Responsible AI and Automation

Understand and interpret the risks and limitations of artificial intelligence at an organizational level

Introduction

Automated decision systems, including artificial intelligence (AI) and robotic process automation (RPA), deliver operational efficiencies, improve decision-making, and free staff from repetitive and often time-consuming tasks. They also create entirely new categories of risk, including the potential for unequal job opportunity impact and the prospect of discriminatory predictions. Left unchecked, these sociotechnical risks can erode public confidence, reduce data security and privacy, and attract unwanted regulatory attention.

Addressing such risks and mitigating bias entails more than just technical acumen—these problems cannot be solved by simply hiring more data scientists or writing more code. Designing AI models that work for people requires dedication to responsible automation. This involves commitment from governance at the highest levels, as well as teams that build responsible AI concepts into every new automation initiative.

Understanding Responsible AI Challenges

AI governance is an emerging and fast-changing field, so the terminology associated with responsible AI is a source of constant discussion. Guidehouse recommends anchoring necessary conversations around the key concepts that follow, while being aware that both the definitions and the technological and social implications associated with responsible AI are still works in progress.

Responsible AI: A responsible approach to automation looks beyond the questions of reducing cost or improving speed to consider long-term outcomes. This requires both technical and policy-minded expertise. The risks of an irresponsible approach to automation resemble those encountered in the cobra effect, which is named after a government bounty program on killing poisonous snakes that created a perverse incentive for citizens to breed the dangerous animals for slaughter. When the loophole was discovered and the program ended, the cobras were set free and actually increased the population's risk of snakebite. Likewise, the irresponsible use of AI may exacerbate the very problems it sets out to fix, for example, by perpetuating racial profiling or distorting already existing stereotypes.

Ethical AI: An ethical approach considers both the inputs and outputs of a model, not simply the prospects for it to deliver an immediate return on investment. Input datasets are susceptible to a wide range of risks, including biased sampling and flawed collection. Outputs may not be held to a high standard of accuracy or reproducibility. The risks inherent to outputs that are not tested for ethics include deploying a model that delivers a low rate of successful predictions or one that causes extremely negative consequences when it is in error. Ethical AI approaches also weigh the potential societal and reputational costs of labor force reductions, not merely the productivity advantages in terms of dollars or hours saved.

Trustworthy AI: Trustworthy AI is championed by leaders who view automation as a balancing act—one that can simultaneously increase efficiency, show commitment to technological excellence, and uphold cultural standards.

How to Follow the Seven Key Requirements of Trustworthy AI Systems

Based upon a framework for responsible AI set out by the European Commission, Guidehouse recommends incorporating these seven principles in any automation project to promote more fairness in AI:

1. **Human Agency and Oversight:** There is a simple pass-fail test for this guideline: Is the automation or AI system in request designed to augment human intelligence and judgment, or to replace it? Oversight must be built in not only at the beginning but also applied throughout the lifespan of an automation or algorithm. No algorithm will ever be 100% accurate, and at inception, it often has room to grow and learn. Humans should be monitoring for drift away from expected results, for the ongoing validity and robustness of data, and for the application of changing standards in AI ethics.

2. **Technical Robustness:** AI systems should combine best-in-class tools. In the trustworthy model, "best-in-class" refers to code that has been tested heavily for production-grade volume and accuracy, as well as flexibility to adapt when ethical issues are raised.

3. **Security:** As data is passed between decision-augmenting systems, the risk of security breach or accidental data exposure rises. Guidehouse recommends tools that have been thoroughly vetted to minimize security risks.

4. **Data Governance:** It is important to ensure that data governance standards are kept up to date with the new demands of responsible AI. This includes understanding how data inputs are sourced and curated, and how privacy can be maintained among previously anonymized data sources when opened to focused analysis.

5. **Explainability:** This plain-language test establishes that the easier it is to explain the workings of an algorithm, model, or automated process, the more likely it is to be trustworthy. Explaining models also helps secure organizational buy-in and can expose flaws (such as perverse incentives).

6. **Fairness:** Fair AI considers the potential sources of harm in a model or automation. This step is important not only to ensure legal compliance but also to keep an organization aligned with its core values. Guidehouse recommends building in fairness at every step of the process, instead of waiting for a discrimination test that may miss deep-rooted sources of bias or error.

7. **Accountability:** An accountable organization stands by the positive and negative effects of the AI systems it deploys and recognizes its ongoing responsibility to both maintain the technical standards of that system and to address the negative consequences. Technical, policy, and subject matter experts must be involved throughout the lifespan of the automation, not simply at the design phase.

