TECHNOLOGY BRIEFING ON AI IN DIGITAL PRIVACY & LIABILITY

In today's digital age, Artificial Intelligence (AI) stands at the forefront of innovation. However, its reliance on vast data sets poses significant challenges for digital privacy, which has emerged as a chief public concern as the technology continues to proliferate. *AI models learn from vast amounts of data*, and the larger and more diverse the data, the more accurate and versatile the AI (e.g., ChatGPT). However, this *data is often collected or purchased from the public domain or private networks without explicit user consent or awareness*.

Data Privacy Basics

The *Cambridge Analytica Scandal (2018)* utilized millions of Facebook profiles without consent to build a software program that could influence voters in elections. This incident underscores the consequences of unregulated data access and its potential for misuse. Both sides of the aisle have championed and expressed the need for stronger regulations, from concerns over wiretapping to debates on internet privacy. See the complementary AI & digital privacy policy memo for more specifics.

- ***Device/Web Architecture & Privacy Leaks***
    - **Devices**: These include smartphones, tablets, computers, and even IoT (Internet of Things) devices like smart fridges or thermostats. They often run apps or browsers that connect to the internet.
    - **Browsers**: Applications like Chrome, Firefox, and Safari that allow users to access websites. Browsers can store cookies, which are small pieces of data that keep track of user preferences or sessions.
    - **Servers**: Physical or virtual machines that host websites, services, or databases. They process requests from devices, like serving up a webpage or storing data from an app.
- ***Data Flow and Monitoring***
    - When a user searches on Google or buys a product on Amazon, the device sends a request to the respective server. The server processes the request and sends back the relevant data or acknowledgment.
    - During this process, information about the user's behavior, device details, location, and more can be logged by the server.
- ***Trackers and Cookies***
    - **Trackers**: These are scripts or pieces of code embedded in websites. They can monitor user behavior, such as which pages you visited, how long you stayed, where you clicked, and more.
    - **Cookies**: Small files saved on the user's device by websites. They can remember login details and user preferences, or track user behavior across different sessions or even different websites.
    - **Third-party Cookies/Trackers**: These are set by domains other than the one the user is currently visiting. They allow tracking across multiple websites, building a broader profile of the user's habits.
- ***Data Monetization and Advertising***
    - **Data Brokers**: Companies that specialize in collecting, processing, and selling user data. They might gather data from various sources, aggregate it, and then sell it to advertisers or other companies.

    - **Targeted Advertising**: Using the data collected by trackers and data brokers, advertisers can display ads specifically tailored to a user's interests, behaviors, and demographics.
    - **Surveillance Capitalism**: This term refers to the business model where companies profit from the extensive collection, processing, and sale of user data without necessarily providing a direct service to the user. It's the idea of capturing free raw material (user data) and translating it into behavioral predictions that are then sold to markets that trade in these future behaviors.

Unique AI Privacy Issues

- ***Data Collection for AI***
    - **Training Data Necessity**: *Machine Learning*, a subset of AI, operates by learning from data. The more data it has, the better it can identify patterns and make predictions. This is similar to how having more examples in a textbook can provide a student with a better understanding of a subject.
    - **Sources**: AI models can be trained on various data types, including text, images, videos, and more. For instance, a voice recognition AI might be trained on thousands of hours of spoken language from different demographics to understand accents, tones, and dialects.
    - **Over-Collection**: While some data collection is essential for functionality, many applications gather additional data that isn't strictly necessary for their primary function. This might be for future features, monetization through data sales, or other less transparent reasons.
- ***Data Storage***
    - **Cloud Servers**: These are virtual servers often distributed across multiple physical locations. They allow for easy access, scalability, and redundancy, which means if one server fails, the data isn't lost.
    - **Encryption at Rest**: Even though data on these servers is often encrypted (converted into a code to prevent unauthorized access), decryption keys can also be vulnerable. If these keys are accessed by malicious actors, they can decode the data.
    - **Data Breaches**: Incidents where unauthorized individuals gain access to confidential data. This could be due to software vulnerabilities, inadequate security protocols, or human errors. Successful breaches can expose sensitive user data, with potential real-world consequences for the affected individuals.
- ***Biases in AI***
    - **Origin of Bias**: Data used to train AI is often collected from real-world sources which may contain inherent biases. If an AI is trained on biased data, it learns those biases.
    - **Training Set Limitations**: Suppose a facial recognition AI is primarily trained on images of individuals from a particular demographic. In that case, it will be more accurate for that demographic and less accurate for others, leading to misidentification or non-recognition issues.
    - **Feedback Loops**: If biased AI systems are deployed in the real world and their outputs are used as new data for further training, they can reinforce and exacerbate their inherent biases. For instance, if an AI system used in law enforcement wrongfully identifies a particular demographic as being more criminal, and this data is fed back into the system, the bias gets strengthened.

