

Cover Page

To: AI Policy Committee

From: Erica Okeh

Date: 09/20/2024

RE: National Data Privacy and Governance
Act (NDPG Act)

Policy Memorandum Template¹

Erica Okeh

Howard University

With
Howard University & Apart Research

Abstract

Artificial Intelligence, specifically generative AI, is lauded as a cornerstone for economic prosperity and innovation in the United States. However, its growing reliance on user data raises significant concerns about ethical data collection and privacy protection. This memorandum identifies the challenges surrounding AI-driven recommendation engines, emphasizing their role in personalization and economic growth alongside the risks posed to consumer data privacy. It evaluates existing regulations, such as the California Consumer Privacy Act (CCPA), identifies their limitations, and proposes national regulations to bolster data governance and user privacy. Recommendations include creating a National Protection Agency, implementing a "Do Not Track" system, enforcing data governance standards, and leveraging Privacy-Enhancing Technologies (PETs). These measures aim to ensure secure, ethical, and transparent AI systems that foster public trust and innovation.

Keywords: Data Privacy, Recommendation Systems, model evaluations,

1. Introduction

This research examines how AI recommender systems can be regulated to balance economic innovation with consumer privacy. It focuses on policy measures to

¹ Research conducted at the [hackathon name with link], 202X

ensure ethical personal data use while maintaining AI effectiveness. The hypothesis proposes that a unified federal framework, integrating privacy-enhancing technologies and transparent practices, can safeguard consumer data, rebuild trust, and support economic growth driven by AI advancements.

Executive Order 14110 mandates AI's role in promoting prosperity, innovation, and security. Generative AI, in particular, is expected to transform industries and boost economic growth. While the U.S. government supports AI initiatives to strengthen the economy in collaboration with industry, academia, and civil society, trust in AI faces challenges as ethical concerns arise—particularly regarding user data exploitation (Diakopoulos & Koliska, 2025).

AI-driven recommendation systems thrive on this demand, suggesting items, shows, or music based on user behavior. These systems are highly effective and profitable, with personalization boosting revenues by 5–15% and 76% of customers expressing frustration at a lack of personalization. The recommendation systems market, valued at \$6.88 billion in 2024, is projected to triple by 2029 (McKinsey & Company, 2023).

However, this growth raises significant privacy concerns. Machine learning algorithms rely on vast datasets, often including sensitive consumer information, sparking debates about data privacy, consent, and security. Critics highlight vulnerabilities in data collection, storage, and usage, pointing to risks of breaches and unethical practices. Without comprehensive regulation, consumers face diminished autonomy and potential misuse of their data by corporations (IT Convergence, 2024).

Personalization remains a key revenue driver, but its reliance on extensive data collection underscores the need for a secure, transparent regulatory framework. Current challenges stem from inconsistent data governance and enforcement mechanisms. Weak data storage solutions, such as data lakehouses, exacerbate vulnerabilities (IT Convergence, 2024). Without intervention, the consequences could include widespread breaches, consumer disengagement, and reputational harm for AI technologies.

Key stakeholders include consumers, corporations, regulators, and advocacy groups. Consumers seek transparency and control over their data, while corporations prioritize innovation and profit. Regulatory bodies must enforce compliance without stifling growth, though power dynamics currently favor large tech companies that dominate data collection and utilization.

2. Methods

Recommendation engines follow a standard pipeline to operate effectively. According to IBM, these engines typically involve five steps: data collection, storage, analysis, pattern detection, and filtering (IBM, 2023). Data collection encompasses user interactions, such as search history, browsing behaviors, and demographic inputs. This data is stored in repositories like data lakehouses, which,

while cost-effective, have weak governance structures and limited data security measures (IT Convergence, 2024).

During the analysis phase, machine learning algorithms process and train on collected data, identifying patterns and correlations to refine recommendations. While effective, these systems highlight the delicate balance between utility and privacy. The reliance on personal data without robust security measures increases exposure to risks, including data breaches and ethical violations.

Previous analysis of recommendation system methods draws on industry reports, privacy laws, and case studies. Analyzing the policy involves evaluating current frameworks like the CCPA (California consumer privacy act) and identifying its limitations allows for the scalability of proposed policies and the adaptability of corporations to new regulations. Weak data governance measures, such as insufficient encryption in large data centers, amplify risks. Existing solutions like the CCPA are limited in scope and enforcement, leaving gaps in consumer protection.

Potential solutions include self-regulation by corporations, state-level laws, and comprehensive federal regulations. Comparative analysis underscores the limitations of decentralized efforts, highlighting the need for a national framework. A federal approach offers the best chance to standardize practices, enhance transparency, and rebuild trust.