New Responsible AI Requirements for Federal Contractors

Federal contractors are strongly advised to pivot to a responsible AI footing to stay ahead of anticipated requirements from several major agencies, including the Department of Defense, the FBI, and the intelligence community. Under the Artificial Intelligence Capabilities and Transparency (AICT) Act, these agencies have been instructed to submit proposed AI systems to rigorous ethical tests, including strict requirements for reproducibility and equity. In 2021, the US Government Accountability Office (GAO) published an extensive report on AI accountability assessment. Guidehouse experts can help federal contractors understand these emerging standards and adapt to meet changing development and deployment requirements.

Guidehouse's Approach to Responsible AI Costs and Rewards

The hidden costs of the incomplete, tech-only approach can be considerable. Unchecked, the expansion of access to data through APIs could expose Social Security numbers or other key personal data to unauthorized use. Data once thought properly masked or anonymized could be decoded. Inaccurate models might reach production or stay in production too long, leading to improper decisions that could create both short- and long-term liabilities.

Working with a partner like Guidehouse that understands contemporary frameworks for responsible automation does more than simply streamline the path to deployment by avoiding last-minute surprises about bias or security risks. Guidehouse has developed the Ultimate AI Lab, an infrastructure for tracking, storing, and evaluating machine learning experiments throughout the tech lifecycle. We work with cloud, AI, and data providers to detect bias, assess data quality and data drift, and improve both the security and explainability of client models.

Guidehouse's approach to responsible AI balances efficiency and technical excellence with fairness and accountability. Our experts understand the new risks of social bias, flawed predictions, and security breaches created by these objectively powerful and exciting technologies. Working with us helps ensure that your AI systems remain fair and secure, as well as productive and effective. We provide safety against these hidden risks early in the planning process—not as an afterthought or add-on.

How Guidehouse Can Help

A human-centered approach to automation can keep your organization growing and evolving without sacrificing long-term shared success for short-term gain. Guidehouse understands the risks as well as the rewards that grow from advanced applications of data science and is committed to working with our clients to build responsible, trustworthy, and reliable products. Using the European Commission requirements for trustworthy AI systems as a framework, we have established our own processes and considerations to ensure ethical development of our AI and RPA solutions, and we apply those principles daily in our Ultimate AI Lab.

As automation and advanced insights grow in sophistication and prominence, so too should the emphasis on accuracy, risk avoidance, and human empowerment. Guidehouse expertise can help steer your automation efforts to promote better outcomes for human workers freed for more innovative, knowledge-driven work, while protecting data governance and security as data sets and computations grow to previously unimaginable volumes.

Guidehouse
Outwit *Complexity*

For more information about fairness in AI systems, visit fairlearn.org or contact a Guidehouse expert for a consultation.

Related Links

Fairlearn.org (https://fairlearn.org/)

Audit-AI (https://github.com/pymetrics/audit-ai)

IBM's AI Fairness 360 (https://ai-fairness-360.org/)

Datasheets for Datasets (https://arxiv.org/abs/1803.09010)

Model Cards for Model Reporting (https://research.google/pubs/pub48120/)

ACLU Washington's Algorithmic Equity Toolkit (https://www.aclu-wa.org/AEKit)

Aequitas Bias and Fairness Audit Toolkit (http://aequitas.dssg.io/)

Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities (https://www.gao.gov/products/gao-21-519sp)

Contact

Bassel Haidar
Director
bhaidar@guidehouse.com

About Guidehouse

Guidehouse is a leading global provider of consulting services to the public sector and commercial markets, with broad capabilities in management, technology, and risk consulting. By combining our public and private sector expertise, we help clients address their most complex challenges and navigate significant regulatory pressures, focusing on transformational change, business resiliency, and technology-driven innovation. Across a range of advisory, consulting, outsourcing, and digital services, we create scalable, innovative solutions that help our clients outwit complexity and position them for future growth and success. The company has more than 16,500 professionals in over 55 locations globally. Guidehouse is a Veritas Capital portfolio company, led by seasoned professionals with proven and diverse expertise in traditional and emerging technologies, markets, and agenda-setting issues driving national and global economies. For more information, please visit **www.guidehouse.com**.

Web: guidehouse.com/technology-solutions/ / @GHTechSolutions / linkedin.com/company/guidehouse-technology-solutions/