RSAConference2022

San Francisco & Digital | June 6 – 9

TRANSFORM

SESSION ID: **LAW-T09**

Designing a Data Ethics and Privacy Program for AI/ML

Bret Cohen

Partner, Privacy and Cybersecurity Practice
Hogan Lovells
bret.cohen@hoganlovells.com

Alexandra Ross

Senior Director, Senior Data Protection, Use & Ethics Counsel
Autodesk
alexandra.ross@autodesk.com

Ilana Golbin

Director, Emerging Technologies & AI
PwC
Ilana.a.golbin@pwc.com

Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA® Conference, RSA Security LLC or any other co-sponsors. RSA Conference does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

What is “data ethics”?

“A branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use”

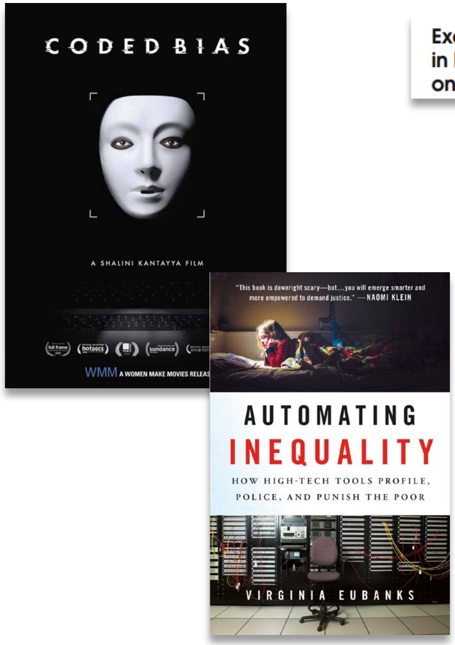
- Open Data Institute

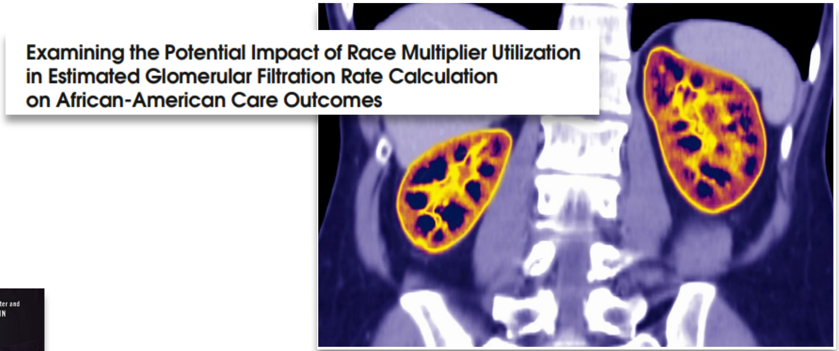
“Data Ethics are the norms of behavior that promote appropriate judgments and accountability when acquiring, managing, or using data, with the goals of protecting civil liberties, minimizing risks to individuals and society, and maximizing the public good.

- US Federal Data Strategy

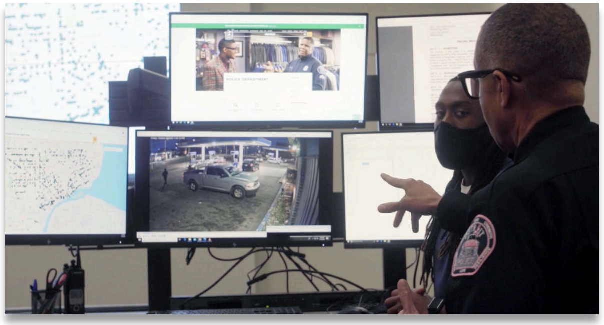
What are the risks?

ofqual


CODED BIAS
A SHALINI KANTAYYA FILM
WMM A WOMEN MAKE MOVIES RELEASE
"This book is downright scary—but....you will emerge smarter and more empowered to demand justice." —NAOMI KLEIN
AUTOMATING INEQUALITY
HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR
VIRGINIA EUBANKS



