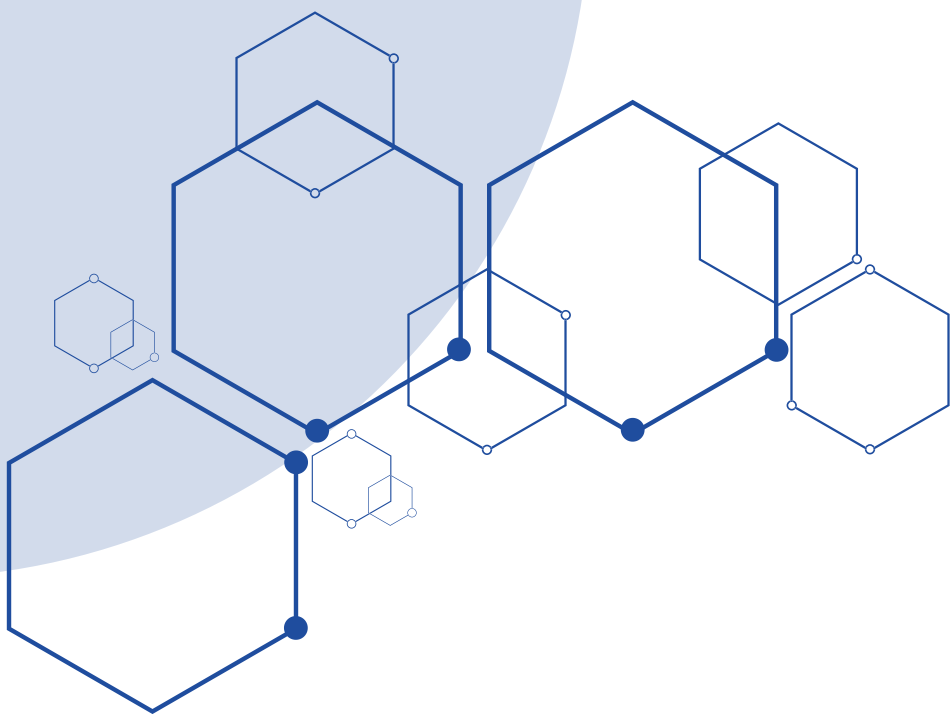


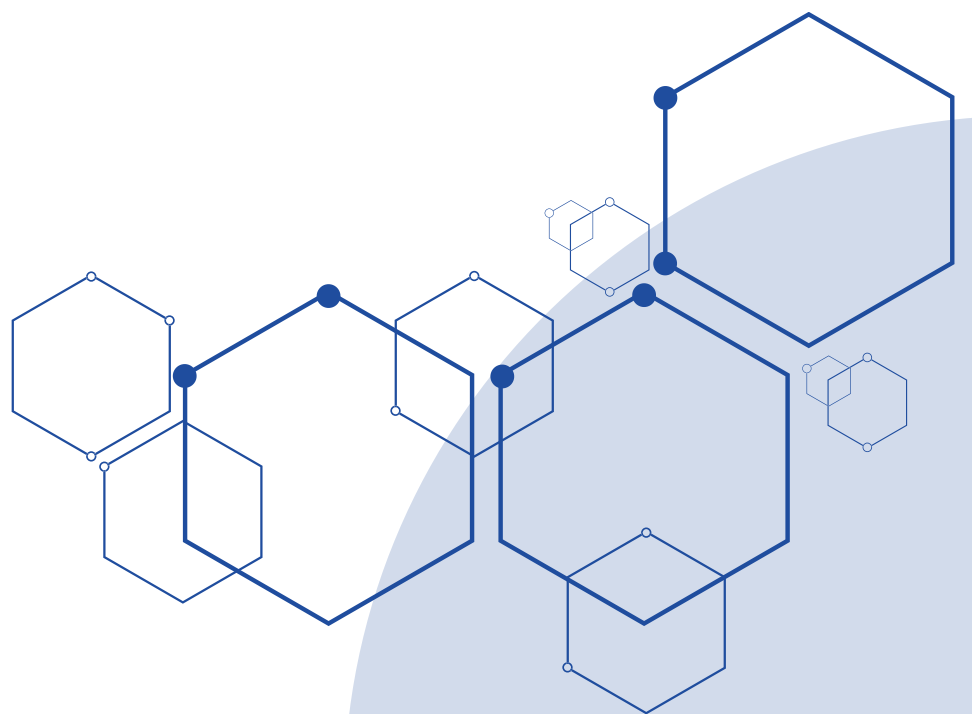


LEADING AI GOVERNANCE

NEW YORK STATE'S ADAPTIVE RISK-TIERED FRAMEWORK

POLICY BRIEF

N. SKY BELL
NBELL62@GATECH.EDU
APRIL 6, 2025



Preface

For the past seven years, I've had the privilege of serving as a public servant in New York State, focusing on bridging the gap between humans and technology by integrating AI advancements into our government systems. This journey has been incredibly fulfilling, allowing me to witness firsthand how AI can truly transform the way we operate and serve our communities. From leading AI initiatives in legal platforms to developing comprehensive deployment guidelines, my passion for this field has only grown stronger over time.

As I began my first semester as a graduate student in the cybersecurity policy track at Georgia Tech, I must admit, I had moments of doubt. I wondered if I truly belonged in such a prestigious program. But then I discovered the AI Safety Initiative's hackathon, and everything fell into place. This event has shown me exactly why I am here and why I chose this path. It has reaffirmed my belief in the importance of responsible AI governance and reignited my passion for this work.

The opportunity to address the urgent challenges in AI governance is both thrilling and inspiring. As AI continues to advance and become more integrated into our lives, it's crucial that we develop thoughtful and innovative policies to guide its use.

My experiences, both as a public servant and now as a graduate student, have given me a unique perspective on the potential of AI and the need for careful oversight. I've seen how AI can revolutionize the world, but I've also recognized the importance of implementing safeguards to ensure these technologies are used responsibly and ethically.

