LEGAL AND ETHICS CHECKLIST FOR AI SYSTEMS

By Lisa R. Lifshitz and Cameron McMaster

Despite the significant benefits that AI systems can offer users and society in general, any organization deploying an AI system must protect core values such as fairness, transparency, and privacy by design. In operation, many AI systems include inherent bias, including discrimination against certain individuals or groups. Unexplainable AI system decisions also raise fundamental questions of accountability, not only with respect to privacy and data protection law but also involving liability in the event of errors and harm to individuals. Given ongoing concerns about the possible malicious use of AI and related risks to privacy and data protection, prior to acquisition of any AI system, prospective purchasers should consider critical guiding principles including accountability, transparency, and intelligibility.

AI systems should be designed responsibly from the very start, applying the principles of privacy by default or privacy by design—or, if you will, "ethics by design." Practically, this includes implementing adequate technical and organizational measures and procedures (proportionate to the type of system being designed or implemented) to ensure that data subjects' privacy and personal information are respected. Developers should be assessing and documenting expected and potential impacts on individuals and society at large during an AI project's entire life cycle and identifying specific requirements for fair and ethical use. Moreover, while the use of AI is to be encouraged, it should not occur at the expense of human or individual rights. This includes respecting data protection or privacy rights—including rights to access, the right to object to processing, and the right to erasure—and guaranteeing, if applicable, an individual's right not to be subject to a decision based solely on automated processing if the decision significantly impacts them. Regardless, individuals should always have the right to challenge AI system decisions.

Given the foregoing, where should an organization that wishes to acquire and use an AI system begin, and what should they review before implementing that system? The following checklist may be of assistance and is intended to serve as a starting point, not an exhaustive compendium, for some of the legal and ethical considerations involved. Expanding on the European Commission's existent Ethics Guidelines for Trustworthy AI, this checklist considers emerging AI applications and their concomitant legal issues.[1] Although this checklist is largely based on European and Canadian privacy law requirements, the considerations involved are applicable to any jurisdiction and are a work in progress.

ENDNOTE

[1] European Commission, "Ethics Guidelines for Trustworthy AI - Independent High-Level Expert Group on Artificial Intelligence" (Apr. 8, 2019), https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

Lisa R. Lifshitz is a partner, Business Law Group, at Torkin Manes LLP and chair of the firm's Technology, Privacy and Data Management Group in Toronto, Canada. Cameron McMaster is an associate at Torkin Manes LLP in Toronto, Canada.

QUESTION	Y	N
HUMAN AGENCY AND OVERSIGHT		
Human Rights		
Was a human rights impact assessment performed? Were possible trade-offs between the different principles and rights ascertained, documented, and evaluated?		
Does the AI system suggest actions or decisions to make, or outline choices to human users?		
• Could the AI system inadvertently impact human users' autonomy by influencing and obstructing their decision-making?		
• Did you evaluate whether the AI system should inform users that its outputs, content, recommendations, or results arise from an algorithmic decision?		
Human Agency		
Is the AI system intended to form part of a work and labor process? If yes, did you discuss how to divide tasks between the AI system and humans whereby human oversight and control is balanced with optimal work and labor conditions?		
• Did you implement safeguards to limit human overconfidence in or overreliance on the AI system?		
Human Oversight		
Did you appoint someone to oversee compliance with data protection principles and laws with respect to collection and use of personal information?		
• Does this appointee have the support of senior management and the authority to intervene on privacy issues?		
• Did you implement a schedule for regular reviews of privacy management and addressing any shortcomings?		
• Did you make available to users the name or title and contact information of the appointee?		
Did you evaluate the suitable level of human control and oversight for the AI system and its use case?		
• Can you explain how humans control, oversee, and interact with the AI system?		
• Does your AI system include a "human in the loop"? When does this person intervene?		
• Did you implement systems and processes to guarantee human control or oversight?		
• Did you implement audit and corrective systems that pertain to the issues of the governance of AI autonomy?		
• Did you review the tests, efficacy of human control and oversight, and audit and corrective systems with counsel to determine the legal risks associated with deploying the AI system and its use case?		
• Did you review the outputs of the final product with counsel to determine possible legal risks of deploying the AI system?		
• If you have modified the use case of your AI system, have you reviewed the change of purpose with your counsel for potential legal risks?		
• Did you implement detection and response systems to evaluate issues pertaining to the AI system's self-learning?		
• Did you implement a procedure to safely cancel an operation where required? If yes, does this procedure cancel the process completely or partly, or does it return control to a human?		
• Did you review detection and response mechanisms and cancel procedures with counsel to determine legal risks?		
SECURITY, TECHNICAL ROBUSTNESS, AND SAFETY		
Resilience to Attack and Security		
Did you evaluate the AI system's security vulnerabilities?		
Did you implement processes and mechanisms to guarantee the integrity and resilience of the AI system against possible attacks?		
Did you test the AI system's behavior in unexpected circumstances, different situations, and diverse environments?		
Data Breach Notification Requirements		
Do you have procedures in place to comply with applicable data breach notification requirements for your AI system as you would for any other data-using system?		