Examining the Potential Impact of Race Multiplier Utilization in Estimated Glomerular Filtration Rate Calculation on African-American Care Outcomes



Global regulation of automated decision-making

- GDPR Art. 22: "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."
  - All automated processing to incorporate **data protection by design** principles (Art. 25)
  - **Notice** to data subject (Art 13 & 14), including meaningful information about the **logic involved** and the **significance and envisaged consequences** for the data subject
  - **Data protection impact assessments** required for automated decision-making based on a "systematic and extensive evaluation of personal aspects"
- Similar rules in Brazil (LGPD), South Africa (POPIA), and China (PIPL)

ADM laws: Coming to a state near you

California (CPRA)

- CPRA rulemaking to produce regulations "governing **access** and **opt-out rights** with respect to business' use of automated decision-making technology, including **profiling** and requiring businesses' response to access requests to include **meaningful information** about the logic involved in such decision-making processes, as well as a **description of the likely outcome** of the process with respect to the consumer"
- "Profiling" means "any form of automated processing of personal information [as defined in rulemaking] to **evaluate** certain personal aspects relating to a natural person, and in particular to **analyze or predict** aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements"

Colorado (CPA) and Virginia (VCDPA)

- Right to **opt-out** for **profiling in furtherance of decisions that produce legal or similarly significant effects** concerning the consumer.
- "Profiling" means any form of automated processing of personal data to **evaluate, analyze, or predict** personal aspects concerning an identified or identifiable individual's economic situation, health, personal preference, interest's reliability, behavior, location, or movement."
- **Data Protection Assessments** for certain high risk profiling activities, including profiling if it presents a **reasonably foreseeable risk** of unfair or deceptive treatment, unlawful disparate impact, financial or physical injury, or other substantial injury to consumers.

Current US regulation of algorithms

- **FTC guidance:**
  - Sale or use of racially biased algorithms may be "unfair"
  - Recent guidance focuses on transparency, unexaggerated claims, and vetting algorithms before launch
  - Watch for claims of data deletion, ability to opt in/out

- **Civil rights/anti-discrimination**
  - Fair Credit Reporting Act (FCRA)
  - Equal Credit Opportunities Act (ECOA)
  - Proposed federal legislation focused on discrimination, e.g.:
    - Algorithmic Justice and Online Platform Transparency Act of 2021 (S.1896 / H.R. 3611)
    - Protecting Americans from Dangerous Algorithms Act (H.R. 2154 / S.230)
    - Justice Against Malicious Algorithms Act
  - AI bills/resolutions introduced in over 20 states, many focused on anti-discrimination principles, notice, risk assessment & reporting

New York credit bias probe

- Following a series of consumer complaints around gender discrimination, lack of transparency, and the use of “black box algorithms,” NY DFS investigated Goldman Sachs’ credit assessment decisions around their Apple Card product
- NYDFS found no evidence of disparate treatment or impact in credit decisions, which were “explainable, lawful, and consistent with the Bank’s credit policy”
- Factors not indicative of bias:
  - Lack of consumer-facing transparency into algorithms that determine creditworthiness
  - Different outcomes for spouses
- Factors that *could* have led to finding of unlawful bias:
  - Policies providing for lower credit limits for women/protected classes
  - Evidence suggesting Goldman judged men and women by different standards
  - Lack of a fair lending program/considering prohibited characteristics of applicants

Pending EU AI regulation

- Focus on data governance, requiring companies to analyze datasets that are used in the training, validation & testing of machine learning, including identifying potential biases, checking for inaccuracies and assessing suitability of the data

- Obligations vary depending on risk of AI system:

Unacceptable Risk	• Manipulation leading to psychological/physical harms or on age/disability vulnerability • Decontextualized or unjustifiably detrimental social scoring system • Indiscriminate facial recognition by police	Prohibition
Higher Risk	• Credit scoring; assessment of workers, students • AI used in critical infrastructure; by judges, public administration, police, border control • Products subject to safety regulations under EU law	Design obligations (human oversight, data management plan, risk assessment, etc.)
More Limited Risk	• Deep fake • Emotion recognition • AI interacting with humans	Transparency obligations
Minimal Risk	• Residual	Voluntary codes of conduct

