

Automatic Retrieval of Privacy Factors from IoMT Policies: ML and Custom NER Approach

Nyteisha Bookert and Mohd Anwar
Symposium on Usable Security and Privacy (USEC) 2023
February 28, 2023



Internet of Medical Things (IoMT)

Collection of devices and applications connected to the internet to provide healthcare services





Goldman Sachs says a digital healthcare revolution is coming — and it could save America \$300 billion



IoMT Market to Reach USD 187.60 Billion by 2028; Surging Expenditure on IoT in Healthcare to Boost Market Growth

**The internet of medical things:
Revolutionising healthcare**



Google to use patient data to develop healthcare algorithms for hospital chain

The company will get access to anonymized patient data

2 million affected by Shields Health Care Group cyberattack

Ransomware attack on Yuma Regional Medical leads to data theft for 700K patients

NYC hospital worker charged with stealing patient info

By THE ASSOCIATED PRESS
Apr 12, 2008 at 3:17 pm

YOUR DATA AND PRIVACY

For seniors using tech to age in place, surveillance can be the price of independence

To age in their own homes, seniors are juggling being watched with being on their own

Hospitals are selling treasure troves of medical data — what could go wrong?

They don't need patient consent to use de-identified data



Privacy Policies

- Promote Transparency
- Notification
- Clear and conspicuous



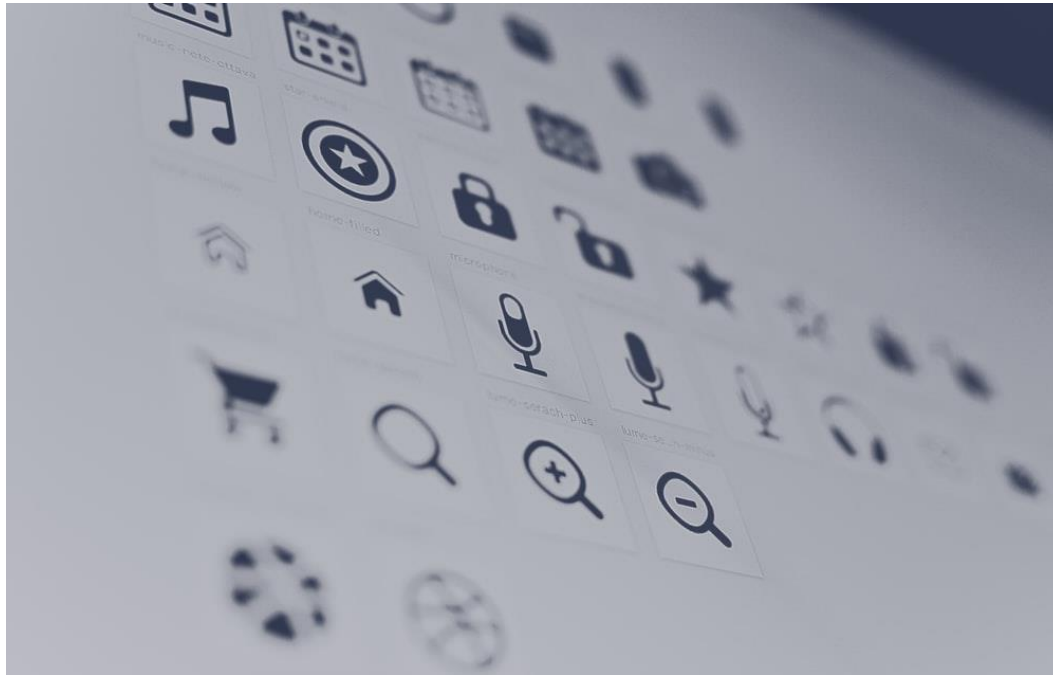
Privacy Policies as a Solution

- ~~Promote Transparency~~
- ~~Notification~~
- ~~Clear and conspicuous~~
- Comprehensive
- Difficult to comprehend
- Vague
- Time-consuming to read





Alternative Privacy Policies



- Companies lack incentives
- Not legally binding



Privacy Policy Analysis

- Themes
 - > Privacy Policy Assessment Criteria
 - > Privacy Policy Content Analysis
- Domains
 - > Websites
 - > Mobile applications
 - > Internet of Things devices