Technologies Preserving Data Privacy

While data, especially on social media platforms and account-based systems, remains vulnerable throughout the development process, several technological solutions can significantly mitigate these risks. Beyond merely offering users opt-in choices, *proactive technical measures can prevent data leaks and curtail unethical data monetization practices*. These technologies have received increased attention and investment over the years—and are especially needed when sensitive data (like patient health records or intellectual property) are involved—but lack enforceability due to their absence in regulatory frameworks and need broader industry adoption.

- ***Federated Learning***
  - In traditional machine learning, raw data from all sources is pooled into one location for training. Federated learning, however, trains AI models at the source (e.g., user devices) and only combines the updated model parameters rather than raw data.
  - **Benefits**: Protects user privacy by not needing to upload raw data, offers real-time model improvements, and can potentially save bandwidth.
  - **Examples**: Tech giants Google and Apple use this technology to create prediction algorithms, such as autocorrect and text recommendations, without saving user data (such as an individual's keystrokes). Companies like *OpenMined* are providing tools to make federated learning more accessible to developers, while startups like *Owkin* use *federated learning in healthcare*, allowing hospitals to collaborate without sharing sensitive patient data.
- ***Differential Privacy***
  - A system ensures differential privacy if the probability of a specific output does not significantly change regardless of the participation of any one individual's data.
  - **Implementation**: Typically, random "*noise*" (nonsense, obfuscatory information) is added to the data or results to mask individual entries.
  - **Examples**: The *U.S. Census Bureau uses differential privacy* to safeguard respondent's data. *Databricks*, a leading data and AI company, also offers solutions that can integrate differential privacy into data analytics workflows.
- ***Homomorphic Encryption***
  - Working Principle: This encryption method allows computation on ciphertexts, generating an encrypted result that matches the outcome of the operations as if they had been performed on plaintext.
  - **Use Cases**: It's especially useful in cloud computing, where a client can encrypt data, send it to a server for complex computation, and get the encrypted result back without the server ever seeing the raw or interim data. IBM is enabling clients to experiment with this technology to protect their IT architecture, products, and data. Startups like Enveil are working on commercial applications of homomorphic encryption, including secure search and analytics on encrypted data.
- ***On-device AI & Local Computation***
  - Instead of relying on cloud servers, the AI model operates directly on the device. It means processing data and drawing conclusions locally. This system is mostly used when the computation being performed is relatively streamlined and simple.
  - **Benefits**: Faster response times (no need to communicate with a central server) and enhanced privacy since data remains on the device.
  - **Examples**: Commonly accessed authentication tools like Apple's *FaceID* and the *'Snips'* platform (acquired by Sonos) that offer an on-device voice assistant, ensuring that voice commands are processed locally without being sent to the cloud. Qualcomm, a major chip manufacturer, is also developing specialized hardware to facilitate on-device AI for mobiles and other gadgets.

The rise of AI amplifies concerns about digital privacy, given its ability to deeply analyze and predict user behavior. As AI's reach grows, it's imperative to implement strong privacy controls to prevent misuse and data breaches. Ensuring that AI upholds individual privacy rights is crucial for harnessing its benefits responsibly.