EU Digital Services Act

- Large online platforms should be accountable, through independent auditing, for compliance with the DSA
  - Auditors can make use of objective sources of information, and should guarantee confidentiality, security and integrity of the platform's information
  - May require access to or reporting of specific data, including:
    - Data necessary to assess risks and possible harms brought about by the platform's systems
    - Data on the accuracy, functioning and testing of algorithmic systems for content moderation, recommender systems, or advertising systems
    - Data on processes & outputs of content moderation or internal complaint-handling systems
- Empowered to require access to, and explanations relating to, data-bases and algorithms of relevant persons, and to interview, with their consent, any persons who may be in possession of useful information and to record the statements made

Common threads across laws

- Increased focused on risks associated with algorithms globally, including new laws and prioritized enforcement related to automated decisions and profiling
- Application may vary depending on harm/impact
  - Rules applying to “decisions that produce legal or similarly significant effects”
  - Differing level of diligence depending on potential harm
- Privacy impact assessments and consumer rights to opt-out or limit decisions, especially for higher-risk applications
- Discriminatory or disparate impact likely to trigger additional regulatory scrutiny
- Laws may require transparency/consent, even for “minimal risk” applications

Ethics and your data protection program

Privacy & Ethics

Both are contextual and about people

In GDPR - Article 22, DPIAs, Legitimate Interest Analysis. Also in US - CCPA/CPRA, VA, CO

Integrating Ethics

Partnership with the ethics team, or

Incorporating ethics into your privacy & security programs

Service Provider Role

Help guide customers

Consider ethical impact of products and build-in ethics by design

Governance harmonization

- Goal: Add governance without additional tax/ overhead/ admin burden
- Solution: Harmonize different types of governance impacting data use:
  - Agree on the target of governance (is it the project? Is it the dataset?)
  - Standardize language (data vs. dataset, what is ML, personal data not PII, etc.)
  - Establish a single technical home (platform) for governance artifacts, process, repository (“let’s all build here”); eventually, create a shared back-end database
  - Make intake processes cross-reference one another (“are you using personal data?” Y “please complete a PIA”); eventually conform multiple intake processes and associated documentation
  - Create shared evaluation processes and escalation paths

