



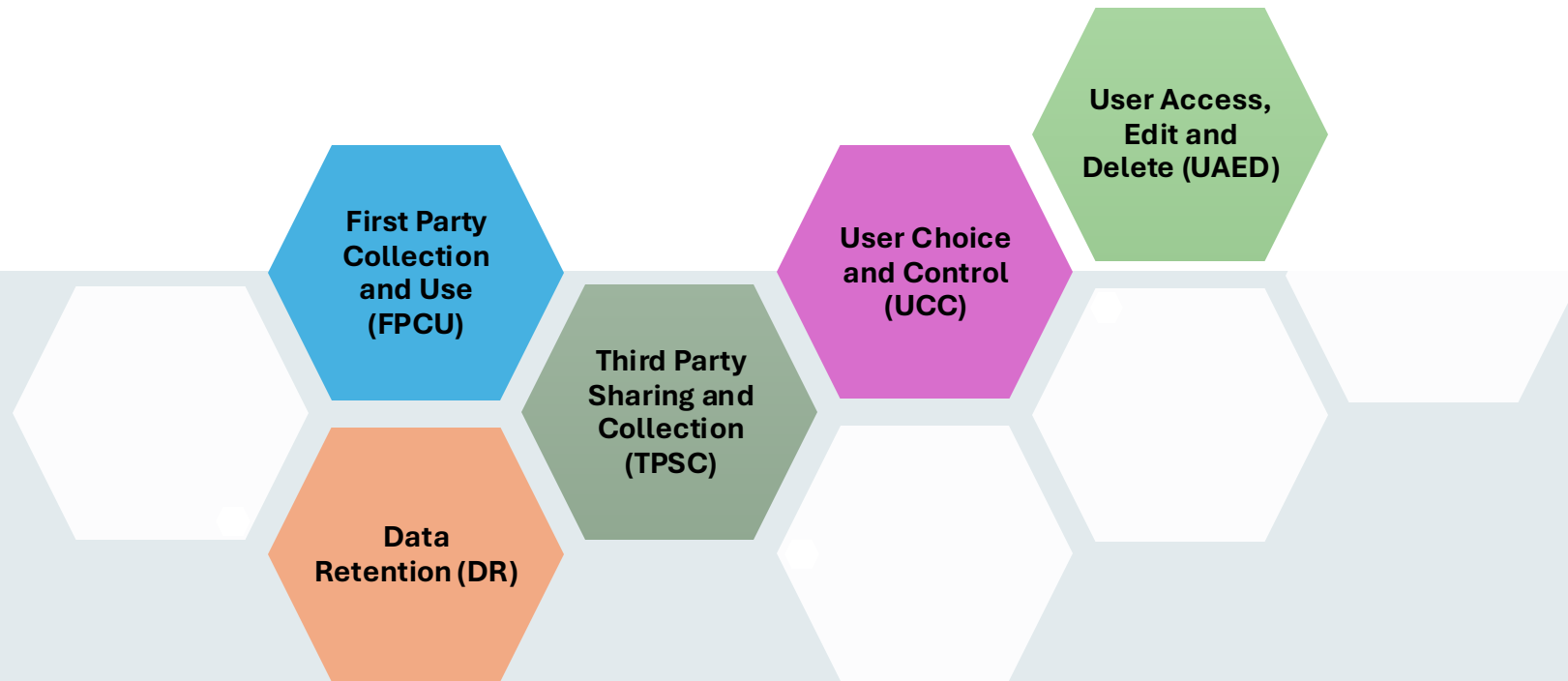
POLICYPULSE

Precision Semantic Role Extraction for
Enhanced Privacy Policy Comprehension

Andrick Adhikari, Sanchari Das, and Rinku Dewri

What is a Privacy Policy?

a contractual document about data practices facing users and regulatory bodies



Comprehension Challenges in Privacy Policies

Length

Complex Language

Ambiguity

Unstructured Text

You have choices when it comes to the technology you use and the data you share. When we ask you to provide personal data, you can decline. Many of our products require some personal data to provide you with a service. If you choose not to provide data required to provide you with a product or feature, you cannot use that product or feature. Likewise, where we need to collect personal data by law or to enter into or carry out a contract with you, and you do not provide the data, we will not be able to enter into the contract; or if this relates to an existing product you are using, we may have to suspend or cancel it. We will notify you if this is the case at the time. Where providing the data is optional, and you choose not to share personal data, features like personalization that use such data will not work for you.