QUESTION	Y	N
Contingency Plan and General Safety		
Does your system have a satisfactory contingency plan if it encounters attacks or other unexpected situations?		
Did you evaluate the likelihood that the AI system could harm or cause damage to users or third parties? Did your evaluation consider the probability, possible damage, impacted parties, and seriousness of the damage/harm?		
• Did you assess the risk to humans, animals, and the environment? Did you provide the necessary information to users and others to mitigate such risk?		
• Did you consider the risks that arise from foreseeable uses of the AI system, such as accidental or malicious misuse? Do you have a plan to mitigate or manage these risks?		
• Did you review product liability, warranty, and consumer protection laws of the jurisdictions?		
• Did your risk assessment consider whether security or network problems, including cyberattacks, could present safety risks, injury, or damage due to the AI system's inadvertent behavior?		
• Did you consider an insurance policy to deal with potential damage from the AI system?		
Did you assess the possible consequences of whether the AI system becomes unavailable, delivers incorrect findings or outputs offensive content (e.g., discriminatory or defamatory)?		
• Did you implement processes to commence alternative/contingency plans?		
• Did you document and test the contingency plans?		
Accuracy of Data		
Did you develop and define an approach to output accuracy required for the AI system and its use case (e.g. the rate of accuracy and its meaning)?		
• Did you test how accuracy is measured and assured?		
• Did you implement processes to guarantee that the data used are thorough and current?		
• Did you implement processes to evaluate requirements for additional data (e.g. to increase accuracy or to reduce and remove bias)?		
Did you evaluate potential harms caused by inaccurate outputs of the AI system?		
Did you define a threshold of the level of inaccurate outputs? Did you implement systems to detect when your AI system surpasses that and a framework to remediate?		
Did you implement processes to increase accuracy of the AI system's outputs?		
Did you implement a process by which personal information is kept accurate, complete and current to satisfy the purposes for which it is used?		
Did you implement a system for individuals to challenge the accuracy and completeness of personal information and have it amended as appropriate?		
Reliability and Reproducibility of Outputs		
Did you create and implement a framework to assess whether the AI system is meeting your objectives, functions, and proposed uses?		
• Did you document whether specific contexts or conditions must be considered to guarantee reproducibility of outputs?		
• Did you implement processes and methods that verify the various aspects of the AI system's reliability and reproducibility of outputs?		
• Did you implement mechanisms that explain when and how an AI system fails in certain situations?		
• Did you document and implement the processes that test and verify the AI system's reliability and reproducibility?		
• Did you create communication mechanisms that assure users of the AI system's reliability?		
PRIVACY AND DATA GOVERNANCE		
Data Protection Law Requirements		
Do you have procedures in place to comply with applicable data protection law requirements for your AI system and any training and testing data as you would for any other data-using system?		

QUESTION	Y	N
Respect for Privacy and Data Protection		
Do you have procedures in place to maintain respect for privacy and data protection in your AI system and any training and testing data as you would for any other data-using system?		
Source and Ownership of Data		
Did you develop, document, and implement policies that ensure that all data were collected by fair and lawful means?		
Did you perform vendor due diligence on your data sources/brokers?		
Did you catalog/record the source for all datasets?		
Did you ensure that all datasets are licensed properly or do not infringe any third-party rights?		
Did you develop, document, and implement data ownership audit mechanisms for data source and title verification?		
Quality and Integrity of Data		
Did you align your system with relevant standards (e.g., ISO, IEEE) or widely adopted protocols for daily data management and governance?		
Did you design and implement oversight systems for data collection, storage, processing, and use?		
Did you evaluate your level of control over the quality and integrity of the third-party data sources used?		
Did you implement methods to guarantee the quality and integrity of your data? Did you assess other processes? Can you explain how you verify whether your datasets have not been compromised or hacked?		
Access to Data		
Do you have rules, processes, and methods to guarantee proper data governance?		
• Did you consider which parties can access data of users and in which circumstances?		
• Did you guarantee that these parties are qualified and required to access the data and that they have the necessary skill and knowledge to understand the data protection policy?		
• Did you implement an oversight mechanism that logs when, where, how, by whom, and for what purpose data were accessed?		
Did you implement a system or process by which individuals whose personal information forms part of a dataset used by the AI system can request to be informed of the existence, use, and disclosure of their personal information and be given access to it within a time period appropriate to your jurisdiction, industry sector, and the nature of the data?		
• Did you provide the name or title and contact information to whom access requests should be sent?		
• Did you provide instructions on how individuals can gain access to their own personal information?		
TRANSPARENCY		
Traceability		
Did you implement traceability measures, such as: • System design and development: • Rule-based AI systems: how the rule set was built; • Learning-based AI systems: training methods, including which training data were collected, selected, and used, and how training was performed. • Testing and validation: • Rule-based AI systems: the scenarios or situations used; • Learning-based model: information about test data used. • Outcomes: • System outputs, as well as other outputs created from other use cases.		
Explainability		
Did you evaluate: • to what extent the output created by the AI system understandable by humans? • to what extent the AI system's output affect your organization's decision-making processes? • the reasons why this specific AI system was used in this particular use case?		
Can you explain why the AI system made a specific choice causing a specific outcome that all users can understand?		