Participating in this hackathon is more than just a educational endeavor for me; it's a chance to contribute to something I truly believe in. I'm excited to bring my insights and enthusiasm to the table, to learn from my fellow participants, and to help shape policies that will benefit our communities. Together, we can ensure that AI is harnessed for the greater good, balancing innovation with the necessary precautions to protect and serve our society.

Nudsara Sky Bell



Executive Summary

The rapid advancement of artificial intelligence (AI) technologies presents unprecedented opportunities and challenges for governance frameworks worldwide. This policy brief proposes an Adaptive Risk-Tiered Framework for AI deployment and application in New York State (NYS) to ensure safe and ethical AI operations in high-stakes domains. The framework emphasizes proportional oversight mechanisms that scale with risk levels, balancing innovation with appropriate safety precautions. By implementing this framework, we can lead the way in responsible AI governance, fostering trust and driving innovation while mitigating potential risks. This framework is informed by the EU AI Act and NIST AI Risk Management Framework and aligns with current NYS Office of Information Technology Services (ITS) policies and NYS legislations to ensure a cohesive and effective approach to AI governance.

Problem Statement & Background

AI technologies are increasingly integrated into critical sectors, raising concerns about safety, ethics, and accountability. The lack of robust oversight mechanisms can lead to unintended consequences, including biased decision-making, privacy violations, and systemic failures. Existing regulatory frameworks struggle to keep pace with rapid AI advancements, necessitating a flexible and adaptive approach to governance (Crawford, 2021).

The NYS Office of Information Technology Services (ITS) plays a crucial role in supporting the critical IT infrastructure for NYS, particularly in the context of AI governance. ITS leads, develops and enforces policies such as [NYS-P03-002 for information security](#) and [NYS-P24-001 for the acceptable use of AI technologies](#), ensuring that AI systems are used responsibly and securely across NYS entities.

However, the current strategy applies a blanket approach to AI deployment, which can inadvertently hinder innovation. To address this, we propose a more nuanced, risk-tiered framework that allows for easier innovation in low-risk areas while implementing stricter oversight and considerations for higher-risk applications.

ITS, as the leading state agency for technology solutions, plays a pivotal role in overseeing and supporting AI governance across NYS. We can provide expert consultation and assistance to other state entities, helping them identify and assess AI solutions effectively. By leveraging frameworks like the NIST AI Risk Management Framework and the risk-tiered approach from EU AI act, we as ITS ensures that potential risks are proactively identified and mitigated before deployment, maintaining our position as a trusted technology partner in navigating the complexities of AI integration for New York State.

Objectives

The primary objectives of this policy are to:

- **Ensure the Safe and Ethical Deployment of AI Technologies:** Establish robust frameworks to safeguard the deployment of AI technologies within New York State (NYS), ensuring they meet safety and ethical standards.
- **Foster Trust in AI Systems:** Promote transparency and accountability in AI governance to build public trust and confidence in AI systems, ensuring stakeholders are informed and engaged in the deployment process.
- **Promote Innovation While Mitigating Risks:** Encourage the development and implementation of innovative AI solutions while proactively identifying and addressing potential risks to maintain a balance between progress and responsibility.
- **Align NYS with Global Best Practices:** Position NYS at the forefront of AI governance by adopting and adhering to international best practices, reinforcing NYS's commitment to responsible and ethical technology deployment.
- **Strengthen NYS ITS Leadership:** Leverage the existing strong organizational structure of NYS ITS to solidify its position as a leader in deploying technology efficiently, innovatively, and responsibly. By integrating these objectives, NYS ITS will not only enhance its operational capabilities but also set a benchmark for other states and organizations in AI governance and innovation.

Scope

This policy applies to all AI applications deployed within NYS, with a particular focus on high-stakes domains such as healthcare, transportation, public safety, employment, criminal justice, and human rights. The framework is designed to be adaptive, scaling oversight mechanisms according to the risk levels of AI applications.

Current Landscape of AI Governance

[The EU AI Act](#) provides a classification of AI risks into unacceptable, high, and limited categories, offering a foundational structure for risk-tiered regulation (European Commission, 2021). However, the unique challenges faced by NYS require a tailored framework that addresses state government needs and contexts. The current landscape of AI governance in the United States is fragmented, with various agencies and stakeholders involved in regulating different aspects of AI (National Security Commission on Artificial Intelligence, 2021). This brief proposes a flexible and robust framework adaptable to the evolving AI landscape. Current approaches often fail to account for the unique characteristics of AI systems, such as their ability to learn and adapt, leading to regulatory gaps (Whittlestone et al., 2019). Furthermore, the rapid pace of AI development makes it difficult for traditional regulatory processes to keep up, necessitating a more agile and responsive approach. This fragmentation can lead to inconsistencies and gaps in oversight, highlighting the need for a cohesive and adaptive framework (Scherer, 2019).

An adaptive framework ensures regulatory oversight is proportionate to risk, enabling effective risk management tailored to each AI system's context and potential impact. It promotes innovation by reducing burdens on low-risk applications. A tiered approach, similar to the EU AI Act, allows for a more nuanced and effective regulatory strategy (European Commission, 2021). This framework also recognizes the need for ongoing evaluation and adjustment, ensuring that regulations remain relevant and effective in the face of rapid technological change. Context-specific regulation is crucial for emerging technologies, allowing frameworks to adapt to unique challenges and opportunities (Marchant et al., 2011). By adopting a dynamic approach, NYS can foster a regulatory environment that encourages responsible AI innovation while safeguarding against potential harms.

ITS, as the leading state agency for technology solutions, plays a pivotal role in overseeing and supporting AI governance across NYS. They provide expert consultation and assistance to other state entities, helping them identify and assess AI solutions effectively. By leveraging frameworks like the [NIST AI Risk Management Framework](#), ITS ensures that potential risks are proactively identified and mitigated before deployment, maintaining our position as a trusted technology partner in navigating the complexities of AI integration.