Establishing data ethics

Observation and learning

Data Ethics Framework for Assessment only

Establish best practices

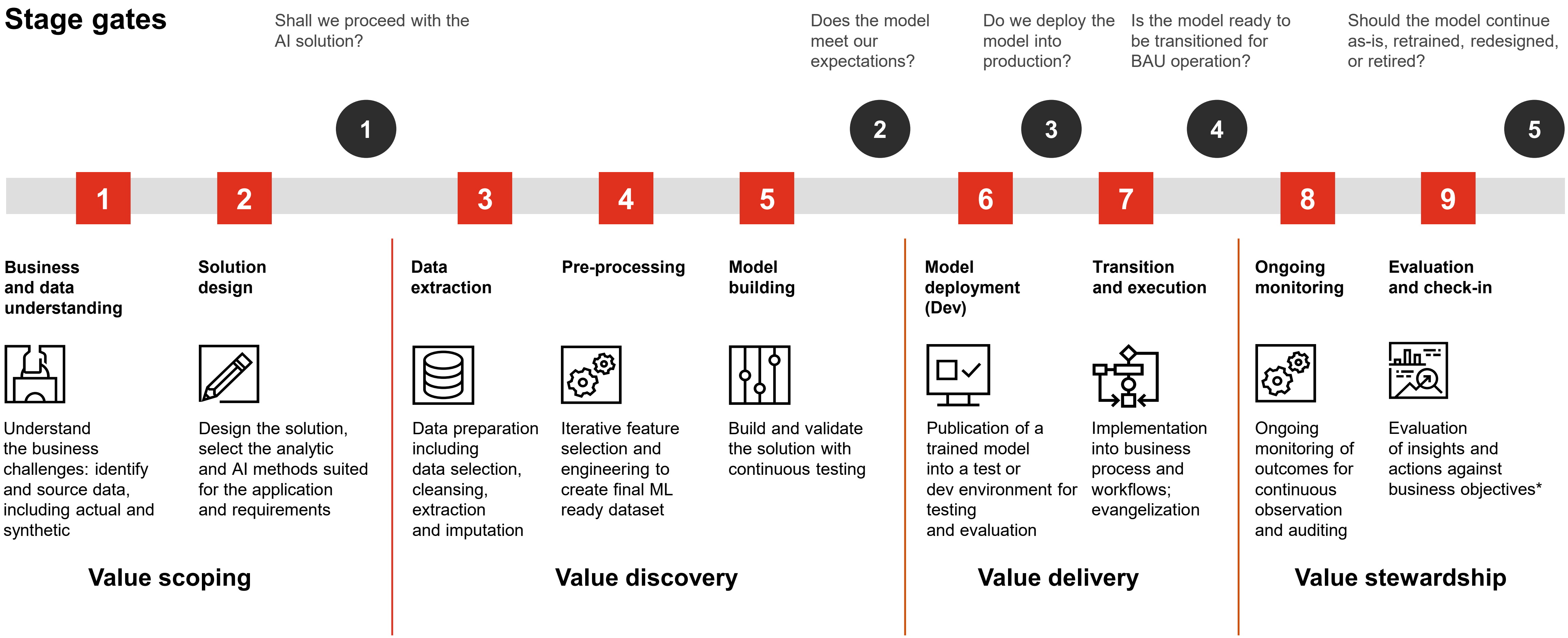
Data Ethics Framework plus **Ethics by Design** to drive behavior

Govern

Data Ethics Framework and Ethics by Design plus ***ex-post* review**

Culture | Communications | Strategy

Privacy and security need to be designed into each step of the AI system development process



Stage gates

1 Shall we proceed with the AI solution?

2 Does the model meet our expectations?

3 Do we deploy the model into production?

4 Is the model ready to be transitioned for BAU operation?

5 Should the model continue as-is, retrained, redesigned, or retired?

Value scoping

1 **Business and data understanding**

Understand the business challenges: identify and source data, including actual and synthetic

2 **Solution design**

Design the solution, select the analytic and AI methods suited for the application and requirements

Value discovery

3 **Data extraction**

Data preparation including data selection, cleansing, extraction and imputation

4 **Pre-processing**

Iterative feature selection and engineering to create final ML ready dataset

5 **Model building**

Build and validate the solution with continuous testing

Value delivery

6 **Model deployment (Dev)**

Publication of a trained model into a test or dev environment for testing and evaluation

