

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

# Robot Lawyers

## Automating Legal Compliance for Transferring Private Data

Stephen Chong
Joint work with Micah Altman and Alexandra Wood

Supported by the NSF and the Sloan Foundation

The Dataverse Project



Use and share data

Data Tags

Data Use Agreements



Social Scientist

?



Privacy Legislation

Analysis in legal memos

Best practices

Statutory text

# Data use agreements (DUAs)

- Contracts that govern transfer of data containing personal information that is subject to some restriction on its use.
  - Limits on use, obligations to safeguard, liability for harm arising from use or misuse, publication requirements and restrictions, privacy rights
  - Often required by statute, regulation, institutional policy, or pre-existing contract

Wood, O'Brien & Gasser, “Contractual Approaches to Research Data Sharing: A Comparative Analysis of Institutional Data Sharing Agreements and Policies”, Working Paper (forthcoming)

# Data use agreements (DUAs)

- Contracts that govern transfer of data containing personal information that is subject to some restriction on its use.
  - Limits on use, obligations to safeguard, liability for harm arising from use or misuse, publication requirements and restrictions, privacy rights
  - Often required by statute, regulation, institutional policy, or pre-existing contract
- What goes in a DUA?
  - Regulations may provide express requirements and guidelines
    - E.g., Health Insurance Portability and Accountability Act (HIPAA) requires some DUAs to include permitted uses and disclosures, individuals permitted to access data, restrictions against identifying or contacting the individuals in data set.
    - E.g., Family Educational Rights and Privacy Act (FERPA) requires some DUAs to describe purpose, scope, and duration of the study
  - Terms derived from agency guidance documents
    - Mandatory requirements or recommendations based on best practice

# DUAs

- Standard data use agreements
  - Boilerplate and clickthrough agreement(s)
  - Typically used by large data repositories
  - Coarse-grained, don't accurately capture permissions/obligations for specific dataset
- Transfer-specific data use agreements
  - Customized for the specific data transfer
  - Typically requires human (lawyer) effort
  - Requires expertise in privacy legislation (i.e., can still be inaccurate)
- Ideally: Automated generation of modular DUAs