Legislative Initiatives

Stop Deepfakes Act (A6540)) – Pending: In Committee Assembly

- **Provenance Data:** Requires generative AI providers to include details on synthetic content, specifying its origin, creation, and any modifications to ensure transparency and accountability.
- **Generative AI Systems:** Defines these systems and mandates the inclusion of provenance data to maintain transparency.
- **Penalties:** Establishes fines up to \$100,000 for intentional or grossly negligent violations to enforce compliance.

The Stop Deepfakes Act addresses concerns about synthetic content generated by AI by mandating that providers include provenance data on such content. This ensures transparency and accountability, allowing the public to distinguish AI-generated material from authentic content. By imposing penalties for non-compliance, the act reinforces the importance of adhering to these transparency measures. Incorporating these provisions can enhance public trust in digital media and mitigate risks associated with misinformation and deception.

LOADING Act (S7543A) – Signed by the Governor

- **Automated Decision-Making Systems:** Defines these systems and restricts their use in high-stakes areas, such as public assistance benefits, unless explicitly authorized by law.
- **Impact Assessments:** Requires regular evaluations of automated decision-making systems to assess fairness, bias, cybersecurity, and public health risks.
- **Disclosure and Oversight:** Mandates state agencies to disclose their use of these systems and conduct impact assessments before implementation.

The Legislative Oversight of Automated Decision-Making in Government Act (LOADING Act) aims to ensure that automated decision-making systems used by state agencies are fair, unbiased, and transparent. It mandates regular impact assessments to evaluate risks and benefits, particularly in critical domains like public assistance and civil liberties. By requiring disclosure and oversight, the LOADING Act promotes responsible and ethical AI use within state government, aligning with broader goals of accountability and transparency in AI governance.

Challenges in High-Stakes Domains

In high-stakes domains such as healthcare, transportation, public safety, employment, criminal justice, human rights, and finance, the consequences of AI failures can be severe and far-reaching. These domains require meticulous attention to detail and robust safeguards to ensure that AI technologies enhance, rather than compromise, public services.

Healthcare

In healthcare, AI algorithms are increasingly used for diagnostic purposes, treatment recommendations, and patient management. However, biased algorithms can perpetuate and exacerbate existing health disparities. For example, a study found that a widely used algorithm for predicting healthcare needs exhibited significant racial bias, leading to fewer resources being allocated to Black patients compared to white patients with the same level of need (Obermeyer et al., 2019). Such biases can result in inequitable treatment outcomes, undermining trust in the healthcare system.

Another study published in PLOS Global Public Health highlights how racial differences in laboratory testing rates can introduce bias into AI models used in healthcare settings (Chang et al.). The research found significant disparities in testing rates between Black and White patients for common laboratory tests, which can lead to biased AI models that may underestimate or overestimate health risks for certain racial groups, perpetuating healthcare inequities. Addressing these biases is crucial for ensuring fair and equitable healthcare delivery, emphasizing the need for careful consideration of how AI models are trained and the data they rely on, especially in high-stakes areas like healthcare. By acknowledging and mitigating these biases, AI can be used to improve healthcare outcomes for all patients.

Transportation

Autonomous vehicles (AVs) hold the promise of revolutionizing transportation by reducing human error, a leading cause of accidents. However, the deployment of AVs introduces significant risks that must be carefully managed. Malfunctions in AV systems can lead to accidents with potentially fatal consequences. For example, in 2018, an Uber self-driving car fatally struck a pedestrian in Arizona due to a failure in the vehicle's detection systems (NTSB, 2019). This tragic incident underscores the critical need for ensuring the safety and reliability of AVs to prevent such events.

Beyond AVs, the integration of AI into transportation systems aims to enhance safety and efficiency, but it also presents challenges that require meticulous attention. AI-powered dashcams and driver coaching systems have demonstrated the potential to reduce accidents and unsafe driving behaviors. However, these systems must be rigorously tested and continuously monitored to ensure they function as intended and do not introduce new risks.

Intelligent traffic management systems use AI to optimize traffic flow and reduce congestion, which can minimize the risk of accidents. However, any failure in these systems could lead to unforeseen hazards. Additionally, AI is employed for predictive maintenance of transportation infrastructure, helping to detect and address potential issues before they result in accidents. Yet, the reliability of these AI systems is paramount to their effectiveness.

These examples highlight the transformative potential of AI in transportation, but they also serve as a cautionary tale. The deployment of AI in high-stakes sectors like transportation necessitates robust regulatory frameworks, continuous monitoring, and stringent safety protocols. Without careful consideration and rigorous oversight, the use of AI in critical applications can lead to fatal consequences, underscoring the importance of a measured and responsible approach to AI integration.

Public Safety

AI applications in public safety, such as predictive policing, aim to enhance law enforcement by identifying potential crime hotspots or individuals likely to commit crimes. However, these systems raise significant ethical and privacy concerns. Predictive policing algorithms have been criticized for reinforcing existing biases in the criminal justice system, leading to disproportionate policing of minority communities (Selbst et al., 2019). For example, in cities like Los Angeles and Chicago, predictive policing tools have been scrutinized for potentially targeting neighborhoods based on historical crime data, which may reflect systemic biases rather than current risk factors (Richardson et al., 2019).

Moreover, the use of facial recognition technology in public spaces has sparked debates about surveillance and civil liberties. While facial recognition can assist in identifying suspects and enhancing security, its deployment raises concerns about mass surveillance and the potential for misuse. Cities like San Francisco and Boston have implemented bans or moratoriums on the use of facial recognition technology by law enforcement due to these concerns (Hill, 2020).

Additionally, AI-driven emergency response systems are being developed to improve public safety by rapidly detecting and responding to incidents. However, the reliance on AI for critical decision-making in emergencies requires robust safeguards to prevent system failures that could compromise public safety. For instance, AI systems used for wildfire detection must be highly accurate to avoid false alarms that could divert resources from actual emergencies (Jacob et al., 2021)..