Excerpt from an actual policy

Addressing Privacy Policy Challenges

Alternative Designs

Short Notices

concise summaries to improve readability

Multi-Layer Policies

structured formats presenting key information first, with details layered below

Graphical Representations

visual elements to aid comprehension

NLP on Privacy Policies

Text Classification

categorizing paragraphs based on privacy specific topics (e.g., OPP-115)

Information Extraction

tailored specifically to requirement (e.g., data type being collected, control choices etc.)

Other Applications

automated query answering, text alignment, summarization

Road So Far...



Information extraction coverage

Deficiency in extracting comprehensive policy artifacts



Paragraph/sentence level classification

Lacks the granularity needed to extract artifacts related to a specific practice



Specialized for a specific task

Another method is needed for another task

PolicyPulse

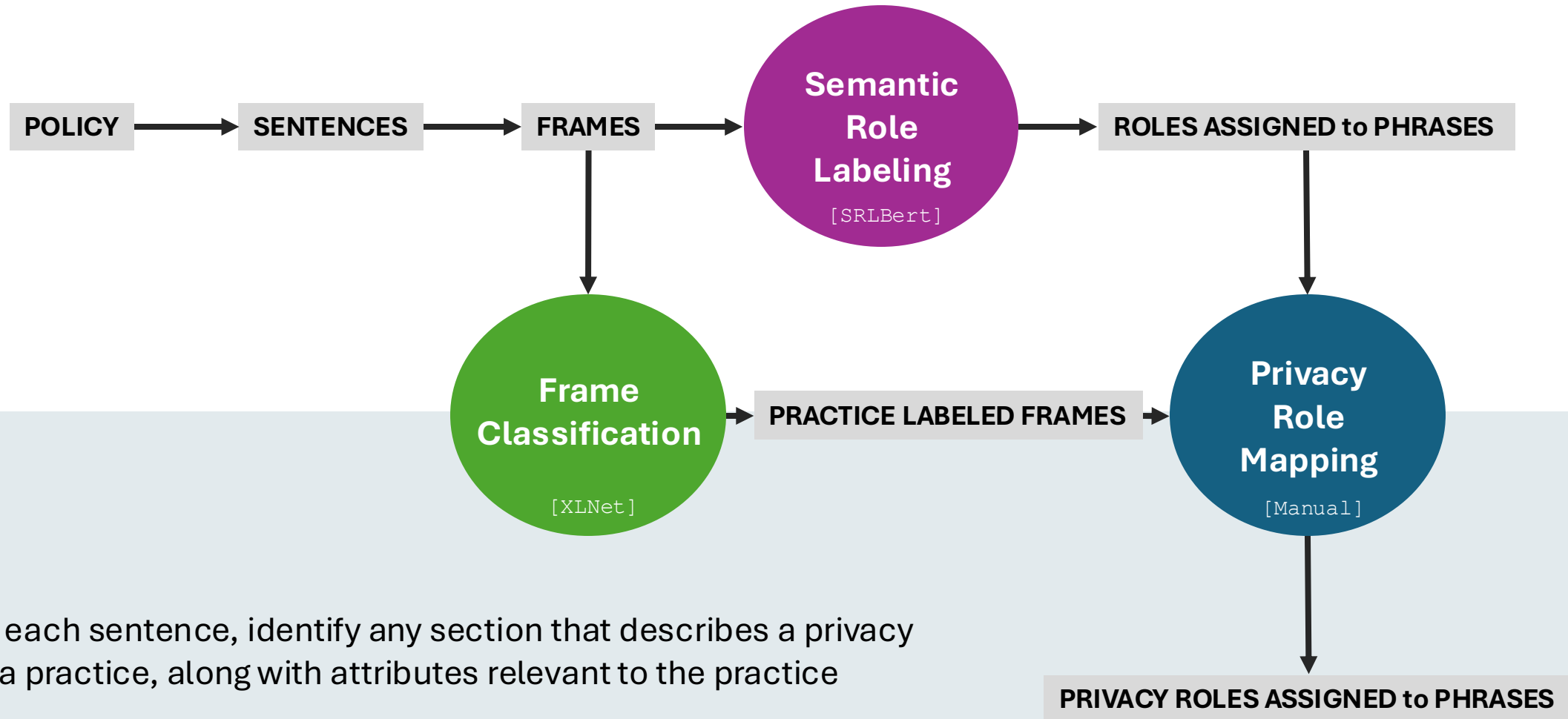


an information extraction pipeline that uses NLP techniques to identify broader privacy-relevant categorical components within a clause in a sentence, while maintaining their semantic relationships