QUESTION	Y	N
Did you consider interpretability as a core element of the design of your AI system from start to finish?		
• Did you actively investigate and consider using the simplest and most interpretable model available for the application in question? Did you implement that model?		
• Did you evaluate whether you can assess your training and testing data? Can you assess it?		
• Did you consider whether you can analyze the model's interpretability after the model's training and development, or you have access to the internal process of the model? Can you analyze the model or have access to its internal process?		
Communication		
Are users informed they are communicating with an AI system and not a human? Do your marketing materials and labeling communicate that to users?		
Did you communicate to users the reasoning behind the AI system's outputs?		
• Did you communicate these reasons this plainly and logically to the intended users of the AI system?		
• Did you consider user feedback to improve the AI system? Did you implement processes to do so?		
• Did you inform users about potential or perceived risks, including bias, resulting from the outputs of your AI system?		
• Based on the intended use case, did you assess how to communicate the above and your approach to transparency to others?		
Did you explain the purpose of the AI system and who or what may benefit from its products and services?		
• Did you identify the use cases for the AI system in a clear, comprehensible, and suitable manner for the intended users?		
• Did you consider human psychology and its likely limitations, including cognitive biases?		
Did you plainly communicate the AI system's attributes, constraints, and possible deficiencies?		
• During its development stage, to the party that is deploying it into a product or service?		
• During its deployment stage, to the user or consumer?		
ACCESSIBILITY AND ELIMINATING BIAS AND HARMFUL OUTPUTS		
Accessibility and Universal Design		
Did you guarantee that the AI system can accommodate a comprehensive scope of individual preferences and abilities?		
• Did you evaluate whether users with different accessibility requirements can use the AI system? How was accessibility designed into the AI system, and how is this tested and verified?		
• Did you guarantee that information about the AI system is available to users employing assistive technologies?		
• During the development stage of the AI system, did you consult with communities with different abilities?		
Did you consider the impact of your AI system on the target users based on their identities / demographics?		
• Did you consider the demographics of your target user and those of the team building the AI system? Is the team responsible for building the AI system representative of your target users? Is it representative of the wider population, considering also other groups who might be tangentially impacted?		
• Did you evaluate whether your AI system's output may contain negative implications that could disproportionately affect persons or groups?		
• Did you solicit feedback from staff in your organization who are representative of different backgrounds and experiences?		
Stakeholder Participation		
Did you assess using a methodology that would include the participation of a diverse and inclusive set of stakeholders in the AI system's design, development, and use?		
Did you plan for the introduction and deployment of the AI system in your organization by communicating and involving impacted workers and their representatives of different communities before launch?		
Unfair Bias Avoidance		
Did you develop procedures to prevent the formation or strengthening of unfair bias in the AI system, including its design and input data?		
• Did you evaluate the possible operational deficiencies to the AI system resulting from dataset composition?		