These examples illustrate the complexities and risks associated with deploying AI in public safety contexts. While AI has the potential to enhance safety and efficiency, it also introduces ethical, privacy, and operational challenges that must be carefully addressed. Robust regulatory frameworks, continuous monitoring, and transparent governance are essential to ensure that AI systems are deployed responsibly and ethically, balancing innovation with the protection of public interests.

Employment

AI-driven hiring tools are designed to streamline recruitment processes by evaluating candidates based on predefined criteria. However, these tools can inadvertently perpetuate biases present in historical hiring data. For instance, Amazon had to scrap an AI-powered hiring tool after discovering that it systematically discriminated against women by favoring male candidates, as it was trained on historical hiring data reflecting past biases (Dastin, 2018). This incident highlights the critical importance of ensuring fairness in AI-driven hiring processes to prevent discriminatory outcomes.

Beyond Amazon's experience, other cases have raised concerns about the use of AI in hiring practices. For example, HireVue, a company offering AI-driven video interview analysis, faced scrutiny over potential biases in its algorithms. Critics argued that the system might disadvantage candidates based on factors unrelated to job performance, such as facial expressions or speech patterns, which could be influenced by cultural or socioeconomic backgrounds (Harwell, 2019).

Additionally, AI-powered resume screening tools have been criticized for potentially excluding qualified candidates who do not fit the mold of traditional hiring patterns. These tools often rely on historical data that may not account for diverse talent pools, leading to the exclusion of candidates from underrepresented groups. Companies must carefully evaluate the criteria used in AI hiring tools to ensure they do not inadvertently perpetuate existing inequalities.

These examples underscore the need for vigilance and transparency in the deployment of AI in hiring practices. While AI has the potential to enhance efficiency and objectivity in recruitment, it also introduces risks of perpetuating or even amplifying biases present in historical data. To mitigate these risks, organizations must implement robust safeguards, conduct regular audits of AI systems, and ensure that the criteria used for evaluation are fair and inclusive.

Criminal Justice

AI applications in the criminal justice system, such as predictive policing and criminal risk assessment tools, aim to enhance law enforcement and judicial decision-making. However, these systems have been criticized for reinforcing existing biases, raising significant ethical and legal concerns. Predictive policing algorithms, for example, have been shown to disproportionately target minority communities, leading to over-policing and potential violations of civil rights (Selbst et al., 2019). These algorithms often rely on historical crime data, which may reflect and amplify systemic biases rather than accurately predicting current crime risks. Criminal risk assessment tools, used to inform sentencing and parole decisions, have also faced scrutiny for perpetuating biases. For instance, the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) tool has been criticized for disproportionately assigning higher risk scores to Black defendants compared to white defendants with similar profiles (Angwin et al., 2016). Such biases can lead to harsher sentences and longer incarceration periods for minority individuals, exacerbating disparities within the criminal justice system.

Moreover, the use of facial recognition technology in law enforcement has raised concerns about privacy and civil liberties. While facial recognition can assist in identifying suspects, its deployment has been linked to issues of mass surveillance and potential misuse. Cities like San Francisco and Boston have implemented bans on the use of facial recognition technology by law enforcement due to these concerns (Hill, 2020). The inaccuracies in facial recognition systems, particularly for individuals with darker skin tones, further compound these issues, potentially leading to wrongful arrests and misidentifications (Buolamwini & Gebru, 2018). Ensuring that AI tools in criminal justice are fair and unbiased is essential to maintaining public trust and upholding human rights. Transparent governance, robust auditing mechanisms, and continuous evaluation of AI systems are necessary to mitigate biases and prevent the perpetuation of injustices. Stakeholders must engage in ongoing dialogues to address ethical considerations and ensure that AI technologies are deployed responsibly within the criminal justice system.

Human Rights

The use of AI in public safety and surveillance raises significant human rights concerns. Facial recognition technology, for instance, has been deployed in public spaces without adequate oversight, leading to privacy invasions and potential misuse. Critics argue that the widespread use of facial recognition can result in constant surveillance, chilling effects on freedom of expression, and disproportionate targeting of marginalized communities (Ferguson, 2019). Additionally, the lack of transparency in how these systems are implemented and regulated exacerbates concerns about accountability and potential abuses of power.

Moreover, AI systems used for monitoring and controlling social behavior, such as social credit systems, can infringe on individual freedoms and human rights. In China, the social credit system utilizes AI to assess citizens' behavior and assign scores that can impact their access to services and opportunities (Creemers, 2018). This system has been criticized for its potential to enforce conformity and suppress dissent, raising questions about the ethical implications of using AI for social control.

The deployment of AI in law enforcement also presents challenges to human rights. Predictive policing algorithms, which aim to forecast crime patterns and identify potential offenders, have been shown to perpetuate and amplify existing biases in policing (Richardson et al., 2019). These systems often rely on historical crime data that may reflect systemic biases, leading to disproportionate surveillance and enforcement in minority communities.

Furthermore, the use of AI in border control and immigration enforcement has raised concerns about the erosion of human rights. AI-powered systems are increasingly used to monitor and analyze migrant populations, raising issues of privacy, discrimination, and due process (Cheney-Lippold, 2017). The opaque nature of these technologies makes it difficult for individuals to challenge decisions made by AI systems, further complicating efforts to ensure fairness and accountability.

Balancing the benefits of AI with the need to protect human rights is essential in deploying these technologies responsibly. Robust regulatory frameworks, transparency, and public engagement are crucial to mitigate the risks associated with AI in public safety and surveillance. Ensuring that AI systems are designed and implemented with a focus on human rights can help prevent abuses and promote a more equitable and just society.

Finance

AI technologies are increasingly being integrated into the financial sector, transforming areas such as fraud detection, risk assessment, and algorithmic trading. While AI offers significant benefits, such as improved efficiency and accuracy, it also introduces new risks and challenges. For example, AI-driven fraud detection systems can enhance security by identifying suspicious activities in real-time. However, these systems must be carefully designed to avoid false positives, which could lead to unnecessary disruptions for customers.

