MEMORANDUM

TO: Federal Trade Commission

FROM: Irene Solaiman, Master in Public Policy 2019, Harvard Kennedy School
Funding provided by Microsoft Corporation
Note: The views and opinions expressed in this memorandum are those of the author

SUBJECT: Consumer Protection in AI/ML (FTC Hearing #7, Q: 9, 14, 15, 16, 21, 23, 25)

**Issue**: The FTC should create a comprehensive nationwide regulatory framework that identifies and protects consumers' personal data used in artificial intelligence and machine learning algorithms while promoting AI/ML development.

**Background**

Current privacy acts span specific sectors or states and can conflict or mislead companies[1]. Existing laws on data protection do not address modern concerns in AI/ML advancements. This necessitates a nationwide framework on data protection in the context of AI and ML. These algorithms consume large amounts of data to better spot patterns and progressively improve their output. Existing applications are already integrated into the average U.S. consumer's daily life from AI assistants like Alexa and Siri to targeted advertisements.

Existing regulations are met with resistance; GDPR, the EU's General Data Protection Regulation, in effect as of May 2018, has met claims of lack of clarity and stifled innovation. Its requiring companies to state how and where data is being used, explain the decision-making process, and enforcing high fines has deterred some firms from providing services to EU citizens[2]. Still, it provides a working framework for iterative updating.

A U.S. regulatory framework must have clear definitions, incentivize cooperation, and be iteratively updated with technological progress.

**Recommendations**

1. **Personal Data (Q9):** Clearly define and give guidelines on personal data.
2. **Explainability (Q14):** Prioritize process and dataset analysis over explainability.
3. **Data Openness and Deletion (Q14):** Encourage data-sharing and mandate minimization.
4. **Individual Consent (Q15):** Give consumers a spectrum of consent in data use.
5. **Educate Consumers (Q16):** Reach out to U.S. consumers on data hygiene and literacy.
6. **Development, not Innovation (Q21, 23):** Ensure improved research over new research.
7. **Enforcement and Compliance (Q25):** Vary fines by violation, applicable worldwide.
8. **Harmonize (Q23, 25):** Overlap regulatory themes with international regulations.

---

[1] O'Connor, N. (2018, January 30). Reforming the U.S. Approach to Data Protection and Privacy. Retrieved from https://www.cfr.org/report/reforming-us-approach-data-protection

[2] Meyer, D. (2018, May 25). AI Has a Big Privacy Problem And Europe's New Data Protection Law Is About to Expose It. Retrieved from http://fortune.com/2018/05/25/ai-machine-learning-privacy-gdpr/

*Recommendation One: Personal Data (Q9)*
In AI/ML, what constitutes personal data is not clear. In creating a regulatory framework, a clear definition is necessary for researchers, developers, and consumer-facing companies to determine what information to prioritize securing. It is also necessary for users to determine what information they should be most conscious of releasing.

Data that links back to the user is a broad definition that is subject to user opinion. For example, an individual's driving patterns used in a self-driving car may or may not be considered personal data. Instead, clearly defined sensitive information under personally identifiable information (PII) guidelines should be prioritized.

Pros:
- Clear definitions prevent loopholes in data labeling.

Cons:
- Regulatory definitions that contradict other regulations or in-company terminology may lead to confusion and difficulty complying with multiple regulations.

*Recommendation Two: Explainability (Q14)*
Most AI/ML algorithms meet the "black box" problem. They are not transparent and the methods by which they reach their ultimate decisions are difficult for even the algorithms' developers to explain. It is especially difficult to explain one data point's contribution to the overall decision made. Furthermore, machine explanations may not exist in human terms[3].

Explainable algorithms would provide little value for consumers who are not data-literate and be difficult to interpret for the data-literate. Certain methods for transparency could compromise accuracy[4]. Analysis should instead focus on the process by which the data is being used and possible biases within the datasets used.

Pros:
- Analyses on process and datasets used are more easily consumable and useful for consumers' informed consent decisions.
- Research will be not be hindered by the need for understandable algorithms, but be more conscious of what data is being used.

Cons:
- Depending on the system, explaining process may rely heavily on explaining the algorithm.