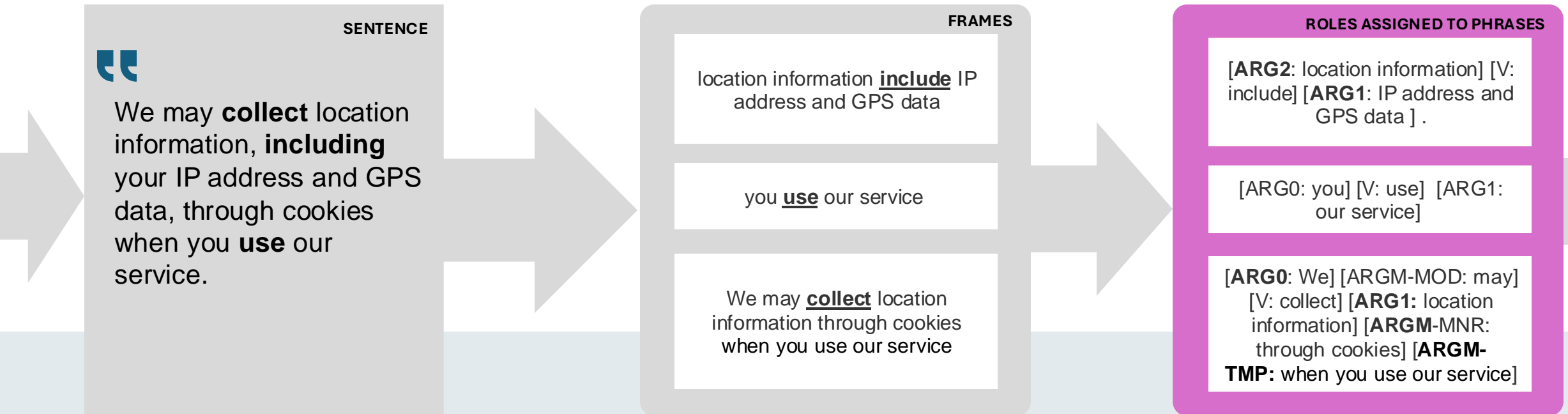


data/code artifacts available from **github.com/crisp-du/ppevo**

PolicyPulse: High Level Overview

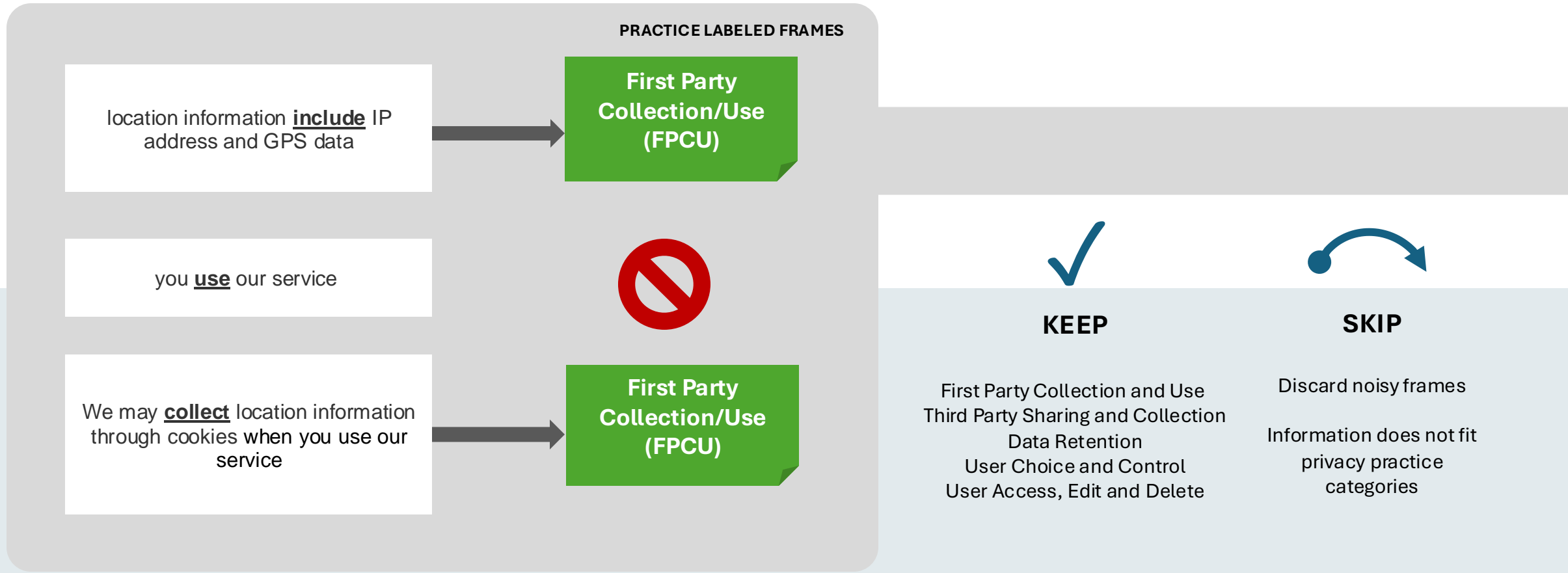