7 **Transition and execution**

Implementation into business process and workflows; evangelization

Value stewardship

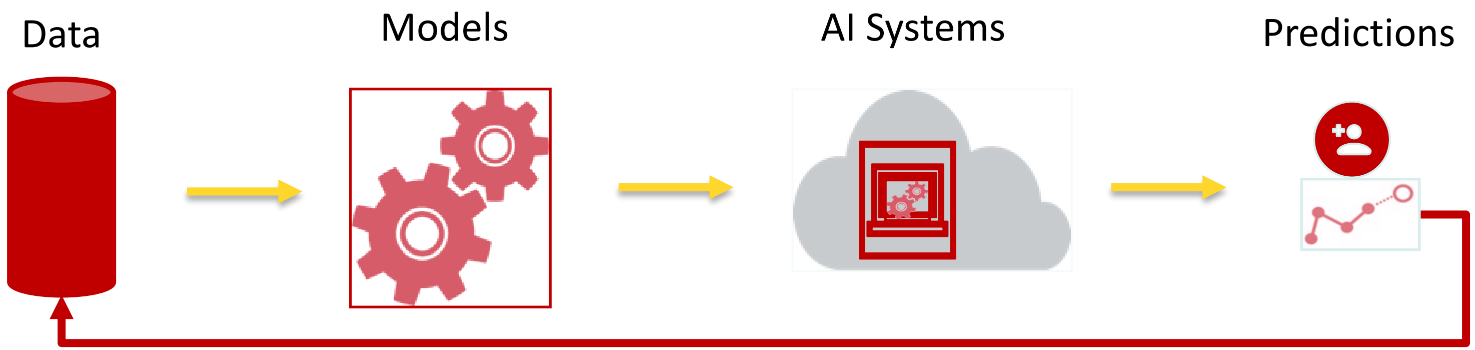
8 **Ongoing monitoring**

Ongoing monitoring of outcomes for continuous observation and auditing

9 **Evaluation and check-in**

Evaluation of insights and actions against business objectives*

Privacy and security concerns exist throughout the AI/ML development process



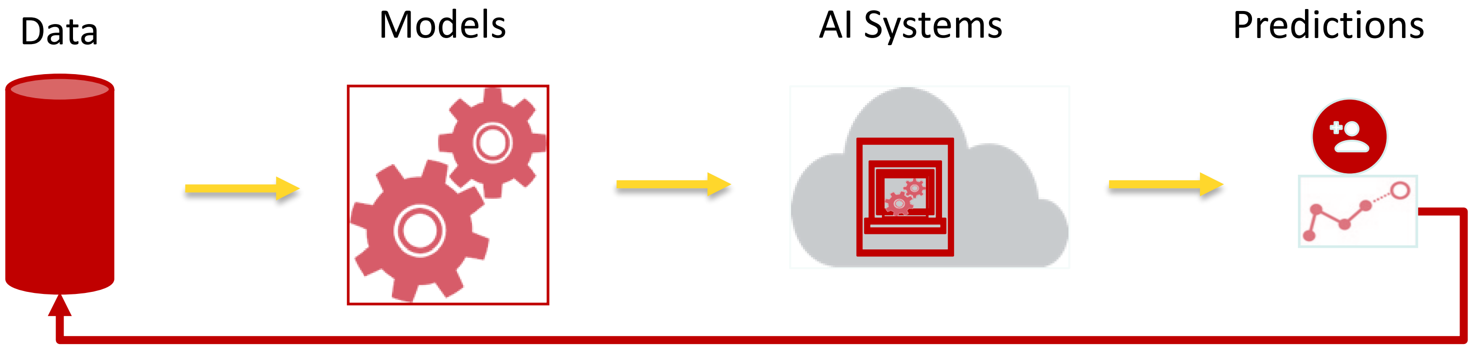
Privacy & confidentiality of training data

Privacy-preserving algorithms and personalized services; Train against robust and variable data

Minimize vulnerability to unauthorized access to output and computation; Anticipate adversarial attacks

Identity-preserving access;
Privacy-preserving inference;
Identify attacks and subversions;
Identify model theft

Privacy and security by design in AI require addressing issues across different vulnerabilities


Data

Models

AI Systems

Predictions

30%

of all AI cyberattacks will leverage training-data poisoning, AI model theft, or adversarial samples to attack AI-powered systems through 2022

- Gartner Research

Data at Rest Vulnerabilities
Data authenticity
Data integrity
Data confidentiality
Data privacy
Data utility
Data quality
Data bias
Data use

Computation Vulnerabilities		
Computation authenticity	Computational integrity	Adversarial attacks
Computational confidentiality	Computational privacy	

Privacy-Enhancing Technologies Vulnerabilities
Verification layer
Storage layer
Communication layer
Processing layer

Identity Vulnerabilities
Identity authenticity
Identity integrity
Identity confidentiality
Identity privacy

Identity-Enhancing Technologies Vulnerabilities
Verification layer
Consensus layer