Social Scientist

The Dataverse Project

Use and share data

Data Tags

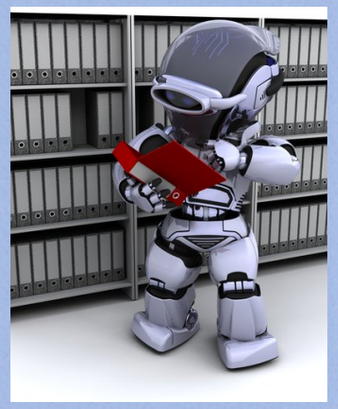
Data Use Agreements

This work

Tag characterization

License generation

Formal model
of regulatory requirements and best practice





Privacy Legislation

Analysis in
legal memos

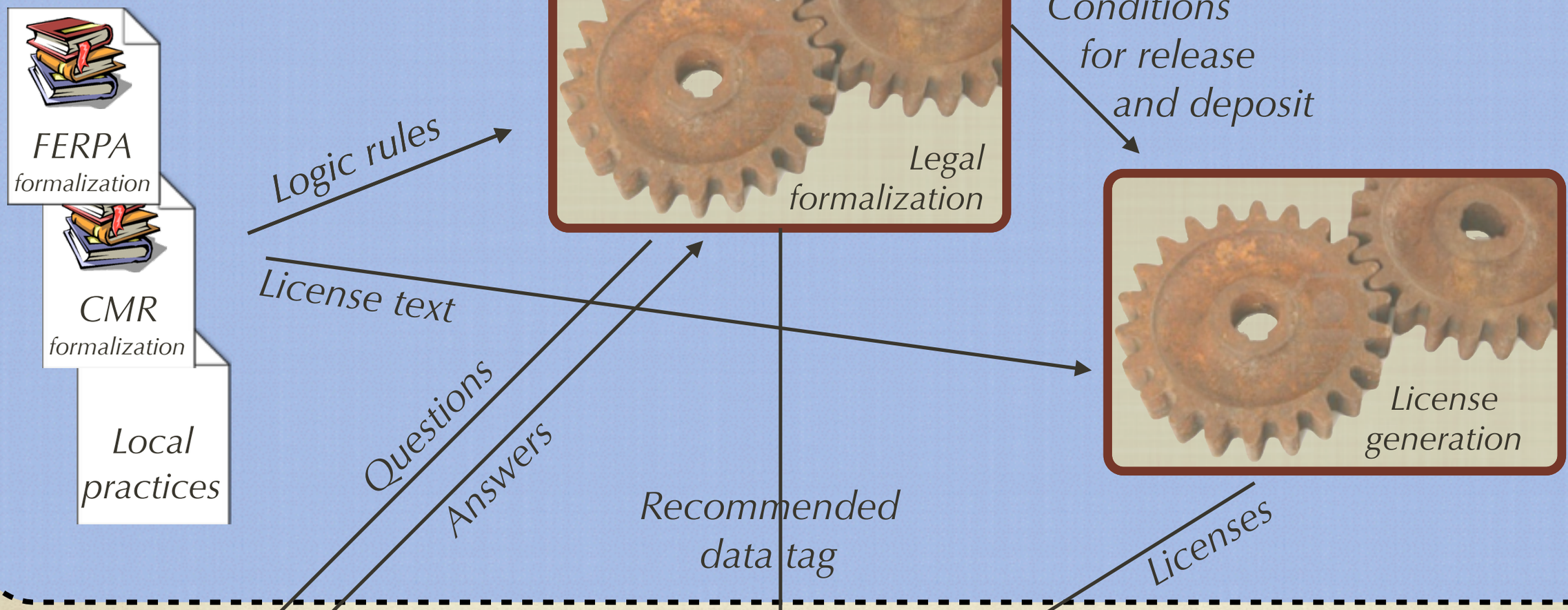
Best practices

Statutory text

# System overview

- Limited domain
  - Data repositories
    - Aspects of privacy legislation related to storing and transferring data
- Focus on subset of conditions that can be automatically handled
  - Not intended to handle 100% of situations
  - May need to be escalated to a human...
- Work in progress...
  - Prototype implementations of components