Semantic Role Labeling (SRL)

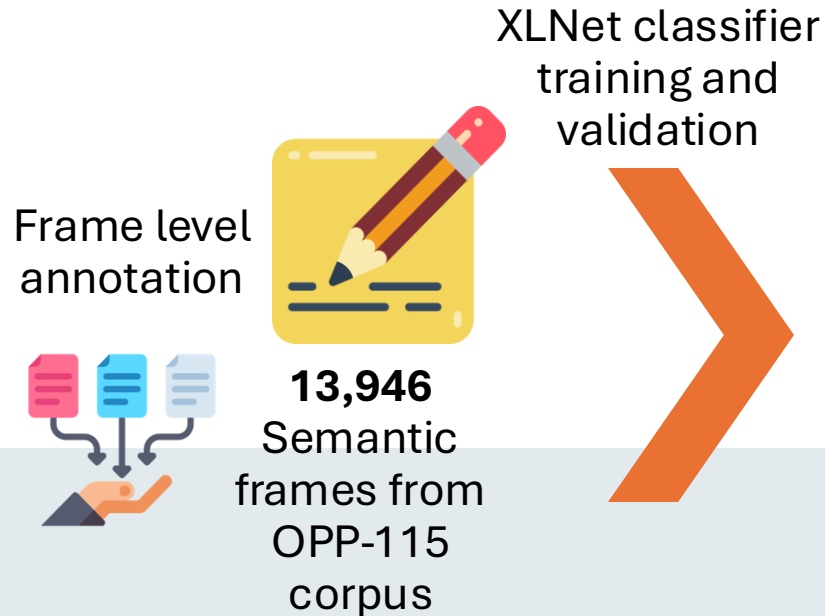


The roles played by different phrases in a sentence or clause with respect to a specific predicate are labeled using SRLBert from AllenNLP

