Responsible AI: From Managing Risk to Driving Business Value

Article

Many organizations have begun to the see the value of mitigating AI risks. News reports from discriminatory hiring processes using AI to privacy violations in facial recognition have put AI on the agendas of boards and on the lips of CEOs and CIOs. However, this push for ethical AI cannot come from the top-down only, it has to be built from the bottom up. Only then can organizations not only avoid pitfalls in AI, but also start to deliver value from their Responsible AI practices.

What is Responsible AI?

Responsible AI covers a wide array of challenges in the AI space. It makes sure AI is legal, ethical, fair, privacy-preserving, secure, and explainable to name a few of the topics covered.

At Lenovo, we established the Lenovo Responsible AI Committee by bringing together a group of 20 people of diverse backgrounds to decide the principles that AI must support in the organization. Together we decided that the six pillars of Responsible AI at Lenovo would be:

1. Diversity & Inclusion
2. Privacy & Security
3. Accountability & Reliability
4. Explainability
5. Transparency
6. Environmental & Social Impact

In this article, we will review each of these pillars and the types of questions your organization can ask about AI projects. Both internal projects and external vendors go through a process of validation with the Lenovo Responsible AI Committee and must get approval from the Committee to be made an offering. After making these commitments internally, Lenovo has externally promised to uphold these pillars in the latest Environmental, Social, Governance (ESG) section of the Annual Report.

Six Pillars of Responsible AI

Although many organizations and governments divided Responsible AI into different categories, they all cover the same basic topics. Some examples include Google, Microsoft, and the European Union. For Lenovo, we cover our six pillars as follows:

- Diversity & Inclusion
- Privacy & Security
- Accountability & Reliability
- Explainability
- Transparency
- Environmental & Social Impact

Diversity & Inclusion

To be considered Responsible AI, the AI project must work for all sub-groups of people. While AI bias can rarely be eliminated entirely, it can be effectively managed. This mitigation can take place during the data collection process—to include a more diverse background of people in the training dataset—and can also be used at inference time to help balance accuracy between different groupings of people.

Common questions include:

- Did you assess and put in place processes to test and monitor for potential biases during the entire lifecycle of the AI system (e.g. biases due to possible limitations stemming from the composition of the used data sets)?
- Where relevant, did you consider diversity and representativeness of end-users and or subjects in the data?

Privacy & Security

AI should protect individual and group privacy, both in its inputs and its outputs. The algorithm should not include data that was gathered in a way that violates privacy and it should not give results that violate the privacy of the subjects even when bad actors are trying to force such errors. The AI application must also be protected against cybersecurity threats and AI-specific security threats such as data poisoning.

Some questions that need to be asked include:

- Did you consider the impact of the AI system on the right to privacy?
- Have you audited your AI system for cybersecurity risk? Did you define risks, risk metrics and risk levels of the AI system?

Accountability & Reliability

Someone should be ultimately responsible when the AI application makes a decision. Unless these decisions are made upfront, it can result in no one taking responsibility for poor outcomes which provides little protection to the customer. Additionally, the AI applications must be reliable. They must not provide wildly different predictions based on minimal changes to the input. When they fail, the user should be able to recognize and react to the failure, instead of it failing silently.

Questions to ask include:

- Have you implemented stress tests regarding your AI System (minimum point of failure, adversarial attacks, etc)?
- Did you establish mechanisms that facilitate the AI system’s auditability and establish a framework for responsibility in case of AI failure (e.g. traceability of the development process, the sourcing of training data and the logging of the AI system’s processes, outcomes, positive and negative impact)?

Explainability

Many AI applications can be "black boxes" in which the input and output are known, but nothing of the decision-making process is understood. While this may be acceptable in certain low-risk applications there are many situations in which the reason for a decision being made is nearly as important and the decision determined. For example, a medical imaging diagnosis application that does not indicate where on the MRI a problem was detected is of limited value.

- Did you explain the decision(s) of the AI system to the users in easy-to-understand way? What algorithms do you use to enhance explainabilty, if any?
- Is the data set that you used a standardized data set with sufficient description?

Transparency

It is often important to know what data was used to make a decision and what version of a model was used to make a decision. When an AI application makes a decision, can the user request to know exactly what data was used to arrive at that decision? If a newer model is released that does not perform well, can the organization go back and correct the poor decisions that the new model made? Furthermore, does the user know he/she is interacting with an AI agent, or is he/she left to wonder if they are dealing with a live person?

Questions to ask include:

- Did you put in place measures that address the traceability of the AI system during its entire lifecycle? Did you put in place measures to continuously assess the quality of the input data to the AI application?
- In cases of interactive AI applications (e.g., chatbots, robo-lawyers), do you communicate to users that they are interacting with an AI application instead of a human?

Environmental & Social Impact

The effects of an AI project should be evaluated in terms of its impact it will have on the environment and on the subjects and users. The results of the decisions of the AI project on the environment should be considered where applicable. One factor that is applicable in nearly all cases is an evaluation of the amount of energy needed to train the required models. Furthermore, social norms such as democratic decision-making, upholding values, and preventing addiction to AI applications should be upheld.

Questions that can be asked:

- Where possible, did you establish mechanisms to evaluate the environmental impact of the AI system's development, deployment and/or use (for example, the amount of energy used and carbon emissions)?
- Did you assess and try to mitigate the societal impact of the AI system's use beyond the end-user and subject, such as potentially indirectly affected stakeholders or society at large?

Driving Business Value

The risks in not paying attention to Responsible AI are well known – lawsuits and poor press from discriminatory practices or privacy violations to name a few. What is less known is that Responsible AI practices can drive business value.

Customers are noticing the AI applications that consider their unique circumstances, backgrounds, and abilities (Diversity and Inclusion) and choosing to use applications that simply work better for them. They are choosing AI applications that preserve their privacy (Privacy) and do not expose them to unnecessary cybersecurity risk (Security).

AI applications to which the customer knows someone stands behind it (Accountability) and that can consistently produce results (Reliability) are more likely to be used. Who wants to use an application that slight changes in the input produce vastly different responses? The same goes for Explainability and Traceability – who wants to use an AI application that can't tell you why it made that particular decision and cannot tell you what data was used to make a decision?

Finally, customers are choosing green applications over inefficient, wasteful ones (Environmental Impact) and applications that they think are supporting the common good (Social Impact). In short, doing Responsible AI allows you to both reach more customers and provide better service to the customers you have which will naturally drive the bottom line. Responsible AI has moved from a nice-to-have to necessary in a few short years.

About the author

David Ellison is the Chief Data Scientist for Lenovo ISG. Through Lenovo's US and European AI Discover Centers, he leads a team that uses cutting-edge AI techniques to deliver solutions for external customers while internally supporting the overall AI strategy for the Worldwide Infrastructure Solutions Group. Before joining Lenovo, he ran an international scientific analysis and equipment company and worked as a Data Scientist for the US Postal Service. Previous to that, he received a PhD in Biomedical Engineering from Johns Hopkins University. He has numerous publications in top tier journals including two in the Proceedings of the National Academy of the Sciences.

Related product families

Product families related to this document are the following:

- Artificial Intelligence

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service. Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION ”AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary. Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk. Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document, LP1833, was created or updated on October 13, 2023.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP1833
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP1833.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at https://www.lenovo.com/us/en/legal/copytrade/.

The following terms are trademarks of Lenovo in the United States, other countries, or both:
Lenovo®

The following terms are trademarks of other companies:

Microsoft® is a trademark of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.