3. Results

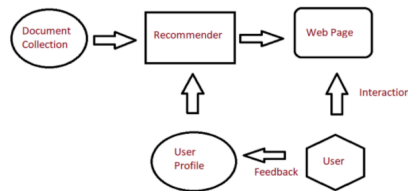


Fig: Recommender System

1

This proposal addresses gaps in existing privacy laws, like the CCPA, by introducing national regulations to strengthen data governance, enforce stricter privacy controls for recommender systems, and raise public awareness of privacy rights. Central to this is creating a National Data Protection Agency to replace overburdened state mechanisms, such as the California Attorney General's office, ensuring consistent privacy enforcement with nationwide jurisdiction and sufficient resources.

A key feature is a nationwide “Do Not Track” signal, enabling users to easily opt out of data collection via browsers or devices. This universal mechanism simplifies the process, curbs invasive practices, and holds companies accountable through enforceable penalties, fostering trust in online interactions.

The proposal also mandates stricter standards for data storage, including standardized metadata, role-based access controls, and end-to-end encryption. Companies would require certifications and regular audits to ensure compliance, addressing risks tied to poor data management (IT Convergence, 2023).

For recommender systems, Privacy-Enhancing Technologies (PETs) like federated learning would protect privacy by keeping user data on local devices and sharing only model updates. Transparency requirements would compel companies to disclose how algorithms function and address risks such as bias or profiling, a critical factor in building trust (Diakopoulos & Koliska, 2025). These measures aim to balance effective system performance with ethical privacy practices.

The proposal includes federal funding for public education campaigns on privacy rights, delivered via schools, social media, and community outreach. A real-time compliance dashboard would improve accountability by showing company compliance statuses and enabling secure reporting of violations, developed with privacy experts to ensure user-friendly and secure design.

Implementation spans three phases: (1) establish legislation and the national agency, (2) enforce data governance standards and the "Do Not Track" signal, and (3) introduce algorithmic transparency and education campaigns. Estimated costs are \$500 million annually for the agency, \$100 million for education, and \$50 million for the compliance dashboard.

Success metrics include reduced privacy violations within five years, universal adoption of the "Do Not Track" signal within two years, and increased consumer awareness measured through surveys. Annual audits and public reporting will ensure accountability. This proposal offers a comprehensive framework to address privacy challenges and build trust in data governance.

4. Discussion

This proposal addresses gaps in existing privacy laws, such as the CCPA, by implementing a national framework to strengthen data governance, enforce privacy controls, and raise public awareness. Short-term outcomes (1-2 years) include increased consumer awareness, widespread adoption of a "Do Not Track" signal, and improved data security. Medium-term impacts (3-5 years) focus on reducing privacy violations, enhancing algorithm transparency, and fostering trust. Long-term goals (5+ years) aim to establish a scalable, standardized privacy framework balancing innovation with ethical practices.

Stakeholders will see varied impacts. Consumers benefit from stronger protections and simplified opt-outs but may lose some personalization. Corporations gain from regulatory clarity but face compliance costs. Governments improve enforcement but incur setup expenses. Potential unintended consequences, like increased compliance

burdens for small businesses or reduced personalization, are mitigated through subsidies, phased implementation, and public education.

The proposal is politically feasible, given public demand for privacy protections, and economically viable with estimated costs of \$650 million annually. Technically, the use of modern encryption, metadata standards, and Privacy-Enhancing Technologies is practical, with support for legacy system updates. Administratively, a National Data Protection Agency ensures efficient oversight and compliance.

This framework strengthens privacy protections, rebuilds public trust, and promotes ethical innovation, positioning the U.S. as a global leader in data governance.

5. Conclusion

The rapid expansion of AI-powered recommendation systems, fueled by sensitive user data, demands immediate national regulations to safeguard consumer privacy and trust. Current protections like the CCPA are insufficient, leaving millions vulnerable to unethical data practices, breaches, and exploitation. By implementing a comprehensive framework—including a National Data Protection Agency, a universal "Do Not Track" signal, and stricter data governance standards—we can protect user rights, ensure accountability, and promote ethical AI innovation. These measures will strengthen public trust, empower consumers with data autonomy, and foster economic growth through responsible technology development. The time to act is now.

6. References

California Office of the Attorney General. (n.d.). *California Consumer Privacy Act (CCPA)*. Retrieved from <https://oag.ca.gov/privacy/ccpa>

Diakopoulos, N., & Koliska, M. (2025). *Prospective algorithmic accountability and the role of the news media*. In M. Noorman & M. Verdicchio (Eds.), *Computer ethics across disciplines: Algorithmic accountability and AI through Deborah G. Johnson's lens*. Springer.

GetRightData. (n.d.). *Challenges and benefits of data lakehouses*. Retrieved from <https://www.getrightdata.com/resources/private-chapter-5-data-lakehouse-challenges-and-benefits>

IBM. (2023). *What is a recommendation engine?* Retrieved from

<https://www.ibm.com/think/topics/recommendation-engine>

IT Convergence. (2024). *Best practices for data lakehouse security and governance*. Retrieved from <https://www.itconvergence.com/blog/best-practices-for-data-lakehouse-security-and-governance>

McKinsey & Company. (n.d.). *Personalization at scale: A proven value driver for growth*. Retrieved from <https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/personalization-at-scale-a-proven-value-driver-for-growth>