# System overview

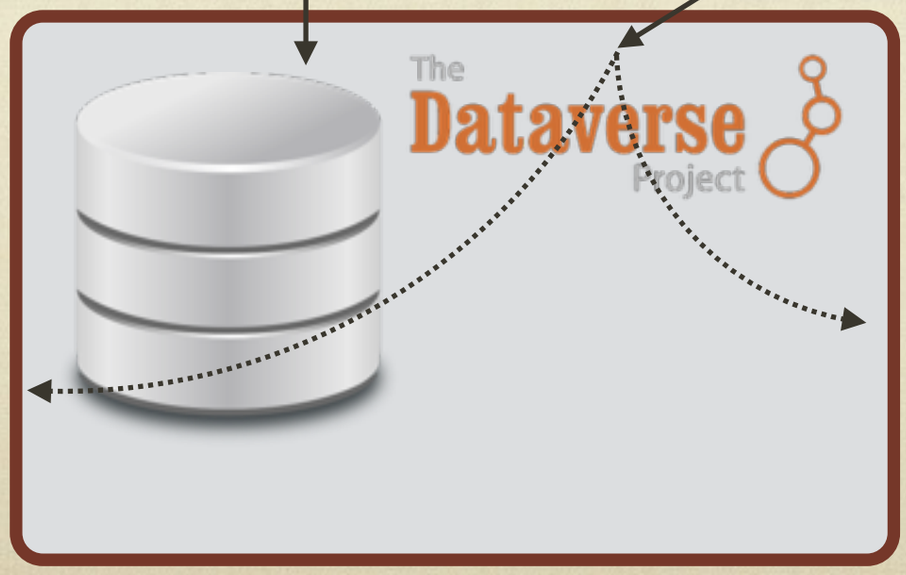





Data Owner

Data deposit

Deposit license



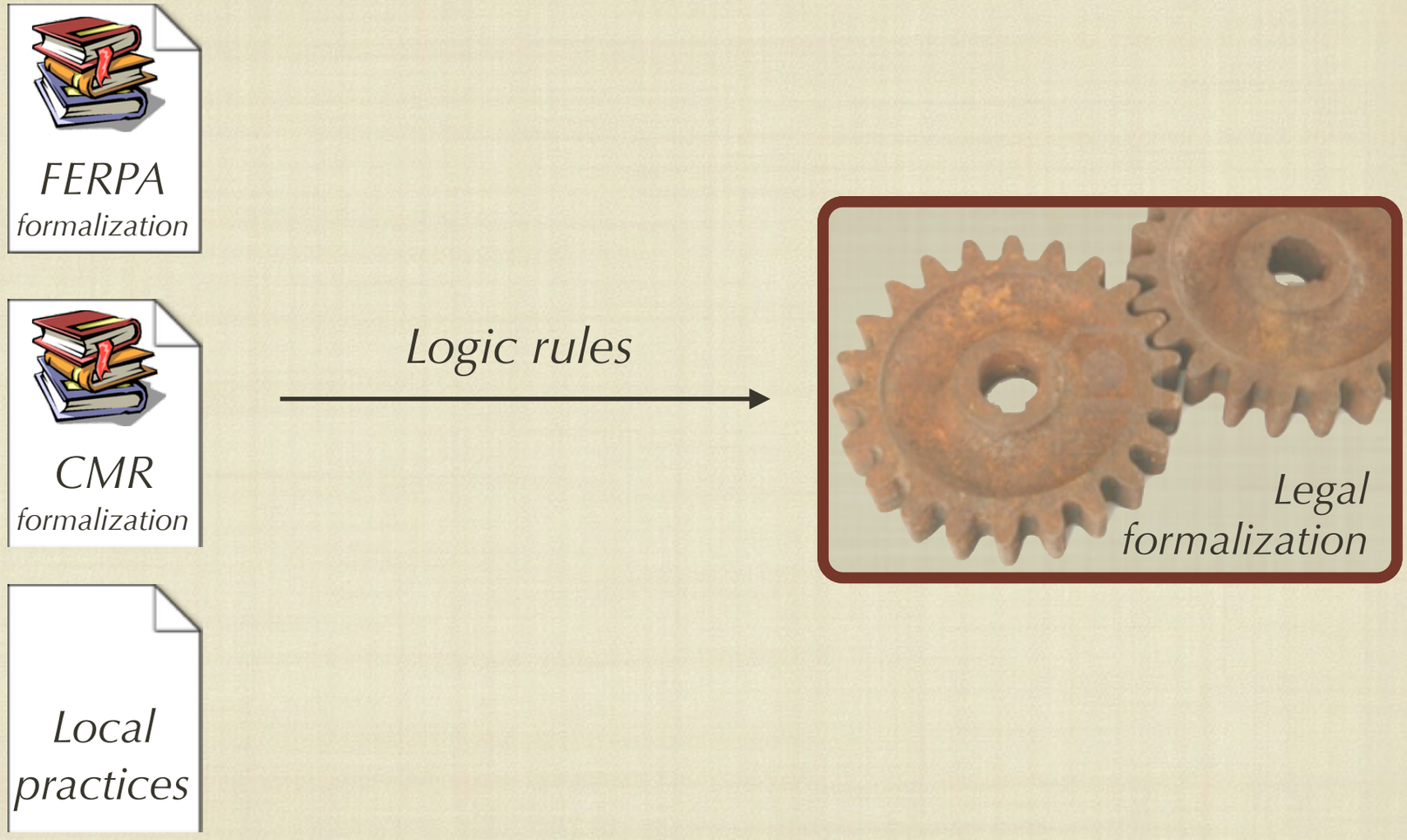

Data use
agreement

Data



Data User

# System overview



# Legal Formalization

- Formalize aspects of privacy legislation
  - Using a logic programming language
- Answer whether legislation/best practice permits or denies specific actions on data sets
  - Expert-system-like ability
- Explore legislation
  - e.g., find conditions where best practice contradictory
- Combines
  - computer science (formal modeling),
  - law (legal research & analysis),
  - social science (survey design),
  - information science (taxonomies)

# Formal model: Actions

`dd` : Data depositor

`r` : Repository

```
Deposit(dd, ds, r, cs)
Accept(r, ds, dd, cs)
Release(r, ds, du, dd, cs)
...
```

