Lecture Title and Date

Privacy, Security, Law and Science - Dov Greenbaum Mar 24.

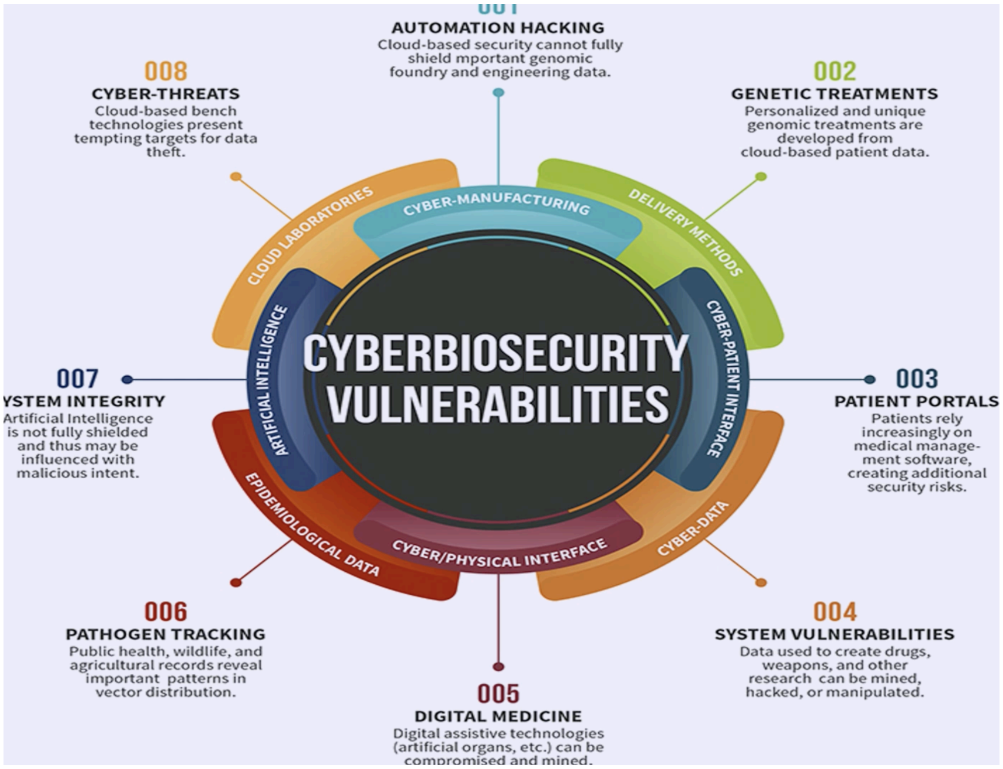
Objectives of the Lecture

1. Understand key concepts of data security in the bioeconomy
2. Recognize privacy concerns related to personal data, especially genetic information
3. Explain the concept of synthetic data and its limitations
4. Identify legal frameworks governing data privacy
5. Understand ethical questions around data ownership, particularly genetic data

Key Concepts and Definitions

Bioeconomy: Economic sectors that employ biological resources or have health outputs, including drug manufacturing, food/agriculture development, energy, and more. Estimated to be worth approximately $4 trillion currently and potentially $30 trillion by 2030 according to the World Bioeconomy Forum.

Cyber Biosecurity: Field focused on assessing vulnerabilities, developing countermeasures, and promoting policies to protect biological data. Biological data in the bioeconomy includes genetic, proteomic, microbiomic, clinical, pharmacogenomic, behavioral and lifestyle, environmental exposure, phenotypic, biobank, and aggregated data.


001
AUTOMATION HACKING
Cloud-based security cannot fully shield mportant genomic foundry and engineering data.
008
CYBER-THREATS
Cloud-based bench technologies present tempting targets for data theft.
002
GENETIC TREATMENTS
Personalized and unique genomic treatments are developed from cloud-based patient data.
CYBER-MANUFACTURING
CLOUD LABORATORIES
DELIVERY METHODS
ARTIFICIAL INTELLIGENCE
CYBER-PATIENT INTERFACE
CYBERBIOSECURITY VULNERABILITIES
007
YSTEM INTEGRITY
Artificial Intelligence is not fully shielded and thus may be influenced with malicious intent.
003
PATIENT PORTALS
Patients rely increasingly on medical management software, creating additional security risks.
EPIDEMIOLOGICAL DATA
CYBER/PHYSICAL INTERFACE
CYBER-DATA
006
PATHOGEN TRACKING
Public health, wildlife, and agricultural records reveal important patterns in vector distribution.
004
SYSTEM VULNERABILITIES
Data used to create drugs, weapons, and other research can be mined, hacked, or manipulated.
005
DIGITAL MEDICINE
Digital assistive technologies (artificial organs, etc.) can be compromised and mined.

Bioconvergence: The need to incorporate more AI into more aspects of biology.

Synthetic Data: Artificially generated data intended to mask underlying real data while preserving statistical properties, making it harder to perform linking attacks.

Digital Twins: Digital representations of physical objects (or potentially people) used for testing and simulation, initially developed by NASA.