[3] Artificial Intelligence, Robotics, Privacy and Data Protection. (2016 October). Retrieved from https://edps.europa.eu/sites/edp/files/publication/16-10-19_marrakesh_ai_paper_en.pdf
[4] Wallace, N. & Castro, D. (2018, March 27). The Impact of the EU's New Data Protection Regulation on AI. Retrieved from http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf

*Recommendation Three: Data Openness and Deletion*
AI feeds on large datasets and functions best in a data-friendly ecosystem; according to the McKinsey Global Institute, nations that promote data sharing and open data sources are more likely to see advances in AI[5]. Anonymizing datasets is often used but is an ineffective way to protect consumers. Re-identification methods popularly exist and are evolving[6]. Instead, data should be minimized and deleted after use in ways that do not inhibit the AI/ML system. Companies must also provide proof of deletion.

Pros:
- Companies must collect less personal data to comply with minimization.
- Data deleted is not easily recoverable or accessible to malicious actors.

Cons:
- Data is difficult to permanently "erase."
- Deletion has been shown to reduce accuracy in or break certain systems where multiple data points are erased. Cumulative data point erasure could change important rules in an algorithm[7], rendering it less effective for future use.

*Recommendation Four: Individual Consent (Q15)*
Consumers are currently only given a binary choice for consent: waive privacy rights or deny data use. Users must be given the option to have a varying degree of consent. Consent should also be explicitly defined for consumer-facing companies that provide advising services like purchase recommendations on e-commerce sites. When submitting personal information that does not fall under PII guidelines, users should also be able to consent to its use.

Pros:
- Consumers have more mobility with how their data is being used and for what services.

Cons:
- Informed consent requires some consumer understanding and likely data literacy, which may advantage and disadvantage certain individuals.

*Recommendation Five: Educate Consumers (Q16)*
Lack of transparency in AI/ML challenges the ability of consumers to make informed decisions. This requires educating consumers through formative core curricula for younger consumers, and mandating consumer-facing companies to brief consumers on their rights and data use before product use.

Pros:
- Consumers will be more likely to understand technical terms and personal impact.

Cons:
- Education outreach will still disadvantage certain demographics, especially overseas citizens.

---

[5] West, D. M., & Allen, J. R. (2018, May 09). How artificial intelligence is transforming the world. Retrieved from https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/#_edn1
[6] Berinato, S. (2015, July 24). There's No Such Thing as Anonymous Data. Retrieved from https://hbr.org/2015/02/theres-no-such-thing-as-anonymous-data
[7] Wallace, N. & Castro, D. (2018, March 27). The Impact of the EU's New Data Protection Regulation on AI. Retrieved from http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf

*Recommendation Six: Development, not Innovation (Q21, 23)*
Regulation should first strive to improve AI/ML research and incorporate ethical standards and engineering principles like privacy by design and privacy by default. Regulation must delineate methods to incorporate these rights while systems are in development phases as opposed to adding security once built.

Pros:
- This ensures prioritizing improving existing research before new developments.
- It sets the stage for future development to incorporate privacy needs.

Cons:
- Building in principles is difficult for existing systems, especially if they already have a large userbase (i.e. Siri, Alexa).

*Recommendation Seven: Enforcement and Compliance (Q25)*
The FTC must determine firms responsible for violations and fine accordingly. Any data collected must first be labeled as PII or otherwise and assigned the correlating level of protection. Fines should vary by violation and level of harm to consumers. This must be non-discriminatory to each firm and establish consumer privacy as a right[8].

Pros:
- Companies will be incentivized to comply with regulations.
- Compliance reestablishes trust in consumer-facing firms and portray the high value of data to consumers; data costs dollars.

Cons:
- Foreign firms that perceive stringent rules and high fines may be disincentivized from offering services to U.S. citizens.
- Level of harm to consumers may be subjective and requires clear guidelines.

*Recommendation Eight: Harmonize (Q23, 25)*
Drawing overarching themes from modern existing data protection regulation prevents patchwork regulation, ensures international compliance, and promotes equal and fair competition.

Pros:
- Companies with an international presence will be less have less incentive to cut out the U.S. market and further incentive to standardize their approach to personal data use.

Cons:
- Foreign regulations are not fitted to the U.S. market; themes must be extracted by applicability.

[8] How Will California's Consumer Privacy Law Impact The Data Privacy Landscape? (2018, August 20). Retrieved from https://www.forbes.com/sites/forbestechcouncil/2018/08/20/how-will-californias-consumer-privacy-law-impact-the-data-privacy-landscape/#193a772fe922