Examples of AI attacks

Data Poisoning

Targeted attacks contaminate the machine model generated in the training phase. Examples are Label contamination attack, Impersonate attack, Evasion attack. Indiscriminate attacks involve training on unknowingly compromised data

Adversarial Perturbation

Attacker stealthily modifies the query to get a desired response which affects model's classification performance or randomly injects noise to misclassify data. He may also Craft inputs to reduce the confidence level of correct classification

Model Inversion, Model stealing and Neural net reprogramming

In model inversion, private features used in machine learning models can be recovered. In model stealing the underlying model is recreated by legitimately querying the model. In Neural net reprogramming machine learning systems are reprogrammed

Adversarial example in Physical domain, Backdoor ML

In physical domain attack, machine learning system in misled in the last layer and Backdoor ML is using outhouse trojaned model which forces targeted mis-classifications

Model safety

There might be unintentional safety concerns that occur due to not only unforeseen conditions but also due to environmental changes and training insufficiencies

Membership inference

Attacker can determine whether a given data record was part of the model's training dataset or not. This can be used to tamper with model performance

Recover Training data

Adversary might recover training data used by model by using queries etc. that defeats fails protection system and used for other illegal activities like ransomware

There are several main privacy techniques – but their usage and maturity vary greatly

Differential Privacy

Ensures that anyone using any database for learning will use an approximate version of that database.

Federated Learning

Trains an algorithm across multiple decentralized edge devices or servers using local data samples, without exchanging them.

Homomorphic Encryption

Permits users to perform computations on its encrypted data without first decrypting it.