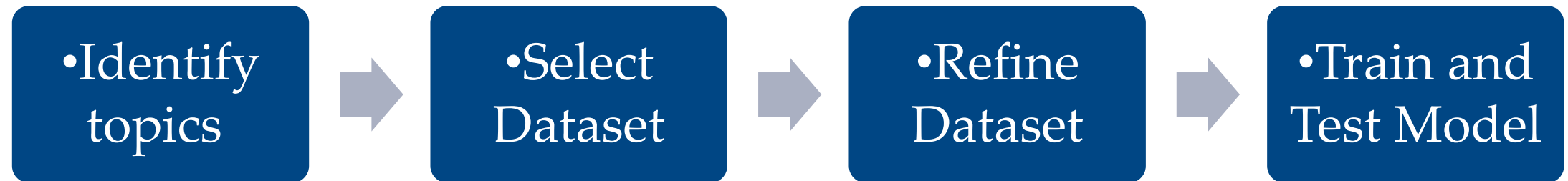


Research Questions & Goals

- Evaluate appropriateness of combining machine learning and custom named entity recognition techniques
- Create IoMT-specific fine-grained annotations



Methodology





Privacy Policy Assessment Questionnaire (PPAQ)

Evaluation criteria framework for the privacy policy of IoMT Devices

01

General

04

Data Retention

07

Children

02

Data Collection

05

Data Security

03

Data Sharing

06

User Choice

(Bookert et al., 2022)



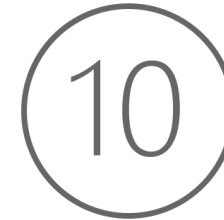
The Online Privacy Policies, set of 115 Corpus (OPP-115)



115 Websites



Legal and privacy
experts



Privacy Factors

(Wilson et al., 2016)



OPP-115 Mapping to PPAQ

PPAQ	OPP-115 Factor
Data Collection	First Party Collection/Use
Data Sharing	Third Party Sharing/Collection
Data Retention	Data Retention
Data Security	Data Security
User Choice	User Access, Edit and Deletion
	User Choice/Control
Children	International and Specific Audiences, Audience Type: Children
Contact Information	Other, Other Type: Privacy Contact Information
Effective Date	Other, Other Type: Introductory/Generic
Change Notification Process	Policy Change



Automatically Identifying PPAQ Categories

Multi-Label Classification Results

ML Model	F1	Precision	Recall
Data Collection	80%	81%	79%
Data Sharing	80%	86%	75%
User Choice	68%	82%	58%
Data Retention	32%	44%	25%
Data Security	67%	81%	56%
Policy Change	86%	97%	77%
Children	85%	100%	74%
Contact Info	63%	69%	58%



Custom Named Entities

- Address
- Children: Age
- Data Retention: Period
- Data Sharing: Recipient
- Date
- Does/Does Not
- Email
- Personal Information Type
- Phone Number
- URL
- User Choice



Refining Dataset

OPP-115 Label: Children

...The Atlantic does not knowingly collect personal information from children under 13 years of age. If you are under 13 years of age, do not provide personal information to The Atlantic without providing us with consent from your parents...



Refining Dataset

OPP-115 Label: Children

...The Atlantic does not knowingly collect personal information from children **under 13 years of age**. If you are **under 13 years of age**, do not provide personal information to The Atlantic without providing us with consent from your parents...



Automatically Identifying Fine-Grained Annotations

Custom NER Preliminary Results

Entity	F1-score	Recall	Precision
Performance	45.48%	50.60%	41.31%
Address	59.26%	80.00%	47.06%
Children: Age	85.71%	96.43%	77.14%
Data Retention: Period	6.67%	7.69%	5.88%
Data Sharing: Recipient	36.71%	49.34%	29.23%
Date	85.71%	88.24%	83.33%
Does/Does Not	31.35%	38.02%	26.67%
Email	88.24%	96.77%	81.08%
Personal Information Type	48.88%	48.05%	49.74%
Phone Number	80.00%	80.00%	80.00%
URL	67.39%	76.54%	60.19%
User Choice	58.36%	63.98%	53.65%



Annotated Privacy Policy

Data Collection

Data Sharing

In the normal course of business, Qualtrics **collects Does/Does Not** **Personal Data Personal Information Type** such as **contact information Personal Information Type** (e.g., **name Personal Information Type**, **address Personal Information Type**, **phone number Personal Information Type**, **email address Personal Information Type**, and your **employer Data Sharing: Recipient**) and **payment details Personal Information Type** for customers.



Future Work

- More granularity for data collection and sharing practices
- Test on IoMT privacy policies
- Automatically detect potential mismatches



Thank you!

Nyteisha Bookert

Ph.D. Candidate

Human-Centered AI Lab

North Carolina A&T State
University

Email: ndbooker@aggies.ncat.edu

Website: nyteishabookert.me