Privacy: The right to decide how the world should view you and what information about you is disclosed. Various interpretations include confidentiality, anonymity, and control over personal information. An invasion of privacy is unauthorized, unwanted, unwarranted, or illegal access to personal information. It also includes public disclosure of information and surveillance. Privacy is important due to personal, social, emotional, psychological, reputational, and economical (such as health/life insurance pricing) consequences that come from invasion of privacy.

Open Source Intelligence (OSINT): Using publicly available information to deduce private information about individuals.

Big Data: Characterized by the "five V's" - volume, velocity, variety, veracity, and value. Often collected from everyday digital interactions.

Main Content/Topics

Morality is all relative: morality in the field of science and technology is all relative and changing over time. For example, IVF was considered unfathomable when it was first developed, but ethically acceptable today.

Security in the Bioeconomy: The bioeconomy encompasses sectors using biological resources including healthcare, pharmaceuticals, and agriculture. Security concerns arise with various types of data: genetic/omic, clinical data, agricultural data, social media-health related data. Databases are only as protected as their weakest link, e.g the Golden State Killer who was identified through genetic genealogy services which allowed investigators to find relatives through DNA databases.

Synthetic Data and Model Collapse: Synthetic data was discussed as a potential solution to data security concerns, as it doesn't directly represent real individuals. However, the lecturer noted potential problems with synthetic data, including: loss of information at the "edges" or rare cases, potential "model collapse" when AI models are trained on synthetic data that was itself generated by AI, leading to progressive degradation of data quality.

Digital Twins: Digital twins are virtual models that simulate real-world objects or systems. Originally developed by NASA for testing spacecraft components (like engines), this concept could theoretically be applied to humans. This would require extensive data collection and centralization, creating additional security concerns.

Brain Data Privacy (cyberneurosecurity): An emerging concern is the collection of brainwave data through medical devices, research studies, and consumer technology (such as Smart Cap with chips implanted). Legal regulations surrounding “freedom of thought” remain ambiguous today.

Genetic Privacy and Ownership: The two primary violations of privacy in genetics are identification and description. One example of identifying genetic information is those used in the police CODA system that’s solely used to identify individuals, and do not contain any phenotypic information. The lecture concluded with a discussion of genetic privacy using 23andMe as a case study:

- The company recently filed for bankruptcy. Some possible reasons include an unsustainable business model (each consumer will probably only ever use it once) and loss of FDA approval for predicting risks of health conditions.
- Previously, they had sold user genetic data to GlaxoSmithKline for $300 million.
- The company also experienced a significant hack leading to a $30 million settlement.
- As a result, the company’s worth fell from $6 billion to nearly zero.
- The lecturer raised the question of whether individuals own their genetic information.

Courts have historically ruled that once biological information is donated to institutions, ownership rights are lost. The lecturer questioned whether contracts that transfer ownership of genetic information should be considered "unconscionable" given the deeply personal nature of genetic data. Moreover, genetic information is inherently shared among biological relatives, potentially compromising the privacy of individuals who have not provided consent. Determining how, and whether, such shared genetic privacy should be protected remains an ethical and legal question.

Discussion/Comments

The lecture covered a wide range of topics related to data security and privacy, particularly focusing on biological and genetic data. The rapid pace and breadth of topics made it challenging to explore any single topic in great depth. The most compelling discussions centered around:

1. The ethical questions of genetic data ownership - who owns your DNA once you've shared it with a company like 23andMe?
2. How seemingly unrelated public information can be combined to reveal private information about individuals
3. The evolution of privacy as a concept and right over time

The topic of digital twins was particularly interesting as it showed how technology originally designed for mechanical testing could potentially be applied to human biological systems, raising complex ethical questions.

List all suggested reading here and please answer:

Are the readings for the class useful? If so, are the specific subsections useful or would change. If not, are there other references you could suggest? Please suggest one.

- The European Union's General Data Protection Regulation (GDPR) documentation
- Reports from the World Economic Forum on the bioeconomy
- Articles on the legal cases surrounding biological material ownership

These readings were useful to get more context about the legal regulations surrounding data privacy. As the lecture covered a lot of different topics, the readings were great supplementary materials to gain more depth into the topics.

Are the readings for the class useful? If no specific readings were mentioned, I would suggest adding:

- Solove, D. J. (2008). "Understanding Privacy" - Harvard University Press
- O'Neil, C. (2016). "Weapons of Math Destruction" - For sections on algorithmic bias and privacy

Other Suggest references for many of the key concepts

- National Human Genome Research Institute. "Privacy in Genomics." *National Institutes of Health*, https://www.genome.gov/about-genomics/policy-issues/Privacy. - Overview of genetics privacy
- Glassman, Michael, and Min Ju Kang. "Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT)." Computers in Human Behavior 28.2 (2012): 673-682. - Discussion on OSINT that the lecturer briefly mentioned
- Kamel Boulos, Maged N., and Peng Zhang. "Digital twins: from personalised medicine to precision public health." Journal of personalized medicine 11.8 (2021): 745. - Discussion on digital twins in medical applications
- Ienca, Marcello, and Pim Haselager. "Hacking the brain: brain–computer interfacing technology and the ethics of neurosecurity." Ethics and information technology 18 (2016): 117-129. - Discussion on neurosecurity