Semantic Frame Classification



Semantic Frame Classifier



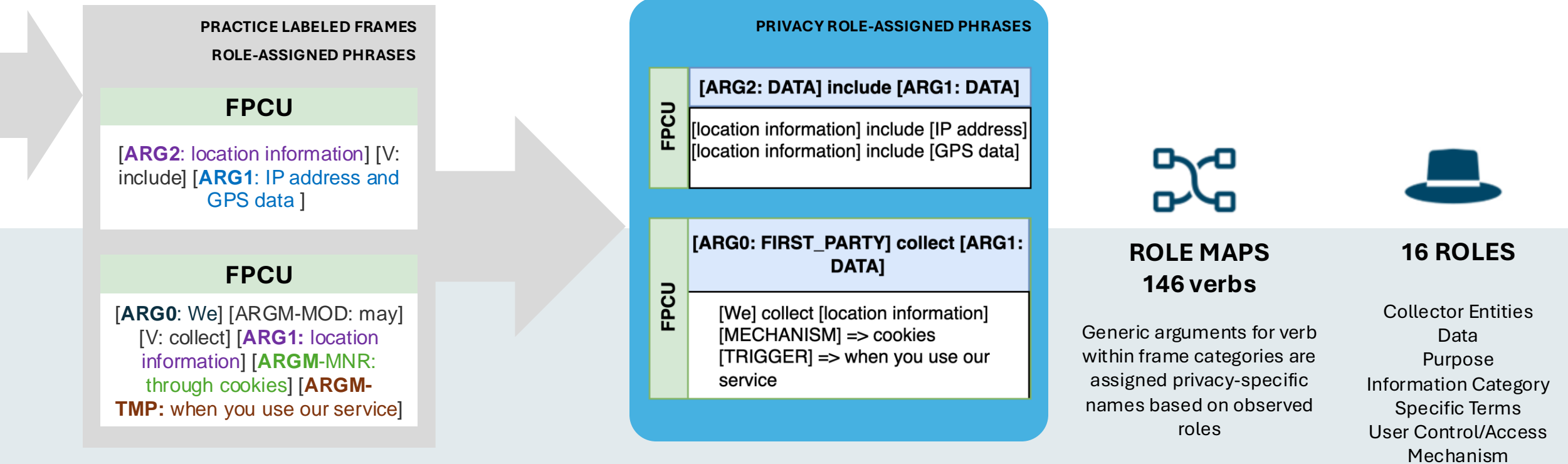
	Precision	Recall	F1
Frame	0.61	0.61	0.58
Frame with Semantic Context (FSC)	0.58	0.57	0.53
FSC with Synthetic training data	0.77	0.89	0.82
FSC Augmented (Two-Level)	0.93	0.95	0.94

Category	Precision	Recall	F1
SKIP	0.99	0.98	0.98
KEEP	0.93	0.96	0.94
macro avg	0.96	0.97	0.96
weighted avg	0.97	0.97	0.97

