently reported processes to ensure that affected individuals have actionable redress and effective remedy in cases where their human rights have been adversely impacted.  You should also determine how to demonstrate this in your assurance case.  **Points to be addressed in COBRA**  Make sure to consider the implications of the mentioned issues when assessing whether the potential impacts of your AI system on the right-holders have been properly mitigated, in view of the principle of transparency and oversight, and the principle of accountability and responsibility. |  |  |