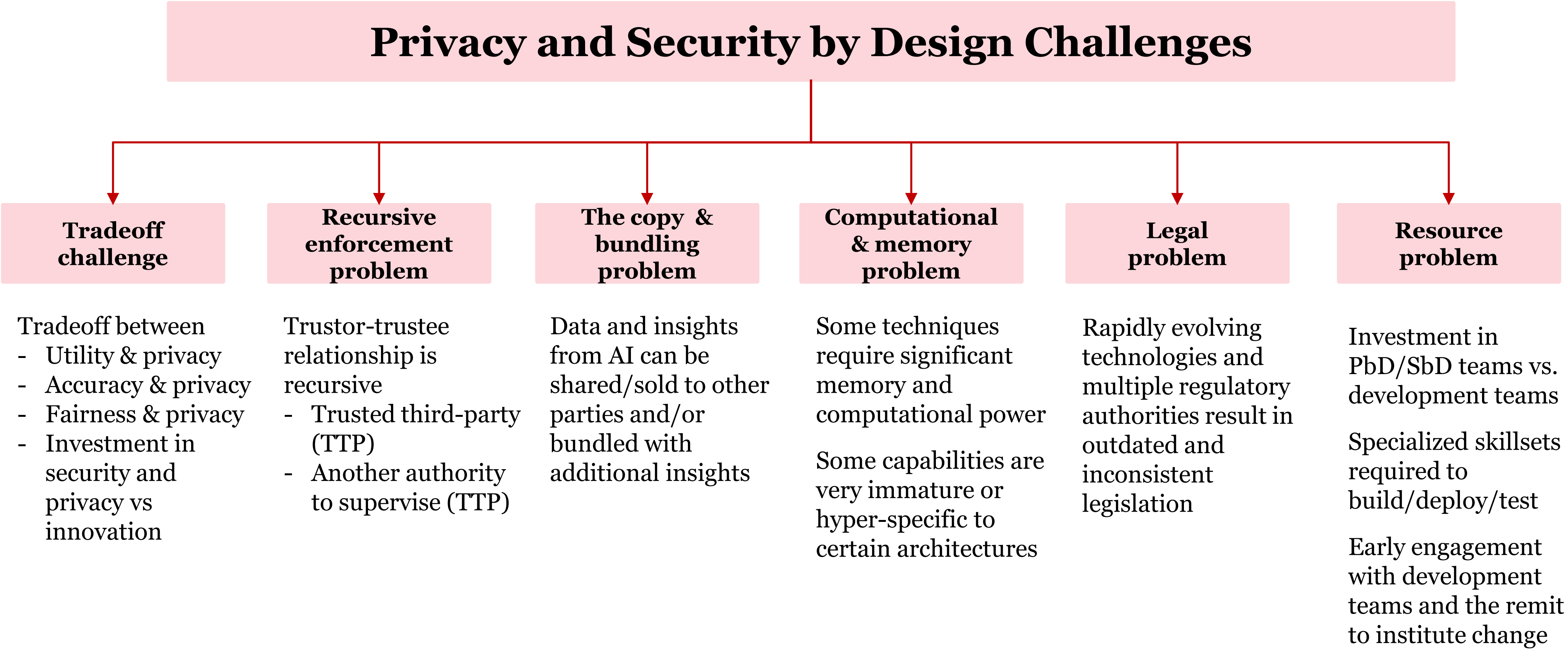
Secure Multi Party Computation

Splits data and assigns the data to multiple trusted third parties so that computation can be done on the split data across third parties without sharing data between each other.

Synthetic Data Generation

Creates statistically similar data that can preserve sensitive data and can be used in ML models when there is a lack of data.

Privacy and security by design still face significant legal and technical challenges



Privacy and Security by Design Challenges

Tradeoff challenge

Tradeoff between
- Utility & privacy
- Accuracy & privacy
- Fairness & privacy
- Investment in security and privacy vs innovation

Recursive enforcement problem

Trustor-trustee relationship is recursive
- Trusted third-party (TTP)
- Another authority to supervise (TTP)

The copy & bundling problem

Data and insights from AI can be shared/sold to other parties and/or bundled with additional insights

Computational & memory problem

Some techniques require significant memory and computational power

Some capabilities are very immature or hyper-specific to certain architectures

Legal problem

Rapidly evolving technologies and multiple regulatory authorities result in outdated and inconsistent legislation

Resource problem

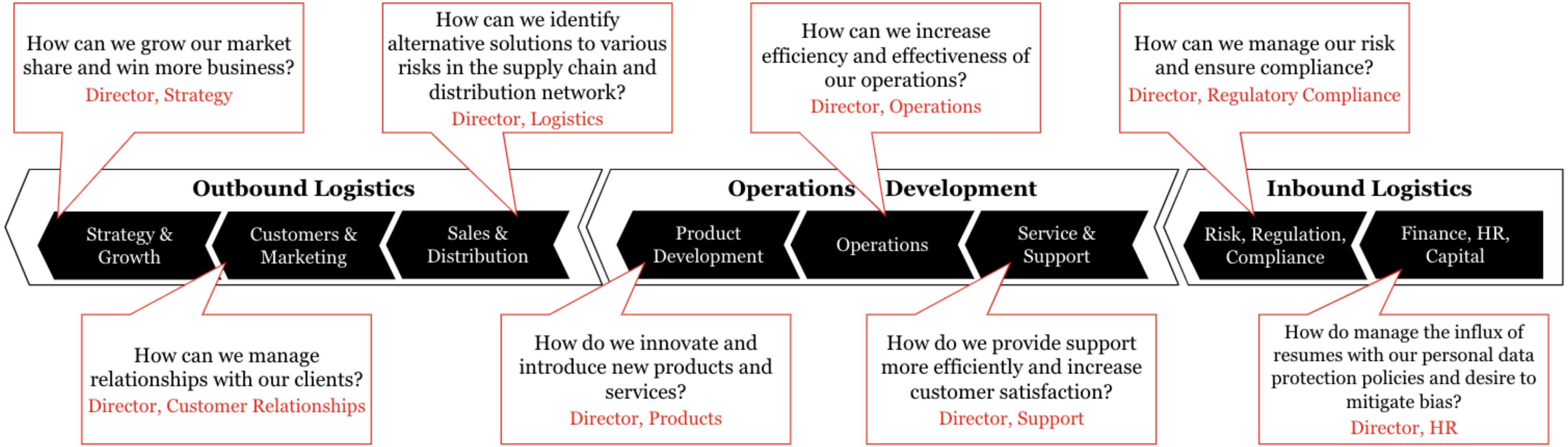
Investment in PbD/SbD teams vs. development teams