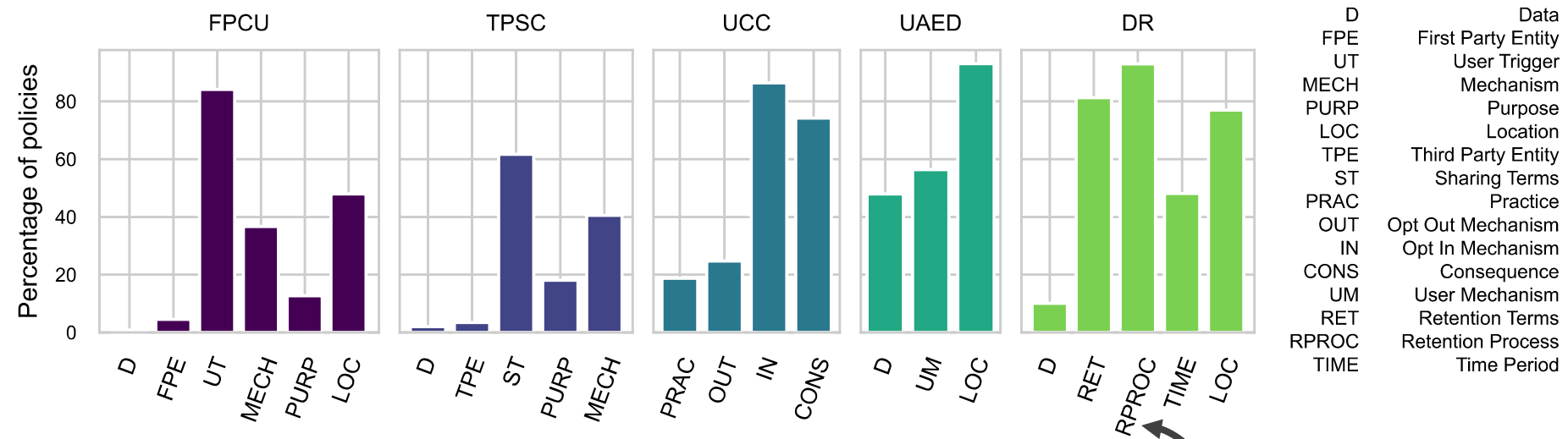
Category	Precision	Recall	F1
FPCU	0.98	0.98	0.98
TPSC	0.99	0.98	0.98
UAED	0.98	0.98	0.98
UCC	0.98	0.98	0.98
DR	0.96	0.97	0.96
macro avg	0.98	0.98	0.98
weighted avg	0.98	0.98	0.98

Nested 10 x 9 cross-validation performance scores with a 9:1 train/test split of annotated semantic frames

Mapping to Privacy Specific Roles



PolicyPulse Application: Completeness Analysis



Percentage of policies with missing roles in a practice category computed from 129,856 policies (PPCrawl corpus)

not giving details on retention process, terms, and location is common

PolicyPulse Application: Alternative Presentation

PRIVACY NOTICE

When you register with Yahoo, we collect

- Personal information, name, address, email address, or phone number
- Device identifiers include IP address, browser version, OS type and version
- Non personal information

With whom do we share?

- Non-affiliated companies
- Marketing partners
- Family Companies



How long do we keep your data?

UNSPECIFIED

Your personal information may be transferred to countries other than your own to process and store data in accordance with our Privacy Policy and to provide you with products and services.

Expanded Information

Yahoo does not rent, sell, or share **personal information** about you with other people or **non-affiliated companies** except to provide products or services you've requested, when we have your permission, or under the following circumstances

types of information	how we use your information								who we share your information with		
	marketing	basic service feature	additional service feature	service operation and security	personalization and customization	analytics/research	merger/acquisition	advertising	third party	partner	advertiser
personal information	!	UTR	UTR	!	-	!	!	-	UTR	UTR	!
device identifier	-	-	-	-	!	-	-	!	-	-	-
geolocation	-	!	-	-	-	-	-	-	-	-	-
contact information	!	!	-	!	-	!	-	-	!	-	-
online activities	-	-	-	-	-	-	-	-	!	-	-
social media data	-	-	UTR	-	-	-	-	-	UTR	-	-
cookie/pixel tag	-	-	!	-	!	UTR	UTR	-	!	-	-
financial information	-	-	-	!	-	UTR	-	-	!	-	-