In risk assessment, AI algorithms can analyze vast amounts of data to predict financial risks and inform investment decisions. However, biases in these algorithms can lead to unfair outcomes, such as denying loans to individuals based on factors unrelated to their creditworthiness. Ensuring transparency and fairness in AI-driven risk assessment is crucial to maintaining trust in the financial system.

Algorithmic trading, which uses AI to execute trades at high speeds, can increase market efficiency but also poses risks of market manipulation and systemic instability. Regulatory frameworks must be in place to monitor and control these activities, ensuring that AI-driven trading does not compromise market integrity.

Additionally, the use of AI in customer service, such as chatbots and virtual assistants, can improve customer experiences by providing quick and personalized support. However, these systems must be designed to protect customer data and ensure privacy. Transparency in how AI systems operate and how customer data is used is essential to building trust with consumers.

These examples highlight the need for robust regulatory frameworks and continuous monitoring in the financial sector to ensure that AI technologies are deployed responsibly. By addressing potential risks and ensuring transparency, fairness, and accountability, the financial industry can harness the benefits of AI while protecting the interests of consumers and maintaining market stability.

Recommendations & Implementation Plan

The rapid advancement of Artificial Intelligence (AI) technologies necessitates a robust governance framework to ensure their safe, ethical, and responsible deployment. This policy recommendation is based on the European Union's AI Act and the National Institute of Standards and Technology's (NIST) AI Risk Management Framework (AI RMF). It aims to establish guidelines for the acceptable use of AI technologies by New York State Entities (SEs), fostering innovation while protecting privacy, managing risks, and promoting accountability, safety, and equity.

1. Establish Risk-Tiered Classification System

Implementing a risk-tiered classification system for AI applications allows for a structured approach to managing and mitigating risks associated with AI deployment. This system categorizes AI applications based on their potential impacts on public safety, health, and fundamental rights.

1.1 Unacceptable Risk

Definition: AI systems that pose unacceptable risks to public safety and fundamental rights should be prohibited. These systems are deemed too dangerous to be deployed under any circumstances.

Examples:

- **AI-Driven Social Scoring Systems:** Systems that evaluate and score individuals based on their behavior and personal data, potentially infringing on individual freedoms and privacy rights.
- **Autonomous Weapons:** AI systems that can select and engage targets without human oversight, raising significant ethical and safety concerns.

Guidance:

- Ensure that any AI system falling under this category is immediately halted and prohibited from further development or deployment.
- Establish clear internal policies and communication channels to identify and report such systems promptly.

1.2 High Risk

Definition: AI systems with significant potential impacts on safety, health, and fundamental rights require stringent testing and certification processes before deployment.

Examples:

- **Medical Diagnostic Tools:** AI applications used for diagnosing medical conditions, where errors could lead to misdiagnosis and harm to patients.
- **Autonomous Vehicles:** Self-driving cars and transportation systems that operate without human intervention, posing risks to public safety if not properly regulated.
- **Critical Infrastructure Management:** AI systems managing power grids, water supply, or transportation networks, where failures could have catastrophic consequences.

Guidance:

- Implement rigorous pre-deployment testing, including simulations and real-world pilot programs, to evaluate system performance and safety.
- Conduct continuous monitoring and regular audits to ensure ongoing compliance and performance. Establish certification processes involving independent third-party auditors to validate safety and efficacy.

1.3 Limited Risk

Definition: AI systems with minimal risks should adhere to transparency and user notification requirements to ensure users are informed about AI involvement in decision-making processes.

Examples:

- **AI-Powered Chatbots:** Systems designed to assist users through automated conversations, which should clearly disclose their AI nature to users.
- **Personalized Advertising Algorithms:** AI tools used to tailor advertisements based on user data, requiring transparency in how data is used and ensuring user consent.

Guidance:

- Ensure transparency by clearly communicating the use of AI in user interactions and obtaining necessary consents.
- Implement user notification requirements to inform individuals when AI systems are involved in decision-making processes affecting them.
- Conduct periodic reviews to assess the ongoing risks and ensure compliance with transparency standards.

Recommendations & Implementation Plan (continued)

2. Strengthen Risk Assessment, Integrate Classification and Implement AI Registry

To ensure the safe and effective deployment of AI systems, it is essential to implement a comprehensive risk assessment, classification, and registration process. This approach should be integrated into existing enterprise application review processes to streamline risk management and deployment, aligning with best practices in AI governance (EU AI Act, 2024; NIST AI RMF, 2023).

2.1 Risk Assessment: Evaluate the potential risks associated with AI systems to ensure they meet safety and performance standards before deployment. Compliance with relevant regulations and standards should be verified through regular audits and performance evaluations (NIST AI RMF, 2023). This step is crucial for identifying potential vulnerabilities and ensuring that AI systems are robust and reliable before they are introduced into operational environments.

2.2 Classification: Utilize the existing enterprise application review process to classify AI systems based on their risk levels. This classification will guide the next steps for deployment, including necessary pre-deployment testing and ongoing monitoring. By categorizing AI systems according to their potential impact, organizations can allocate resources effectively and prioritize high-risk systems for more stringent oversight (EU AI Act, 2024).

2.3 AI Registry: The objective of maintaining a registry of AI systems is to track their deployment, performance, and compliance status effectively. This involves creating and regularly updating an AI inventory that includes details such as risk classification, deployment status, and compliance records. The inventory should be kept current to reflect any changes in the status or risk levels of AI systems. This registry will require effort and build upon the current enterprise applications approval process, ensuring that ITS remains at the forefront of managing AI use. Transparency and reporting mechanisms should be established to continuously improve systems based on real-world outcomes, fostering trust and accountability in AI deployment (Holstein et al., 2019).

By integrating these processes, organizations can effectively manage AI risks, ensure compliance with safety standards, and foster trust in AI technologies among stakeholders. This comprehensive approach supports the responsible deployment of AI systems, particularly in critical sectors, while promoting innovation and public safety.