QUESTION	Y	N
• Did you establish a strategy to increase diversity and inclusivity within the teams responsible for system design, development, and deployment?		
• Did you assess the diversity and representativeness of users in the data? Did you test for particular communities or difficult or uncommon use cases?		
• Did you apply existing technologies to enhance your understanding of the data, model, and its performance?		
• Did you implement systems to screen for possible biases during the design, development, deployment, and use stages of the system?		
Did you implement a process that lets others, such as users, identify issues pertaining to bias, discrimination, or inadequate performance of the AI system?		
• Did you create a process and framework that enables the communication of the issues to the appropriate party at your organization?		
• Did you think about other people possibly indirectly impacted by the AI system, in addition to the users?		
Did you evaluate whether any outcome variability could occur under the same input conditions?		
• If yes, did you assess the possible causes of the outcome variability?		
• With respect to variability, did you develop a system that measures the possible impact of such variability on human rights?		
Did you implement an acceptable working definition of “fairness” that you employed in designing an AI system?		
• Is your definition a commonly used one? Did you evaluate other definitions before selecting this one?		
• Did you implement metrics to measure and test the applied definition of fairness?		
• Did you develop systems and tools to guarantee fairness in your AI system? Did you consider third-party systems and tools?		
Harmful Impacts		
If your AI system retrieves and displays information when engaged with a user (including chatbots), did you develop, document, and implement your AI system and use case so that it does not inadvertently disclose private information, sensitive information, confidential information, or private facts without appropriate consent of the holders?		
• Did you perform sufficient tests to determine the reasonable ranges of outputs of the AI and consider whether these disclose private facts?		
• Did you have legal counsel test the AI system and analyze its outputs to assess the potential risks of disclosing private information?		
• Did you thoroughly test the AI system and use case to determine that data leakage is not possible?		
• Did you design, document, and implement mechanisms to receive user notices of potential disclosure of private information?		
If your AI system generates content for public consumption (including being a chatbot), did you develop, document, and implement your AI system and use case so that it does not publish third-party information that damages the reputations of third parties in the eyes of a reasonable person?		
• Did you perform sufficient tests to determine potential libel or slander that the AI system could generate?		
• Did you have legal counsel test the AI system and analyze its outputs to assess the potential risks of generating potential defamation or slander?		
• Did you design, document, and implement mechanisms to receive user notices of potential libel or slander?		
If your AI system generates content for public consumption (including being a chatbot), did you develop, document, and implement your AI system and use case so that it does not publish third-party information that damages the reputations of third parties in the eyes of a reasonable person?		
• Did you consult with counsel regarding whether your use case could potentially generate deceptive marketing messages? Did that counsel provide you with an assessment?		
Did you consider whether your AI system or use case could conduct fraud against its users and develop, document, and implement measures to prevent such activities?		
If your AI system or use case physically interacts with the world, did you consider whether your AI system or use case could cause harm or physical injury to its users?		

QUESTION	Y	N
• Did you conduct sufficient testing to determine the various errors that could cause injury to users?		
• Did you develop, document, and implement mechanisms to override the operation of the AI system to protect the safety of the user?		
• Did you conduct sufficient testing of override mechanisms to determine their failure rate?		
• Did you consult with counsel to assess the legal risks produced by the deployment and use of the AI system?		
• Did you communicate the inherent risks of using the AI system and its use case to users in a clear way?		
If your AI system or use case performs fraud detection for financial institutions, did you design, develop, and implement mechanisms and strategies to prevent adverse impacts on false positives?		
Legal Compliance		
If your AI system generates information content for the consumption of the public (such as being a chatbot), did you acquire appropriate licensing to reproduce copyrighted materials and use trademarks?		
If your AI system performs robo-trading or securities trading on behalf of users, are you licensed to do so by the appropriate securities regulator in the jurisdictions in which it trades?		
If your AI system directly communicates commercial messaging with users, did you review with counsel whether the organization is complying with other relevant legislation (e.g., Canada's Anti-Spam Legislation)?		
If your AI system collects personal information and other data points about users to generate scoring profiles, did you review with counsel whether you are complying with consumer report and other licensing and legal requirements?		
ACCOUNTABILITY		
Auditability		
Did you develop systems and tools that enable the AI system's auditability, including ensuring traceability and logging of the AI system's procedures and outputs?		
In situations where the AI system's application affects human rights, did you confirm that it is possible to independently audit the AI system?		
Minimizing and Reporting Negative Impact		
Did you document an assessment of the risks or impacts of the AI system that considers different stakeholders that are directly and indirectly impacted by the AI system?		
Did you train and educate your staff to establish accountability processes and practices?		
• Did you only train the developers of the AI system? Was the training given to teams in stages of bringing the product to market/use? Were senior management present at the trainings?		
• Did these trainings cover the legal issues and compliance requirements that result from developing, deploying and marketing an AI system?		
• Did you assess forming a review board to consider AI ethics matters or a similar committee to evaluate accountability and ethics practices, including the grey areas?		
Did you anticipate any kind of outside guidance or did you implement third-party auditing procedures to monitor ethics and accountability, in addition to internal practices?		
Did you develop procedures for third parties (e.g., suppliers, consumers, distributors/vendors, regulators) or staff to report possible vulnerabilities, risks, or biases in the AI system?		
Documenting Trade-offs		
Did you develop a system to discover relevant interests and values connected to the AI system and the possible trade-offs between them?		
Did you develop a methodology to determine such trade-offs? Did you record the trade-off determinations?		
Ability to Redress		
Did you develop an acceptable set of systems and procedures that permit for redress of any harm or adverse impact?		
Did you implement systems to inform users and third parties about methods of redress?		