Specialized skillsets required to build/deploy/test

Early engagement with development teams and the remit to institute change

Source: Revealing the Landscape of Privacy-Enhancing Technologies in the Context of Data Markets for the IoT: A Systematic Literature Review. *In https://arxiv.org/pdf/2107.11905.pdf*

AI is used across the enterprise, and is key to harnessing future applications



How can we grow our market share and win more business?
Director, Strategy

How can we identify alternative solutions to various risks in the supply chain and distribution network?
Director, Logistics

How can we increase efficiency and effectiveness of our operations?
Director, Operations

How can we manage our risk and ensure compliance?
Director, Regulatory Compliance

Outbound Logistics

Strategy & Growth

Customers & Marketing

Sales & Distribution

Operations

Development

Product Development

Operations

Service & Support

Inbound Logistics

Risk, Regulation, Compliance

Finance, HR, Capital

How can we manage relationships with our clients?
Director, Customer Relationships

How do we innovate and introduce new products and services?
Director, Products

How do we provide support more efficiently and increase customer satisfaction?
Director, Support

How do manage the influx of resumes with our personal data protection policies and desire to mitigate bias?
Director, HR

And come to organizations through…

Vendor Solutions

Data & Analytics Platforms

Open Source Software

Bespoke Development

Consider future applications that combine many of these, e.g. the metaverse. Are you prepared?

Apply what you have learned today

- Assess your organization & use of AI/ML
- Leverage existing data protection programs & governance
- Don’t forget vendor risks
- What to do tomorrow?
  - Start the conversation: what is our AI governance strategy?
- What to do next month?
  - Join the conversation: where privacy pros aren’t already included in the AI governance strategy, find a way to insert privacy. Where privacy pros are included in the AI governance strategy, identify where privacy should take the lead and where it should follow.
  - Are our AI risks on the data or modeling side? Are the benefits of our AI being mapped and measured?
- What to do by next year?
  - Create or disseminate your AI governance strategy within the organization

Questions?

Resource List

- Hogan Lovells Chronicle of Data Protection: Series on AI Regulation
  - The emerging regulatory environment (14 April 2021)
    - https://www.engage.hoganlovells.com/knowledgeservices/news/ai-algorithms-part-1-the-emerging-regulatory-environment
  - The EU releases its new regulation on artificial intelligence (15 April 2021)
    - https://www.engage.hoganlovells.com/knowledgeservices/news/ai-algorithms-part-2-the-eu-releases-its-new-regulation-on-artificial-intelligence
  - Why the EU's AI regulation is a groundbreaking proposal (3 May 2021)
    - https://www.engage.hoganlovells.com/knowledgeservices/news/ai-algorithms-part-3-why-the-eus-ai-regulation-is-a-groundbreaking-proposal
  - The FTC's guidance on AI (14 June 2021)
    - https://www.engage.hoganlovells.com/knowledgeservices/news/ai-in-the-us-the-federal-trade-commissions-guidance-on-ai
  - UK government announces plan to regulate artificial intelligence (4 October 2021)
    - https://www.engage.hoganlovells.com/knowledgeservices/news/uk-government-announces-plan-to-regulate-artificial-intelligence-as-part-of-new-national-ai-strategy
- Salesforce: Privacy and ethical use principles guiding our COVID-19 response
  - https://www.salesforce.com/content/dam/web/en_us/www/documents/legal/Privacy/privacy-and-ethical-use-principles-guiding-our-covid-19-response.pdf
- PwC: Responsible AI, Maturing from Theory to Practice
  - https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/pwc-responsible-ai-maturing-from-theory-to-practice.pdf