3. Implement Proportional Oversight Mechanisms

Implementing oversight mechanisms that scale with the risk levels of AI applications ensures that AI systems are managed appropriately based on their potential impact. This approach allows for proportional regulation, fostering innovation while mitigating risks.

3.1 Unacceptable Risk

- **Prohibition:** AI systems classified as unacceptable risk should be prohibited from development and deployment. This includes systems that pose significant threats to public safety, fundamental rights, or ethical standards.
- **Policy Enforcement:** Establish clear policies and communication channels to identify and halt the development of such systems promptly. Ensure that any identified unacceptable risk systems are immediately reported and prohibited from further use.
- **Regulatory Compliance:** Ensure compliance with international standards and regulations that prohibit the deployment of high-risk AI systems without adequate safeguards (EU AI Act, 2024).

3.2. High Risk

- **Regulatory Sandboxes:** Create controlled environments for testing high-risk AI systems before full-scale deployment. These sandboxes allow for innovation while ensuring that potential risks are managed and mitigated (Jenkins & Kosseff, 2020).
- **Strict Testing and Certification:** Implement rigorous pre-deployment testing, including simulations, stress tests, and real-world pilot programs, to evaluate system performance under various conditions. Verification processes should involve independent third-party audits to ensure objectivity and transparency.
- **Continuous Monitoring:** Establish frameworks for ongoing monitoring, including regular audits and performance evaluations, to ensure AI systems remain safe and effective over time.
- **Liability Frameworks:** Develop frameworks to hold developers accountable for AI harms, ensuring responsible innovation and providing mechanisms for compensating affected parties (Dignum, 2019).

Recommendations & Implementation Plan (continued)

3. Implement Proportional Oversight Mechanisms

3.3.Limited Risk

- Transparency Requirements: Ensure that AI systems with limited risk adhere to transparency and user notification requirements. This includes clearly communicating the use of AI in user interactions and obtaining necessary consents.
- Periodic Reviews: Conduct periodic reviews to assess the ongoing risks and ensure compliance with transparency standards. These reviews should focus on evaluating system performance and addressing any emerging issues.
- User Engagement: Foster user engagement through transparent communication and participatory processes. This engagement builds trust and ensures that AI systems align with public values and expectations (Ribeiro et al., 2016).

By establishing these oversight mechanisms, organizations can effectively manage AI risks, promote responsible innovation, and ensure the safe and ethical deployment of AI technologies across various risk levels.

4. Foster Inter-State Cooperation

Fostering inter-state cooperation is crucial for creating a cohesive and effective AI governance framework. By leveraging existing relationships with other state bodies, such as through community links and the National Association of State Chief Information Officers (NASCIO), New York State (NYS) can share best practices and harmonize regulatory approaches. This collaboration ensures that AI technologies are deployed responsibly and ethically across the nation, transcending geographical boundaries.

Through these partnerships, NYS can lead conversations and encourage the exchange of successful strategies and lessons learned, helping states avoid common pitfalls and adopt proven methods for effective AI regulation. A harmonized regulatory environment ensures consistency in AI governance across state lines, which is particularly important for technologies that operate beyond individual state jurisdictions.

By actively engaging with other states and fostering a collaborative approach, NYS can drive innovation, enhance public trust in AI, and ensure that AI technologies benefit all stakeholders while mitigating potential risks. This leadership role positions NYS as a model for responsible AI governance, influencing national and potentially international standards and practices. Through platforms like NASCIO, NYS can facilitate the development of consistent standards and guidelines for AI deployment, reinforcing its position as a global leader in ethical and innovative AI governance.

5. Promote Organizational Engagement and Transparency

Encouraging organizational engagement in AI governance through transparent communication and participatory processes is crucial for building trust and ensuring that AI systems align with public values and expectations (Ribeiro et al., 2016). This engagement involves creating avenues for stakeholders within state agencies to actively participate in shaping AI policies and practices. Initiatives to foster engagement can include:

- Stakeholder Workshops: Organize workshops involving diverse stakeholders, including government agencies and industry experts, to discuss AI use cases, risks, and benefits. These workshops can help identify potential challenges and opportunities associated with AI deployment.
- Educational Campaigns: Launch educational campaigns to raise awareness about AI technologies and their implications within state agencies. These campaigns should provide clear, accessible information to help stakeholders understand how AI systems work and how they might impact operations.
- Statewide Training Programs: Develop and implement statewide training programs led by ITS to educate government employees on AI governance, risk management, and best practices. This training will ensure that all state agencies are equipped with the knowledge and skills necessary to responsibly deploy and manage AI systems.
- Collaborative Framework: Establish a collaborative framework where all state agencies can work together under the leadership of ITS. This framework should facilitate the sharing of best practices, lessons learned, and innovative approaches to AI governance. Regular inter-agency meetings and collaborative projects can help build a cohesive strategy for AI deployment across the state.

By promoting organizational engagement and transparency, ITS can lead the way in fostering a collaborative environment where state agencies work together to responsibly and effectively manage AI technologies. This approach not only builds trust among stakeholders but also positions New York State as a leader in ethical and innovative AI governance.

Impact Assessment

The implementation of the Adaptive Risk-Tiered Framework for AI deployment and application in New York State (NYS) represents a transformative opportunity to lead the nation in responsible AI governance. By adopting this framework, NYS can set a new standard for balancing innovation with robust risk management, ensuring that AI technologies enhance public services and protect citizens' rights.

Enhanced Safety and Public Trust

The framework's emphasis on rigorous risk assessment and mitigation will significantly reduce the likelihood of AI-related incidents in critical sectors such as healthcare, transportation, and public safety. By proactively identifying and addressing potential risks, NYS can prevent catastrophic failures, such as those seen with autonomous vehicle accidents or biased healthcare algorithms. This proactive approach will not only safeguard public safety but also build trust among citizens, fostering greater acceptance and support for AI technologies.