`ds` : Dataset

`du` : Data user

`cs` : Condition set
*(provides further details about action)*

# Permitted or Denied

- Actions can be **permitted** or **denied** by legislation

```
Permitted(leg, a)
Denied(leg, a)
```

`leg` : Legislation

`a` : Action

  - Or neither permitted or denied
- E.g., `Denied(ferpa,`

```
        Release(harvardDataverse,
                cs152grades-2015sp,
                jon@doe.com,
                chong@seas.harvard.edu,
                [dataverseClickthrough]))
```

# Example formalization

```
Let dd be the data depositor
Let du be the data user
Let ds be the data set
Let r be the repository
Let cs be a set of conditions

IF  CMR:depositorInScope(dd, ds)
AND CMR:identifiable(ds)
AND NOT (CMR:secure(r)
         AND CMR:isAcceptableConditionsForRelease(cs))
THEN DENIED(cmr, Release(r, ds, du, dd, cs))

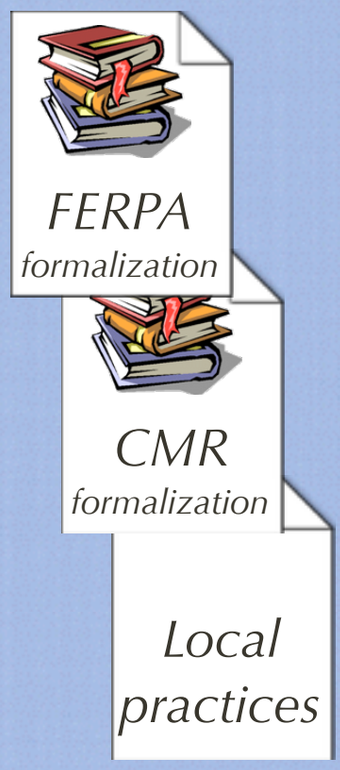

Let l be a license
Let cs be a set of conditions
IF  License(l) ∈ cs
AND licenseImplies(l, CMR:TransmissionEncrypted)
THEN CMR:isAcceptableConditionsForRelease(cs)
```

# Formalization Process

- Process should be reproducible

| | | | | |
|---|---|---|---|---|
| Define Use Cases In Scope | Data Deposit | Retention | Transformation | Dissemination |
| Parse Use Case Components | Actors: data controller, repository, user | Actions: accept data, store, release,… | Entities: Data set, Record,… | |
| Legal Review of Law | Identify Restrictions on Use Cases | Identify Restrictions on Actors, Actions, Entities | | |
| Expert Coding of Use Case Actions | Map rules permitting and restricting action to law-specific characteristics | Map law-specific characteristics to license text affirmations | Map law-specific characterstics to general properties | |