! information will be used for the purpose or shared
 UTR some user action will trigger use or sharing of the information
 - information will not be used for the purpose

PolicyPulse Application: Question Answering

Example question	PolicyPulse on Question Category & Privacy role		Frame with Highest Semantic Similarity	Extracted Answer [ARGUMENT]
What information is shared with advertisers?	TPSC	DATA	We use the [ARGUMENT] to help advertisers and other partners measure the effectiveness and distribution of their ads and services, and understand the types of people who use their services and how people interact with their websites, apps, and services	information we have (including your activity off our Products, such as the websites you visit and ads you see)
How long is my information kept?	DR	TIME PERIOD	We store data [ARGUMENT]	until it is no longer necessary to provide our services and Products, or until your account is deleted—whichever comes first

Road Ahead...



PolicyPulse

Granular analysis to facilitate broader information extraction



Refinements

Capturing relationships across sentences; classification-based role mappings; more categories and roles



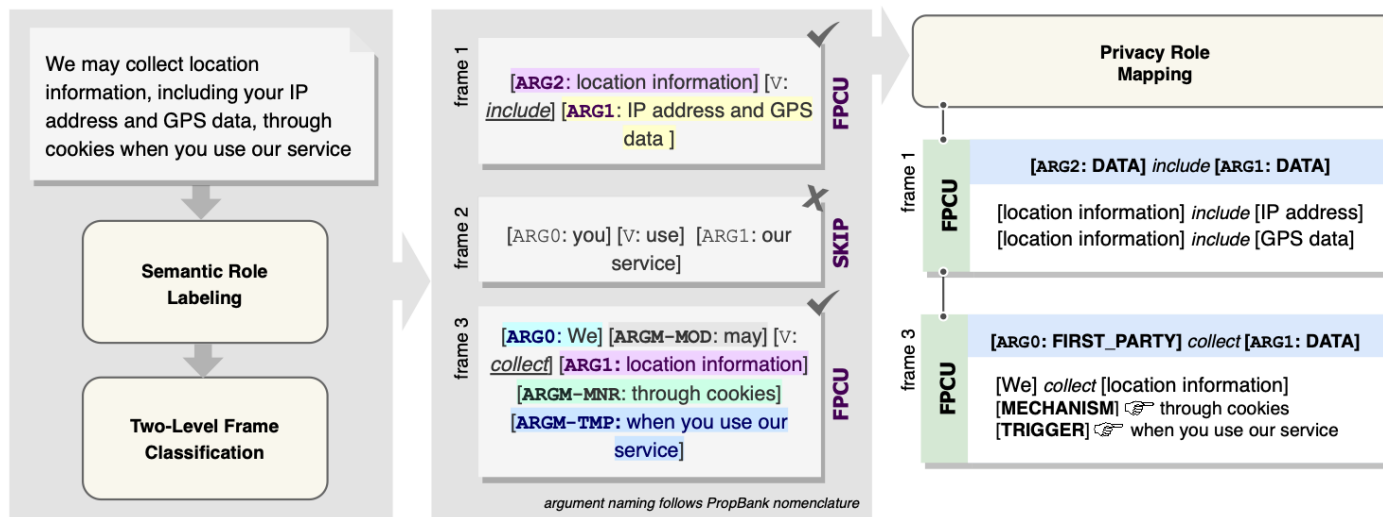
Canonical representation

Reduce extracted data to a minimum set



New analytics

Identify novel uses of data extracted at granular level



data/code artifacts available from



github.com/crisp-du/ppevo

OPP-115 frame annotations

XLNet frame classifier

Privacy role maps for 146 verbs

PolicyPulse

Precision Semantic Role Extraction for
Enhanced Privacy Policy Comprehension

Andrick Adhikari
andrick.adhikari@du.edu

Sanchari Das
sdas35@gmu.edu

Rinku Dewri
rinku.dewri@du.edu