Innovation and Economic Growth

By implementing a tiered regulatory approach, NYS can foster an environment that encourages innovation while mitigating risks. Low-risk AI applications will face fewer regulatory hurdles, allowing startups and established companies to develop and deploy new technologies rapidly. This balanced approach will position NYS as a hub for AI innovation, attracting investment and driving economic growth. The framework's adaptability ensures that regulations remain relevant and effective, supporting continuous technological advancement.

Global Leadership and Influence

Adopting this framework will position NYS as a global leader in AI governance. By aligning with international best practices, such as the EU AI Act and the NIST AI Risk Management Framework, NYS can influence global standards and shape the future of AI regulation. This leadership role will enhance NYS's reputation and provide opportunities for collaboration with other jurisdictions, further strengthening its influence on the global stage.

Ethical and Equitable AI Deployment

The framework's focus on transparency, accountability, and ethical considerations ensures that AI technologies are deployed responsibly and equitably. By addressing biases in AI systems and promoting fairness, NYS can mitigate the risks of perpetuating existing inequalities. This commitment to ethical AI deployment will enhance public trust and demonstrate NYS's dedication to protecting the rights and interests of all citizens.

Robust and Adaptive Governance

The proposed framework provides a structured yet flexible approach to AI governance, allowing NYS to adapt to the rapidly evolving technological landscape. By establishing clear guidelines for risk assessment, classification, and oversight, the framework ensures that AI systems are subject to appropriate levels of scrutiny and regulation. This adaptive governance model will enable NYS to respond effectively to emerging challenges and opportunities, maintaining its position at the forefront of AI innovation.

Conclusion

The rapid evolution of artificial intelligence (AI) presents both unparalleled opportunities and significant governance challenges. New York State (NYS) is poised to lead the nation in establishing a robust and adaptive framework for AI deployment. By implementing the proposed Adaptive Risk-Tiered Framework, NYS can set a national standard for responsible AI governance, harmonizing innovation with rigorous safety measures. The NYS Office of Information Technology Services (ITS) is instrumental in this effort, leveraging its technological expertise and commitment to ethical AI deployment to ensure that AI systems are used responsibly and securely across all sectors.

Through collaboration with organizational and national stakeholders, NYS can share best practices and harmonize regulatory approaches, enhancing AI governance both within the state and across the nation. This initiative will position NYS as a global leader in ethical and innovative AI deployment, fostering a transparent and participatory governance process that aligns AI systems with public values and expectations. By championing responsible innovation while safeguarding public safety and fundamental rights, NYS will not only benefit its residents but also set a precedent for other states, establishing itself as a beacon of responsible AI governance on the global stage.

Appendix A : Risk Classification Criteria & Examples

Risk Category	Criteria	Examples
Unacceptable Risk	- AI systems that pose a clear threat to the safety, livelihoods, and fundamental rights of individuals. These systems are prohibited due to their potential for significant harm. - Real-time remote biometric identification systems in public spaces. - Systems that exploit vulnerabilities, leading to harmful behavior. - Untargeted scraping of internet or CCTV material to create facial recognition databases.	- Social scoring systems that evaluate individuals based on behavior or personal traits. - AI-driven manipulation that influences behavior without consent (e.g., subliminal advertising). - Predictive policing that assesses future criminal offenses based on personal traits.
High Risk	- AI systems that can pose serious risks to health, safety, or fundamental rights. These systems require strict regulatory compliance and oversight before deployment. - AI solutions used in education that determine access to education or career paths. - AI tools for employment and worker management. - AI systems used in migration and border control management.	- AI safety components in critical infrastructures (e.g., transport, energy). - AI-based recruitment tools that sort CVs for job applications. - AI systems used in law enforcement for evaluating evidence. - AI applications in the administration of justice and democratic processes.
Limited Risk	- AI systems that present a risk of manipulation or deceit. These systems must be transparent, and users must be informed about their interaction with the AI. - Generative AI systems that create deep fakes, which must be clearly denoted as such. - Systems that could influence user decisions without clear disclosure.	- Chatbots that interact with users, requiring disclosure of their AI nature. - AI-generated content that must be identifiable to users.

Appendix B: Comparative Analysis of Alternate Regulatory Approaches

Introduction

The regulation of artificial intelligence (AI) presents unique challenges due to its rapid evolution and diverse applications. This appendix provides a comparative analysis of alternate regulatory approaches, including self-regulation, centralized regulatory authorities, and uniform regulatory standards. By examining the strengths and weaknesses of each approach, this analysis highlights why the proposed adaptive risk-tiered framework is the most promising for AI governance in New York State (NYS).

Self-Regulation

Self-regulation involves allowing AI developers and deployers to establish and adhere to their own standards and guidelines. This approach relies on industry-led initiatives and voluntary compliance with best practices.

Strengths:

- Flexibility: Allows for rapid adaptation to technological advancements without the need for legislative changes.
- Industry Expertise: Leverages the technical knowledge and experience of industry professionals.
- Innovation: Encourages innovation by reducing regulatory burdens and allowing for experimentation.

Weaknesses:

- Inconsistency: May lead to inconsistent standards and practices across different organizations.
- Accountability: Lacks enforceable mechanisms to hold developers accountable for harms.
- Public Trust: May not adequately address public concerns and expectations regarding AI safety and ethics.

Centralized Regulatory Authorities

Centralized regulatory authorities involve establishing a single governmental body responsible for overseeing AI development and deployment. This approach centralizes decision-making and enforcement powers.

Strengths:

- Consistency: Ensures uniform standards and practices across the industry.
- Accountability: Provides clear mechanisms for holding developers accountable for harms.
- Public Trust: Can address public concerns more effectively through centralized oversight.

Weaknesses:

- Bureaucracy: May lead to slow adaptation to technological changes due to bureaucratic processes.
- Resource Intensive: Requires significant resources for effective oversight and enforcement.
- Innovation Barriers: May impose uniform restrictions that hinder innovation in low-risk areas.

Uniform Regulatory Standards

Uniform regulatory standards involve establishing a set of rules and guidelines that apply equally to all AI applications, regardless of their risk levels. This approach aims to provide a level playing field for all developers.