# System overview




Logic rules

License text

Questions

Answers




Conditions
for release
and deposit




Recommended
data tag

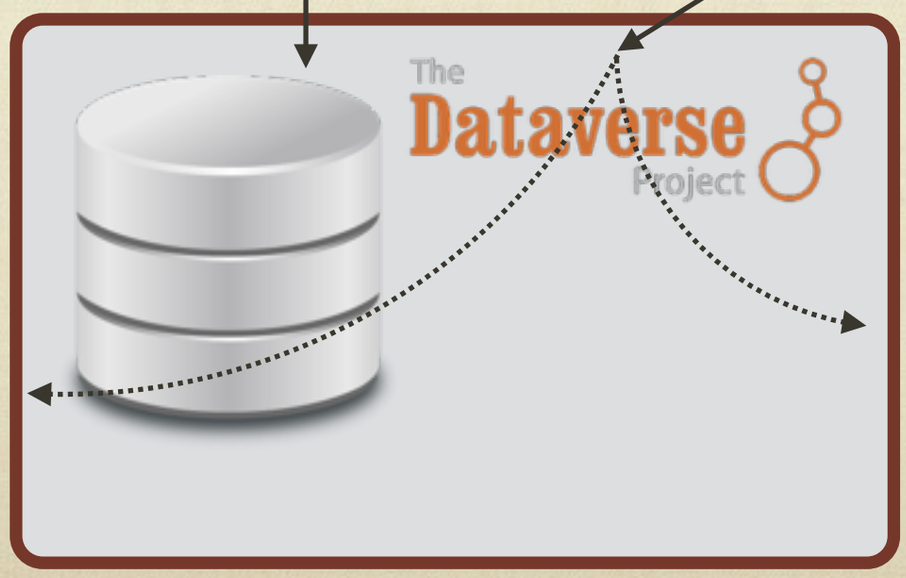
Licenses



Data Owner

Data deposit

Deposit license



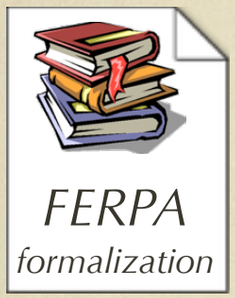

Data use
agreement

Data



Data User

# System overview







Local practices

License text




Conditions for release and deposit




Licenses

# License generation

- Input:
  - Which laws are relevant
  - What restrictions are required on data transfer
    - E.g., must be stored encrypted
  - License template
    - Outline/structure of license
  - License terms
    - Text
    - Conditions under which to include
  - License specifics
    - Name of repository, name of recipient, dates of study, ...
- Output:
  - Human-readable license

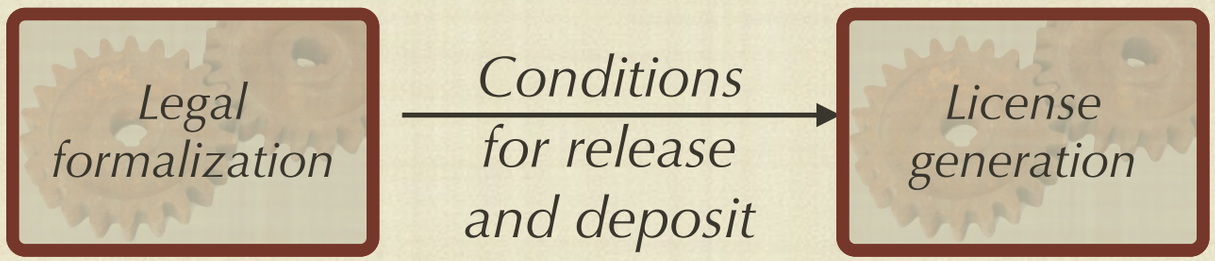
# Demo of License Generation

# Related Work

- Formalization of legislation (esp. privacy)
  - "The British Nationality Act as a Logic Program" Sergot et al. (1986)
  - "Privacy and Contextual Integrity: Framework and Applications" Barth et al. (2006)
  - "Privacy APIs: Access control techniques to analyze and verify legal privacy policies" May at al. (2006)
  - "Analyzing regulatory rules for privacy and security requirements" Breaux and Anton (2008)
  - "Reasoning about conditions and exceptions to laws in regulatory conformance checking" Dinesh et al. (2008)
  - "A formalization of HIPAA for a medical messaging system" Lam et al. (2009)
  - "Experiences in the Logical Specification of the HIPAA and GLBA Privacy Laws" DeYoung et al. (2010)
  - "PriCL: Creating a Precedent. A Framework for Reasoning about Privacy Case Law" Backes et al. (2015)
  - "A Semi-Automated Methodology for Extracting access control rules from the European Data Protection Directive" Faterna et al. (2016)
  - ...
- Automated generation of licenses
  - ???

# Current status

- Need to rework legal formalization
  - Prolog insufficient to capture reasoning for tag recommendations
- Need to integrate legal formalization and license generation




- Summer interns exploring extensions:
  - Support economic data protected by NDAs
  - Support MOOC datasets
  - Formalization for HIPAA-covered medical research data

# Future work

- From Prototype to Toolkit
  - Templates for data deposit, retention statements.
- Add new laws and domains
  - Including HIPAA, best practice, Common Rule.
- Legal review and publication
  - External legal review and formal publication of vetted rules bases.
- Modular license design
  - More principled automated composition of modular terms
- Integrating differential privacy
- Reproducing formalization
  - Develop coding rules and measure interrater reliability
- Integration with DataTags
- Including other Actions & Actors
  - Currently limited to small set of actions/actors most useful to repositories.
- Interview questions

# Robot Lawyers:
## Automating Legal Compliance for Transferring Private Data