Strengths:

- Fairness: Ensures that all developers are subject to the same regulatory requirements.
- Simplicity: Simplifies compliance by providing clear and consistent standards.
- Public Trust: Can enhance public trust by ensuring that all AI systems are subject to the same level of scrutiny.

Weaknesses:

- Over-Regulation: May impose unnecessary burdens on low-risk AI applications, stifling innovation.
- Under-Regulation: May fail to adequately address the unique risks posed by high-risk AI applications.
- Adaptability: May not adapt well to the rapid pace of technological change.

Proposed Adaptive Risk-Tiered Framework

The proposed adaptive risk-tiered framework involves categorizing AI applications into different risk tiers and tailoring regulatory oversight to the specific risks posed by each tier. This approach allows for flexible and dynamic risk management strategies.

Strengths:

- Proportionality: Ensures that regulatory oversight is proportionate to the risks posed by different AI applications.
- Adaptability: Allows for rapid adaptation to technological advancements through continuous monitoring and updates.
- Innovation: Promotes innovation by reducing unnecessary regulatory burdens on low-risk applications.
- Public Trust: Enhances public trust by providing transparency and accountability through a publicly accessible AI registry.

Weaknesses:

- Complexity: May require more sophisticated regulatory mechanisms and expertise.
- Implementation Challenges: May face challenges in implementing and enforcing tiered regulatory standards.

Conclusion

The proposed adaptive risk-tiered framework offers a balanced approach to AI governance, addressing the limitations of self-regulation, centralized authorities, and uniform standards. By tailoring regulatory oversight to the specific risks posed by different AI applications, this framework promotes responsible innovation while ensuring public safety and trust. This approach is particularly well-suited to the dynamic and evolving nature of AI technologies, making it the most promising regulatory strategy for AI governance in NYS.

References

1. Calo, R. (2014). *Digital market manipulation*. George Washington Law Review, 82(4), 995-1037. Retrieved April 6, 2025, from https://www.gwlr.org/wp-content/uploads/2014/10/Calo_82_41.pdf
2. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica. Retrieved April 6, 2025, from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
3. Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. Conference on Fairness, Accountability and Transparency, 1, 77-91. Retrieved April 6, 2025, from <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>
4. Chang T, Nuppnau M, He Y, Kocher KE, Valley TS, Sjoding MW, Wiens J. *Racial differences in laboratory testing as a potential mechanism for bias in AI: A matched cohort analysis in emergency department visits*. PLOS Glob Public Health. 2024 Oct 30;4(10):e0003555. doi: 10.1371/journal.pgph.0003555. PMID: 39475953; PMCID: PMC11524489. Retrieved April 6, 2025, from <https://pmc.ncbi.nlm.nih.gov/articles/PMC11524489/>
5. Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. NYU Press.
6. Crawford, K. (2021). *The AI governance challenge*. Nature, 593(7858), 171-175.
7. Creemers, R. (2018). *China's social credit system: An ecosystem of trust*. In China's Social Credit Systems (pp. 1-22). Brill. Retrieved April 6, 2025, from http://www.iberchina.org/files/2018/social_credit_china.pdf
8. Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. Retrieved April 6, 2025, from <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>
9. Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Moral Way*. Springer.
10. Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. arXiv preprint arXiv:1702.08608. Retrieved April 6, 2025, from <https://arxiv.org/pdf/1702.08608>
11. European Commission. (2021). *Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts*. COM(2021) 206 final. Retrieved April 6, 2025, from <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>
12. Ferguson, A. G. (2019). *The rise of Big Data policing: Surveillance, race, and the future of law enforcement*. NYU Press.
13. Hill, K. (2020). *San Francisco bans facial recognition technology*. The New York Times. Retrieved April 6, 2025, from <https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html>
14. Holstein, K., Cohen, M., Austen, C., Chatzimiltiadis, A., Daumé III, H., & Wallace, J. R. (2019). *Implications of user agency in AI-infused decision making: Rethinking shared decision making in the era of artificial intelligence*. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1-13.
15. Jacob, R., Hertweck, J., & Karjalainen, T. (2021). *AI for wildfire detection: Challenges and opportunities*. Journal of Environmental Management, 287, 112345. Retrieved April 6, 2025, from <https://fireecology.springeropen.com/AIWFM>
16. Jenkins, S., & Kosseff, J. (2020). *The role of regulatory sandboxes and innovation hubs in an era of rapid technological change*. Journal of Internet Law, 23(10), 3-12.
17. Jobin, A., Ienca, M., & Vayena, E. (2019). *The global landscape of AI ethics guidelines*. Nature Machine Intelligence, 1(9), 389-399.
18. Marchant, G. E., Sylvester, D. J., & Abbott, K. W. (2011). *Global governance of nanotechnology*. Routledge.
19. Mittelstadt, B. D. (2019). *Principles alone cannot guarantee ethical AI*. Nature Machine Intelligence, 1(11), 501-507. Retrieved April 6, 2025, from <https://doi.org/10.1038/s42256-019-0114-4>
20. National Security Commission on Artificial Intelligence. (2021). *Final report*. Retrieved April 6, 2025, from https://assets.foleion.com/eu-central-1/de-uploads-7e3kk3/48187/nscai_full_report_digital.04d6b124173c.pdf
21. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). *Dissecting racial bias in an algorithm used to manage the health of populations*. Science, 366(6464), 447-453.
22. Reuters. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Retrieved April 6, 2025, from <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>
23. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why should I trust you?": Explaining the predictions of any classifier*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.
24. Richardson, R., Schultz, J., & Crawford, K. (2019). *Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice*. New York University Law Review, 94, 1-73.
25. Scherer, M. U. (2019). *Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies*. Harvard Journal of Law & Technology, 32(2), 354-402.
26. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). *Fairness and abstraction in sociotechnical systems*. Proceedings of the Conference on Fairness, Accountability, and Transparency, 59-68.